STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **13**, April 2025, pp 1679–1689.
Published online in International Academic Press (www.IAPress.org)

Periodic Exponential Autoregressive Models for Rainfall Forecasting in Algeria

Sabah Becila *, Mouna Merzougui

Department of Mathematics, LaPS Laboratory, Badji Mokhtar Annaba University , Algeria

Abstract This study examines the utilization of periodic exponential autoregressive $(PEXPAR)$ models in analyzing rainfall time series data from Algeria. The method of Gaussian quasi maximum likelihood for parameter estimation is used. By comparing its forecasting performance with $SARIMA$ models, we observe a slight improvement with $PEXPAR_{12}(1)$, suggesting its potential efficacy in capturing seasonal variations and nonlinear behavior in precipitation data.

Keywords Periodic exponential autoregressive model, SARIMA model, Quasi-maximum likelihood, Rainfall in Algeria.

AMS subject classifications 62F12; 62M10

DOI: 10.19139/soic-2310-5070-2000

1. Introduction

This paper explores the application of Periodic Exponential Autoregressive models to analyze rainfall time series data from Algeria. Traditionally, hydro-climatic time series, such as rainfall, have been modeled using Seasonal Autoregressive Integrated Moving Average $(SARIMA)$ models due to their ability to capture seasonal characteristics. However, rainfall time series often exhibit nonlinear behavior, prompting the exploration of alternative models like the Exponential Autoregressive $(EXPAR)$ model, as demonstrated in previous research, [13] and related references. Understanding rainfall patterns in Algeria is crucial due to their profound implications for agriculture, water resource management, and environmental sustainability. The variability and complexity of rainfall data demand sophisticated modeling approaches to accurately capture its dynamics. There is a growing interest in $EXPAR$ models [17], [9], [18]). Recent work by [5] highlighted the suitability of $EXPAR$ models for modeling and forecasting sunspot annual numbers, showcasing the broader applicability of such models beyond hydro-climatic data. Moreover, the periodicity phenomenon in time series modeling has a rich history, dating back to [11], who introduced new models leveraging periodicity. This approach has been successfully applied in various fields, as evidenced by [15], [10] [14], [12], among others. $PEXPAR$ models represent a significant advancement, offering a generalization of $EXPAR$ models by allowing parameters to vary periodically with time, see [16]. This flexibility is particularly advantageous for modeling seasonal data characterized by nonlinear features and periodic autocovariance structures. In this study, we employ the quasi maximum likelihood estimation $(QMLE)$method to estimate

*Correspondence to: Sabah Becila (Email: sabah.becila@univ-annaba.dz). Department of Mathematics, LaPS Laboratory, Badji Mokhtar Annaba University, Algeria.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

$PEXPAR$ models, extending the approach from classical $EXPAR$ models, [6]. The use of $QMLE$ ensures consistent and asymptotically normal estimators, aligning with the nonlinear least square estimator (NLS) principles, [7].
The paper is organized as follows. Section 2 describes the $PEXPAR$ model and defines the $QMLE$ estimator. Section 3 presents a concise simulation study to assess the performance of the proposed methodology. Finally, the application of $PEXPAR$ models for analyzing rainfall data in Algeria is discussed in Section 4. Section 5 concludes the paper by summarizing key findings and outlining directions for future research.

2. QML Estimation of the $PEXPAR_S(1)$ Model

The process $\{Y_t;\ t \in \mathbb{Z}\}$ is said to follow a Periodic Exponential Autoregressive $PEXPAR_S\,(1)$, with period S $(S \geq 2)$, if it satisfies:

$$Y_t = \left(\alpha_{t,1} + \alpha_{t,2} \exp\left(-\gamma_t Y_{t-1}^2\right)\right) Y_{t-1} + \varepsilon_t,\ \ t \in \mathbb{Z}, \tag{1}$$

where $\{\varepsilon_t;\ t \in \mathbb{Z}\}$ is an independent and periodically distributed $(i.p.d.)$ process with mean 0 and finite variance σ_t^2 with a probability density $f\,(.)$ which is not specified and $\gamma_t > 0$. The autoregressive parameters $\alpha_{t,1}, \alpha_{t,2}\ \ \forall t \in \mathbb{Z}$, the nonlinear parameter γ_t and the innovation variance σ_t^2 are periodic, in time, with period S , i.e,

$$\alpha_{t+kS,1} = \alpha_{t,1}\ , \alpha_{t+kS,2} = \alpha_{t,2},\ \gamma_{t+kS} = \gamma_t \text{ and } \sigma_{t+kS}^2 = \sigma_t^2\ , \forall k,\ t \in \mathbb{Z}.$$

We define the conditional mean and conditional variance of the model. Let

$$\mu_t = E\left(Y_t / \mathcal{F}_{t-1}\right),$$

and

$$\Delta_t^2 = Var\left(Y_t / \mathcal{F}_{t-1}\right),$$

where $\mathcal{F}_{t-1} = \sigma\left\{Y_{t-s}, s \geq 1\right\}$ is the σ-algebra generated by the past of Y_t up to $t-1$. Then

$$\begin{aligned} \mu_t &= \left(\alpha_{t,1} + \alpha_{t,2} \exp\left(-\gamma_t Y_{t-1}^2\right)\right) Y_{t-1}, \\ \Delta_t^2 &= E\left(\left(Y_t - \mu_t\right)^2 / \mathcal{F}_{t-1}\right) = \sigma_t^2. \end{aligned}$$

The conditional mean is stochastic and depends on the past information. However, the conditional variance is deterministic and not stochastic. Our model primarily allows for predicting values but does not incorporate stochastic modeling of their variabilities. It is important to note that other periodic models exist that encompass both the conditional mean and variance stochastically. Notable examples include the periodic random coefficient AR (RCA) model, Aknouche and Guerbyenne (2009a, 2009b), periodic $ARMA$ models with periodic $GARCH$ components, Aknouche and Bibi (2009), and periodic autoregressive conditional duration models, Aknouche et al. (2022). These models provide a comprehensive framework by jointly modeling both the conditional mean and variance.
By letting $t = i + S\tau, i = 1, 2, ..., S$ and $\tau \in \mathbb{Z}$, equation (1) can be rewritten in the equivalent form:

$$Y_{i+S\tau} = \left(\alpha_{i,1} + \alpha_{i,2} \exp\left(-\gamma_i Y_{i+S\tau-1}^2\right)\right) Y_{i+S\tau-1} + \varepsilon_{i+S\tau},\ i = 1, ..., S,\ \tau \in \mathbb{Z}. \tag{2}$$

The vector of parameters

$$\theta = \left(\theta_1', ..., \theta_S'\ \right)' \in \mathbb{R}^{3S} \text{ where } \theta_i = \left(\alpha_{i,1}, \alpha_{i,2}, \gamma_i\right)', i = 1, ..., S,$$

belongs to a parameter space of the form:

$$\Theta^S \subset (]-1,+1[\times]-1,+1[\times]0,\infty[)^S.$$

The variance σ_t^2 can be considered as a nuisance parameter. The true value of the parameters is unknown and denoted by:

$$\theta_{i,0} = \left(\alpha_{i,1}^0, \alpha_{i,2}^0, \gamma_i^0\right)', i = 1, ..., S \text{ , and } \theta_0 = \left(\theta'_{1,0}, ..., \theta'_{S,0}\right)' \in \mathbb{R}^{3S}.$$

We want to estimate the true parameter θ_0 from observations $Y_1, ..., Y_n$ where $n = mS$. Given the initial value Y_0, the conditional log likelihood of the observations evaluated at θ depends on f which is unknown. The QML estimator is obtained by replacing f by the $N\left(\mu_t, \sigma_t^2\right)$:

$$L_n(\theta, Y_1, \ldots, Y_n) = -\frac{mS}{2}\log(2\pi) - \frac{m}{2}\sum_{i=1}^{S}\log\left(\sigma_i^2\right) - \sum_{i=1}^{S}\sum_{\tau=0}^{m-1}\frac{\left(Y_{i+S\tau} - \left(\alpha_{i,1} + \alpha_{i,2}\exp\left(-\gamma_i Y_{i+S\tau-1}^2\right)\right)Y_{i+S\tau-1}\right)^2}{2\sigma_i^2},$$

assuming $\sigma_i \neq 0$.

Let $\widehat{\theta}_n$ be the QML estimator, it can be observed that maximizing L_n is equivalent to the following minimization problem with respect to θ :

$$\widehat{\theta}_n = \arg\inf_{\theta \in \Theta^S} Q_n(\theta),$$

where

$$Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\frac{\left(Y_t - \left(\alpha_{t,1} + \alpha_{t,2}\exp\left(-\gamma_t Y_{t-1}^2\right)\right)Y_{t-1}\right)^2}{\sigma_t^2}.$$

We need the following assumptions to show the asymptotic properties of the estimator.

$A1$: The parameter space Θ^S is a compact set.

$A2$: $i)$ Y_t is periodically strictly stationary. A sufficient condition is

$$\max\left(|\alpha_{i,1}|, |\alpha_{i,1} + \alpha_{i,2}|\right) < 1 \text{ for } i = 1, ..., S.$$

$ii)$ The periodically white noise is such that $E\left(\varepsilon_t^2\right) < \infty$, for any $t \in \mathbb{Z}$.

$A3$: The parameter $\theta_{i,0}$ belongs to the interior $\overset{\circ}{\Theta}$ of Θ.

$A4$: $E(\varepsilon_t^6) < \infty$, for any $t \in \mathbb{Z}$.

Assumption $A3$ is standard for asymptotic normality and $A4$ is necessary for the existence of the variance of the score vector.

Theorem 1

Under assumptions $A1 - A2$, the quasi maximum likelihood estimators $\widehat{\theta}_{i,m}$ are strongly consistent, i.e. almost surely

$$\widehat{\theta}_{i,m} \to \theta_{i,0} \text{ as } m \to \infty.$$

And under all the assumptions above,

$$\sqrt{m}(\widehat{\theta}_{i,m} - \theta_{i,0}) \underset{m \longrightarrow \infty}{\overset{D}{\longrightarrow}} N(\underline{0}_3, 2J_i^{-1}),$$

where

$$J_i = E\left(\frac{\partial^2 \ell_{i,\tau}(\theta_i)}{\partial\theta_i \partial\theta_i'}\right),$$

is a positive definite matrix and

$$\ell_{i,\tau}(\theta_i) = \frac{\left(Y_{i+S\tau} - \left(\alpha_{i,1} + \alpha_{i,2}\exp\left(-\gamma_i Y_{i+S\tau-1}^2\right)\right)Y_{i+S\tau-1}\right)^2}{\sigma_i^2}.$$

Proof

Since the conditional QML estimator is equivalent to the nonlinear least squares (NLS) estimator, the proof follows the same approach as detailed in [7]. The consistency is established through an ergodicity argument, and the normality relies on a central limit theorem applied to martingale differences. □

Remark: For fixed i, we obtain the QML estimates of the classical $EXPAR(1)$ model in [6].

3. Simulation results

To assess the asymptotic properties of the QML estimator in finite samples, we conduct a small-scale simulation study. We generate time series from the $PEXPAR_2(1)$ models with sample sizes of $n = 300$, 600 and 1000 Monte Carlo replications. The mean values, biases and standard deviations of the QML estimations are reported across these different sample sizes. Additionally, we incorporated estimated asymptotic standard errors (ASE) derived from the asymptotic variances specified in Theorem 1 into our analysis. Table 1 presents the parameter estimations for $\theta = (-0.5, 0.3, 0.9; 0.2, -0.6, 1)'$. In Figure 1 and Figure 2, we display boxplots and Q-Q plots of the estimation errors, respectively, for the model with 100 replications and $n = 500$. Our simulation program, implemented in R using the nlm function, is sensitive to initial values. To mitigate this sensitivity, we introduce a random perturbation on the true parameter during initialization. Practically, we set $\widehat{\gamma}_i = \dfrac{\log \epsilon}{\max\limits_{0\leq\tau\leq m-1} Y^2_{S\tau+i-1}}$, where ϵ is a small number, and utilize linear least squares estimation for the remaining parameters as initial values. From the results in Table 1, we observe that the parameters are well estimated, with either the bias or standard deviation decreasing as n increases, or both, indicating the consistency of the estimates. Moreover, incorporating the estimated asymptotic standard errors provides additional insights into the precision of these estimations under asymptotic conditions. The ASE values offer a measure of expected accuracy, indicating how closely the estimated parameters are likely to approximate the true values as the sample size increases towards infinity. Additionally, Figures 1 and 2 demonstrate that the estimation errors are centered around 0 and exhibit a normal distribution.

$QMLE$	$\widehat{\alpha}_{1,1}$	$\widehat{\alpha}_{1,2}$	$\widehat{\gamma}_1$	$\widehat{\alpha}_{2,1}$	$\widehat{\alpha}_{2,2}$	$\widehat{\gamma}_2$
$n = 300$	-0.4958	0.3026	0.8996	0.2013	-0.5929	1.0023
$bias$	0.0041	0.0026	-0.0003	0.0013	0.0070	0.0023
sd	0.0323	0.0350	0.0144	0.0337	0.0382	0.0144
ASE	0.0001	0.0001	$5e-07$	0.0002	0.0002	$2e-06$
$n = 600$	-0.4969	0.3007	0.8994	0.2013	-0.6009	1.0004
$bias$	0.0030	0.0007	-0.0005	0.0013	-0.0009	0.0004
sd	0.0294	0.0287	0.0142	0.0325	0.0305	0.0147
ASE	$7e-05$	$7e-05$	$3e-07$	0.0001	0.0001	$e-06$

Table1: Estimation results for $PEXPAR_2(1)$

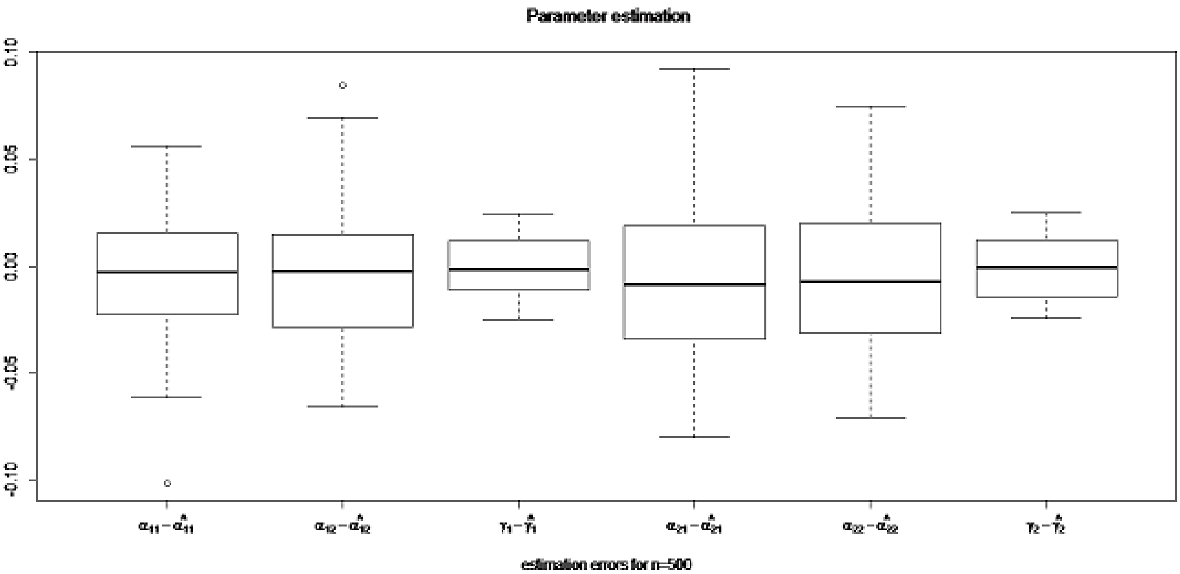

Figure 1. The boxplots of the errors of estimation for 100 replications and n=500.

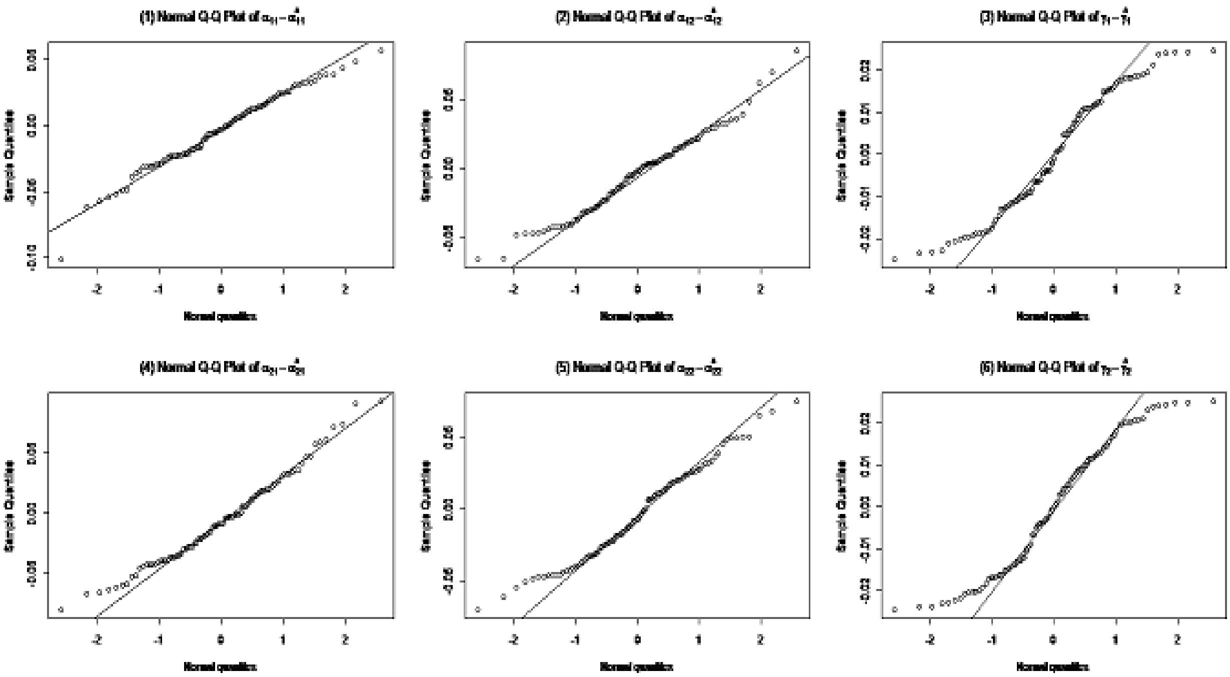

Figure 2. The Q-Q plots of the errors of estimation for 100 replications and n=500.

4. Application: Rainfall in Algeria

This section investigates the modeling of monthly rainfall data in Algeria spanning from January 1901 to December 2016. We employ both linear ($SARIMA$) and nonlinear ($PEXPAR_{12}(1)$) models to analyze

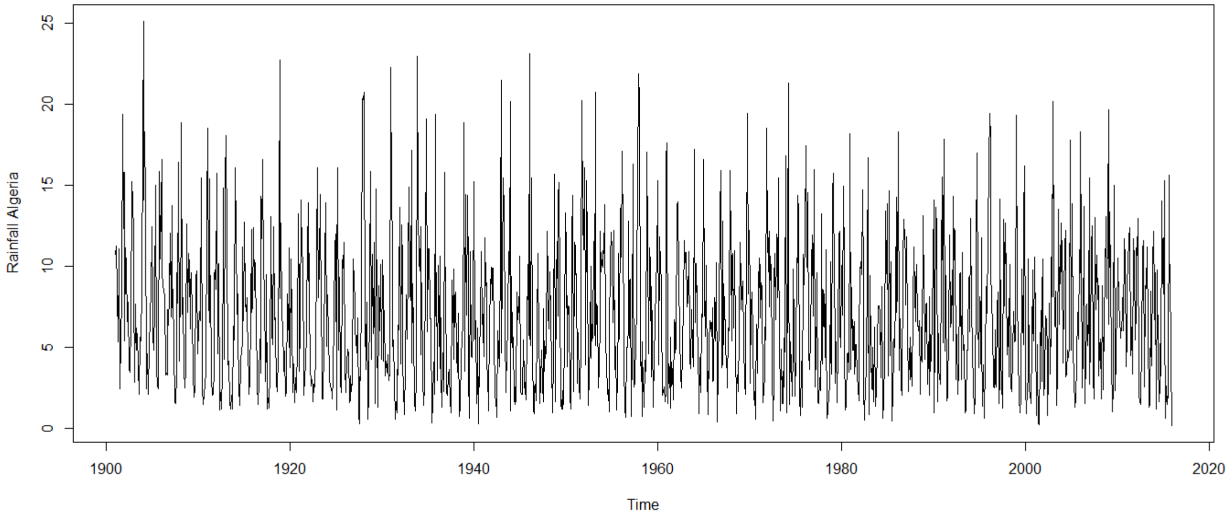

Figure 3. Monthly rainfall in Algeria

the dataset, comprising 1392 observations, sourced from the World Bank's Climate Knowledge Portal. We left the data for the year 2016 to out-of-sample forecasting validation.

SARIMA modelling

While rainfall exhibits seasonal and nonlinear characteristics, the visual inspection of the time series (Figure 3) does not conclusively confirm non-stationarity. However, further analysis using monthplot and boxplot (Figure 4) reveals significant seasonal behavior, characterized by noticeable fluctuations in the mean across seasons. Additionally, the monthplot suggests that the conditional variance may exhibit stochastic characteristics, as noted by referee feedback. The correlogram (Figure 5) confirms the presence of seasonality (multiples of 12), but without a clear decay, indicating non-stationarity. To address this, we apply a log transformation to stabilize the variance, as depicted in Figure 5. Subsequently, by examining the correlogram of the seasonal differences series $\Delta_{12} \log Y_t$, shown in Figure 6, we identify several SARIMA models and select $SARIMA\,(0,0,1)\,(0,1,1)_{12}$ using the [8] methodology,

$$\left(1 - L^{12}\right) \log Y_t = \left(1 - \underset{(0.0120)}{0.96}\, L^{12}\right) \left(1 + \underset{(0.0273)}{0.15}\, L\right) \varepsilon_t.$$

where the standard errors are provided in brackets. The resulting residuals are uncorrelated ($p - value = 0.7771$, Box-Ljung test) but not Gaussian ($p - value < 2.2e - 16$, Shapiro-Wilk normality test), indicating the potential for a nonlinear model. Forecasted values are reported in Table 3 and visually represented in Figure 8.

Periodic EXPAR modelling

The nonlinearity of the data is confirmed by the Keenan test on the log-transformed data $p - value = 7.36e - 11$. After centering the data by subtracting the seasonal means, parameter estimation results are presented in Table 2. The estimated residual variances are provided in Table 3, allowing for a comparison between $SARIMA$ and $PEXPAR$ models. Notably, for $SARIMA$: $\sigma^2_{SARIMA} = 0.3825$ and for $PEXPAR$:

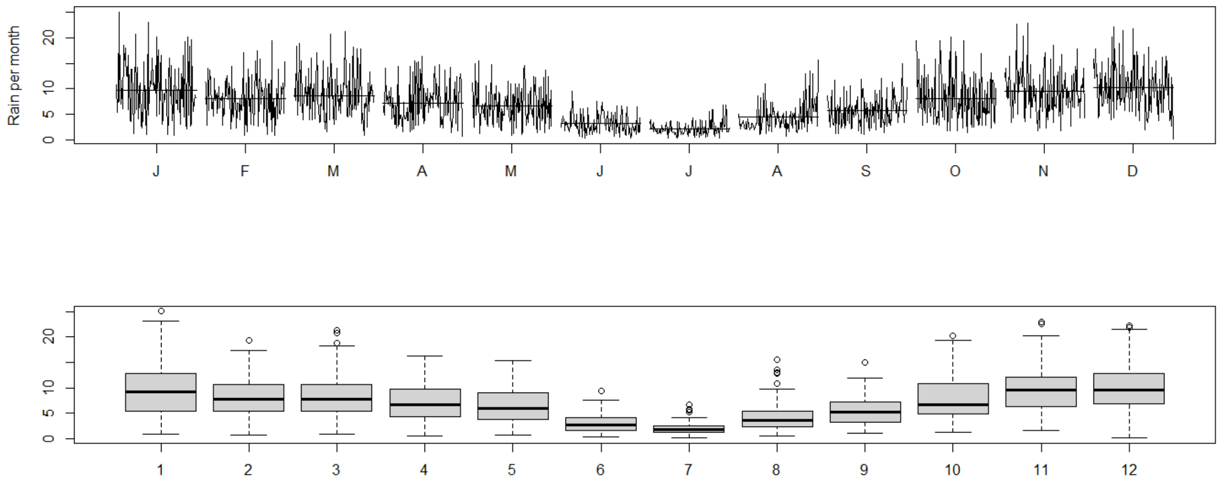

Figure 4. Monthplot and Boxplot of the rainfall

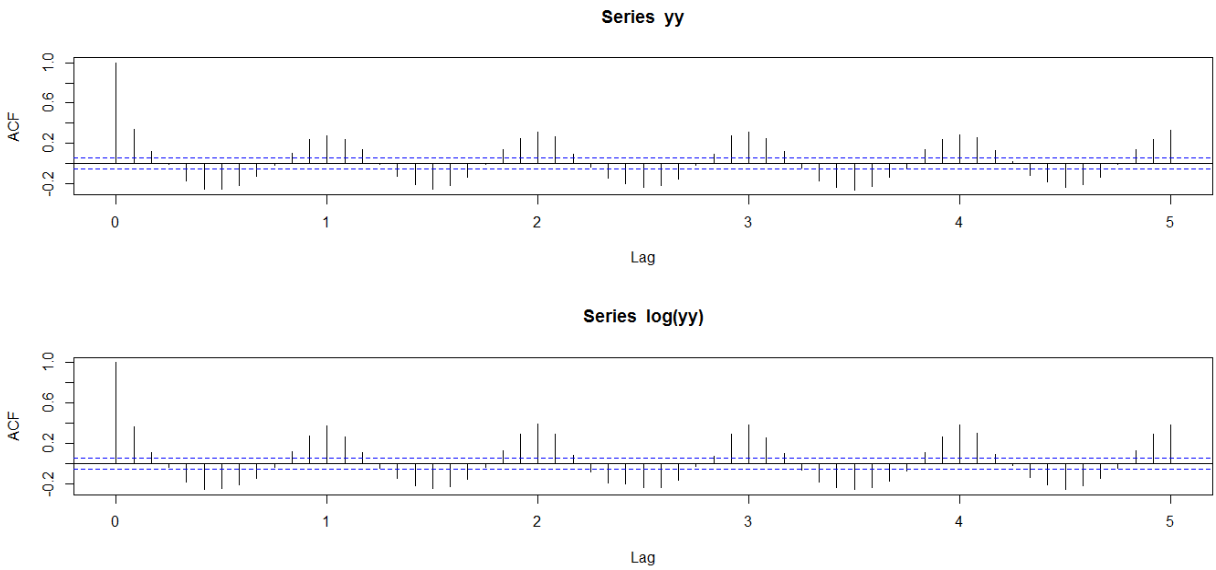

Figure 5. Correlogram of the rainfall

$\sigma^2_{PEXPAR} = 0.3608$, as the mean of values in Table 4. For the out-of-sample forecast, reported in Table 3, we obtain $\sigma^2_{SARIMA} = 3.9687$ and $\sigma^2_{PEXPAR} = 3.5076$. The $PEXPAR$ model demonstrates a marginally lower residual variance compared to $SARIMA$, further highlighting its potential efficacy in capturing the complexities of the rainfall data. In conclusion, our analysis suggests that the periodic $EXPAR$ model holds promise for capturing the seasonal variations and nonlinear behavior inherent in rainfall data, offering an alternative to traditional linear $SARIMA$ models.

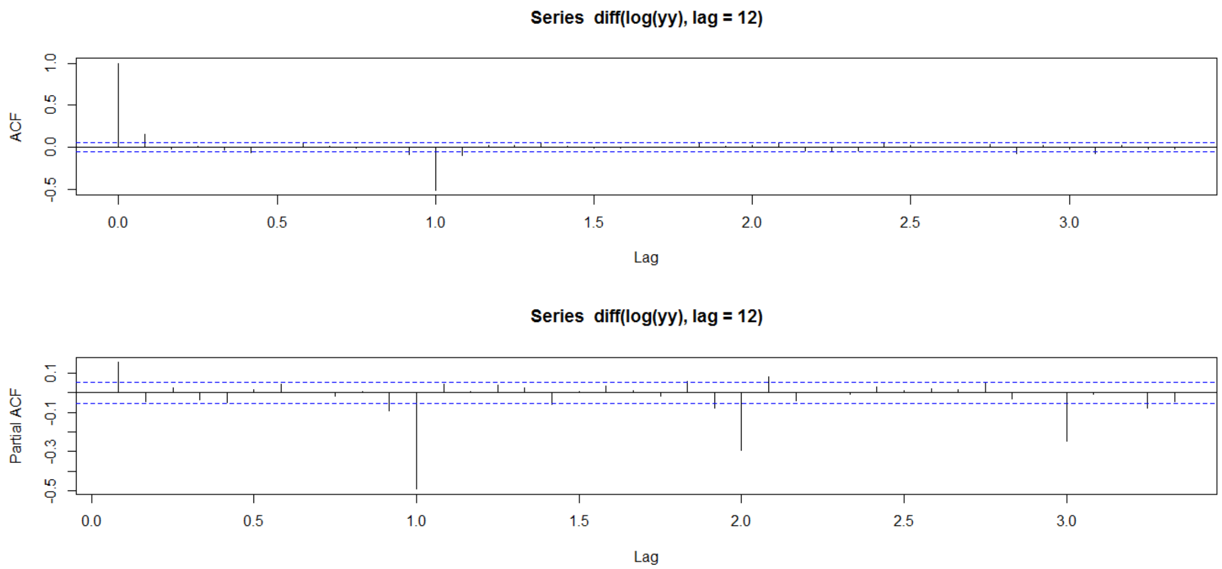

Figure 6. Correlogram of the seasonal differences rainfall

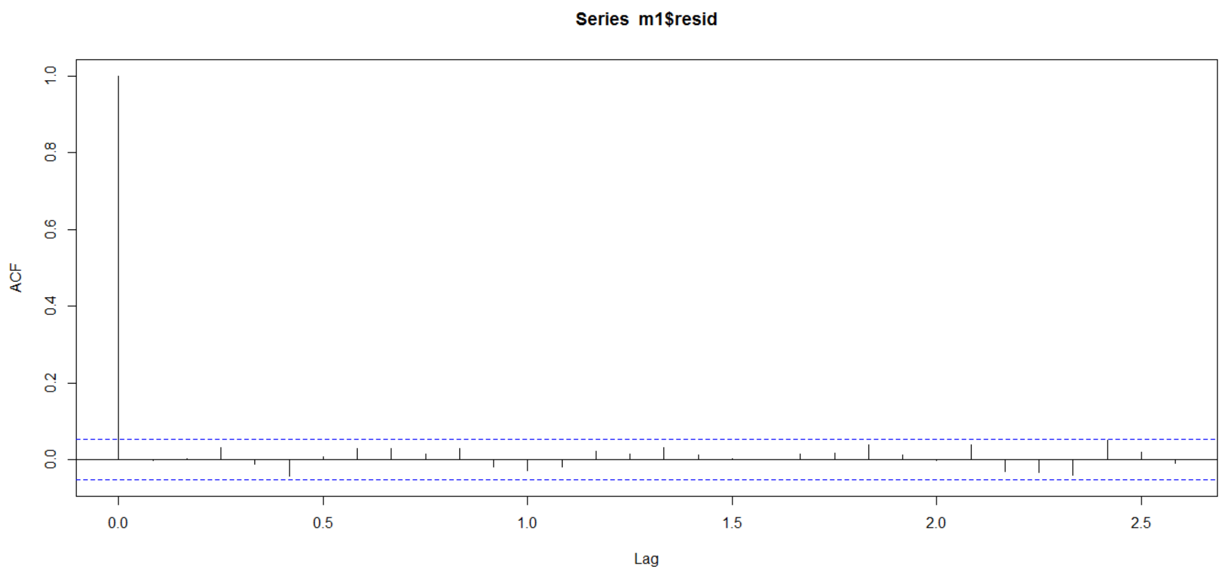

Figure 7. Correlogram of the residuals

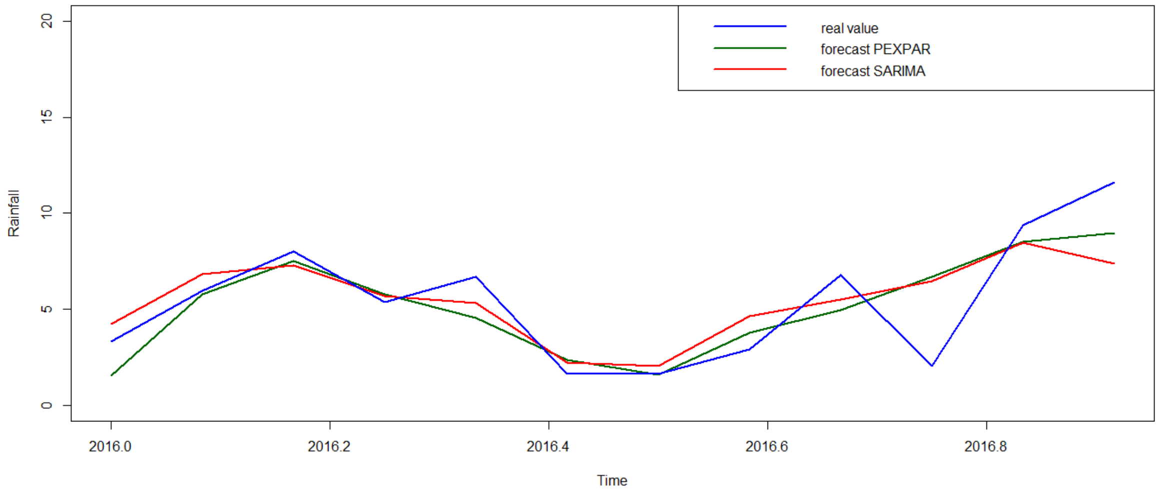

Figure 8. Comparison between real values and forecasted values

i	$\alpha_{i,1}$	$\alpha_{i,2}$	γ_i
1	0.4151	-0.7616	0.9105
2	0.1191	-0.6820	2.3728
3	0.3862	-0.3623	1.7407
4	0.2505	4.8749	65.2184
5	0.1713	4.5140	62.0735
6	-0.4483	0.7883	0.3903
7	0.1934	1.6171	21.6310
8	0.0235	-0.3598	7.3441
9	0.2336	0.4372	3.7746
10	0.3660	-1.1060	30.8242
11	0.3544	-0.5875	1.2350
12	0.1062	-0.8745	49.5967

Table 2: Parameter estimation of $PEXPAR_{12}\,(1)$

$Month$	$Real\ values$	$PEXPAR_{12}\,(1)$	$SARIMA$
Jan	3.3404	1.5434	4.2472
Feb	5.9697	5.7643	6.8530
Mar	8.0031	7.5117	7.2937
Apr	5.3921	5.8031	5.6884
May	6.6844	4.5539	5.3476
Jun	1.6623	2.3689	2.2397
Jul	1.6439	1.5962	2.0294
Aug	2.9051	3.7706	4.6669
Sep	6.7740	4.9689	5.5177
Oct	2.0406	6.6870	6.4840
Nov	9.3678	8.5109	8.4678
Dec	11.6125	8.9596	7.3785

Table 3: Forecasted values from $PEXPAR_{12}\,(1)$ and $SARIMA$.

Month	1	2	3	4	5	6
σ^2	0.4677	0.3620	0.2710	0.3680	0.3975	0.4743
Month	7	8	9	10	11	12
σ^2	0.3919	0.3506	0.2876	0.3363	0.2608	0.3621

Table 4: Monthly residual variances for $PEXPAR_{12}\,(1)$ model.

5. Conclusion

This study focused on modeling the rainfall series of Algeria using a $PEXPAR$ model, allowing for different coefficients in each season. The estimation of $PEXPAR_{12}(1)$ was addressed using the Gaussian quasi maximum likelihood method, and the consistency and asymptotic normality of the estimators were confirmed through simulated series. Comparing the forecasts generated by $SARIMA$ and $PEXPAR_{12}(1)$ models revealed a slight superiority of the latter over the former. These findings underscore the potential of $PEXPAR$ models in capturing the seasonal variations and nonlinear behavior of rainfall data, offering valuable insights for forecasting and decision making in hydro-climatic studies. To extend this research, future investigations could focus on developing a periodic $EXPAR - GARCH$ model. This expanded framework would incorporate both the periodic autoregressive structure of $PEXPAR$ models and the stochastic modeling of conditional variance inherent in $GARCH$ (Generalized Autoregressive Conditional Heteroskedasticity) models. By integrating these elements, the periodic $EXPAR - GARCH$ model could provide a more comprehensive approach for capturing the complex dynamics and variability of rainfall data, thereby enhancing predictive accuracy and decision support in hydro-climatic studies.

Acknowledgement

We would like to extend our heartfelt thanks to the referees, whose comments and suggestions worked like magic on our article, significantly transforming it. Their invaluable insights illuminated our work and provided a fresh perspective, enabling us to present more robust and compelling results.

REFERENCES

1. A. Aknouche, B. Almohaimeed, and S. Dimitrakopoulos, *Periodic autoregressive conditional duration*, Journal of Time Series Analysis, 43, 5–29, 2022.
2. A. Aknouche, and A. Bibi, *Quasi-maximum likelihood estimation of periodic GARCH and periodic ARMA-GARCH processes*, Journal of Time Series Analysis, 28, 19–46, 2009.
3. A. Aknouche, and H. Guerbyenne, *On some probabilistic properties of double periodic AR models*, Statistics and Probability Letters, 79, 407–413, 2009a.
4. A. Aknouche, and H. Guerbyenne, *Periodic stationarity of random coefficient periodic autoregressions*, Statistics and Probability Letters, 79, 990–996, 2009b.
5. N. Azouagh, and S. El Melhaoui, *An Exponential Autoregressive model for the forecasting of annual sunspots number*, Electronic Journal of Mathematical Analysis and Applications, 7, 3, 17–23, 2019.
6. A. Amondela, and C. Francq, *Concepts and tools for nonlinear time series modelling*, In Handbook of Computational Econometrics. Edts D. A. Belsley and E. J. Kontoghiorghes, Wiley, 2009.
7. S. Becila, and M. Merzougui, *Nonlinear Least Squares estimation of the Periodic EXPAR(1) model*, Communications in Statistics-Theory and Methods, 2020. DOI: 10.1080/03610926.2020.1839099.
8. G. E. P. Box,and G. M. Jenkins, *Time Series Analysis, Forecasting and Control*, 2nd edi., Holden-Day San Francisco, CA, 1976.
9. G. Y. Chen, M. Gan, and G. L. Chen, *Generalized exponential autoregressive models for nonlinear time series: Stationarity, estimation and applications*, Information Sciences 438, 46–57, 2018.
10. P. H. Franses, and M. Ooms, *A Periodic Long Memory Model for Quarterly UK Inflation*, International Journal of Forecasting, 13, 119–128, 1997.
11. E. G. Gladyshev, *Periodically correlated random sequences. Soviet*, Soviet. Math., 2, 385–88, 1961.
12. F. Hamdi, and S. Souam, *Mixture periodic GARCH models: Applications to exchange rate modeling*, 5th International Conference on Modeling, Simulation and Applied Optimization (ICMSAO).IEEE, 2013.

13. A. Lama, K. N. Singh, H. Singh et al, *Forecasting monthly rainfall of Sub-Himalayan region of India using parametric and non-parametric modelling approaches*, Model. Earth Syst. Environ, 2021. https://doi.org/10.1007/s40808-021-01124-5.
14. P. A. W. Lewis, and B. K. Ray, *Nonlinear modelling of periodic threshold autoregressions using TSMARS*, ournal of Time Series Analysis 23, 4, 459–471, 2002.
15. A. I. McLeod, *Diagnostic checking periodic autoregression models with application*, Journal of Time Series Analysis 15, 2, 221–233, 1994.
16. M. Merzougui, H. Dridi, and A. Chadli, *Test for periodicity in restrictive EXPAR models*, Communication in Statistics-Theory and Methods, 45:9, 2770–2783, 2016.
17. T. Ozaki, *Non-linear time series models for non-linear random vibrations*, Journal of Applied Probability, 17, 84–93, 1980.
18. H. Xu, F. Ding, and E. Yang, *Three-stage multi-innovation parameter estimation for an exponential autoregressive time-series model with moving average noise by using the data filtering technique*, International Journal of Robust and Nonlinear Control; 1–19, 2020.