IJEGEO

International Journal of Environment and Geoinformatics 11(3):019-023 (2024)

Research Article

Nakagami Distribution for Modeling Monthly Precipitations in Van, Türkiye

Kübra Bağcı Genel

[1] Econometrics, Faculty of Economics and Administrative Sciences, Van Yüzüncü Yıl University, Van, Türkiye

E-mail: kubrabagci@yyu.edu.tr

Received 13.02.2024
Accepted 31.08.2024

How to cite: Bağcı Genel, K.(2024). Nakagami Distribution for Modeling Monthly Precipitations in Van, Türkiye, *International Journal of Environment and Geoinformatics (IJEGEO)*, 11(3): 019-023. doi. 10.30897/ijegeo.1434719

Abstract

Precipitation patterns are intricately influenced by geographic factors and local environmental conditions. Statistical distributions are one of the methods that help investigate precipitation characteristics at different sites. Van is a province that ranks among the provinces with the lowest precipitation in the Eastern Anatolia region of Türkiye, receiving an annual rainfall of around 400 mm. In this study, 63 years of monthly average precipitation data from Van, are modeled employing various well-known statistical distributions including the Nakagami distribution. The Nakagami distribution is one of the flexible distributions used in describing data from various fields. In estimating the parameters of the considered distributions maximum likelihood estimation method is utilized. Comparisons are made using various goodness of fit criteria including root mean squared error, coefficient of determination, and Kolmogorov-Smirnov test. According to the results, the Nakagami distribution is found to be the most suitable statistical distribution for modeling precipitations in Van province. Additionally, precipitation values for 10, 25, 50, and 100-year return periods are obtained.

Keywords: Drought, Precipitation, Nakagami distribution, Van.

Introduction

Precipitation is a critical natural phenomenon that significantly impacts various systems, including agricultural practices, public water resources, and energy. In addition, the frequency of extreme weather events is increased with climate change (Tramblay et al., 2020; Spinoni et al., 2020). Droughts, flash floods, erosion, and other phenomena represent significant natural hazards that are linked to variations in precipitation patterns.

Given Türkiye's diverse climate types, the risk of drought may be elevated in particular areas. Van is one of the provinces with the least rainfall in the Eastern Anatolia region, with an annual rainfall of approximately 400 mm. While Van province experiences the highest precipitation during the spring season increasing flood risk, the summers in Van are typically characterized by dry conditions. The region predominantly sustains itself through agriculture and animal husbandry consequently the local livelihood is heavily influenced by the quantity of precipitations. In addition to the need for agricultural irrigation, the pearl mullet population, an endemic species living in the region, is affected by the drought in the Lake Van basin. Several studies have explored modeling precipitation data using distribution functions, addressing various purposes such as regional frequency analysis and forecasting of drought. For extreme value analysis, the precipitations are modeled using distributions such as generalized extreme value (GEV), generalized logistic, and generalized Pareto. For the case of drought analysis, there are various drought indices for the characterization and estimation of drought. Accurate drought estimation depends on selecting a suitable pdf for precipitation characteristics observed in certain regions. Since the choice of pdf can significantly influence the accuracy of drought indices, it is essential to highlight that while the (standard precipitation index) SPI based on the gamma distribution has been widely adopted, (Angelidis et al., 2012; Naresh Kumar et al., 2009; Sönmez et al., 2005) other distribution functions have also been employed for drought estimation (Angelidis et al., 2012; Moccia et al., 2022; Stagge et al., 2015) considering the precipitation characteristic of a specific area. When the related literature is examined, it is seen that the focus is generally on a limited set of distributions and many advantageous flexible distributions are overlooked. Kassam et al., (2021) provided a comprehensive review of the distributions used in modeling precipitations and applied 37 pdfs' for this purpose to data obtained from Northern Cyprus. According to their results, contrary to general expectations, the most commonly used distributions, such as beta, Log-Pearson, and exponential distributions were the least successful in modeling rainfall data. Meanwhile, the Burr, Wakeby, and Nakagami distributions provided the best fit to total rainfall data, based on goodness-of-fit tests.

Some of the other works considering modeling precipitations using alternative probability distributions are overviewed as follows. For the case of regional frequency analysis, Seckin et al., (2011) conducted a flood frequency analysis analyzing flood events in six regions of Türkiye and found that the 3-parameter log-normal and Pearson type III distributions are the most suitable models for the precipitation data. Similarly, Galoie et al., (2013) conducted a comparative analysis of four commonly used rainfall frequency distributions

(GEV, Gumbel, log-Pearson type 3, and lognormal) to identify the best-fit probability distribution for rainfall events. Alam et al., (2018) focus on frequency analysis to identify the most suitable probability distribution model for particularly maximum monthly rainfall in 35 locations in Bangladesh using 30 years of data. In addition, they calculated return periods for 10, 25, 50, and 100 years, providing valuable information for developing more accurate models of flooding risk and damage. Likewise, Tosunoglu & Gurbuz, (2019) conducted a study to identify the best-fitted probability distribution functions for annual rainfall in Türkiye and improve rainfall quantile estimates under various return periods. Ten widely used distributions are evaluated for data from 155 gauge stations (1975–2014) and they concluded that the gamma, log-logistic, two-parameter log-normal, normal, and Weibull distributions demonstrated the best fit for 78% of the data series, also estimated rainfall quantiles for different return periods (25, 50, and 100 years). Ghiaei et al., (2018) also utilized a regional frequency analysis in the Eastern Black Sea region in Türkiye and concluded the GEV distribution provided the best fit for precipitation data. Ozonur et al., (2021) analyzed monthly rainfall data of Brazil, using eight probability distributions including two- and three-parameter Weibull, Rayleigh, gamma, lognormal, and maximum Gumbel distributions. In addition, they estimated the parameters of the given distributions using maximum likelihood estimation (MLE) and showed that the three-parameter lognormal distribution generally exhibited superior fit compared to others, with the two-parameter lognormal distribution providing the least favorable fit among the considered probability distributions.

For the case of drought analysis, Angelidis et al., (2012) employed the normal and log-normal probability distributions, in addition to the commonly used gamma distribution, for calculating a popular drought index SPI. Their research found that particularly for SPI at 12 or 24 months, the log-normal or normal distributions can be suitable alternatives to the gamma distribution for monthly precipitation spanning 76 years. Similarly, Guenang et al., (2019) investigated the impact of various statistical distribution functions (gamma, Weibull, exponential, and lognormal) on the accuracy of the SPI at different time scales. Their research highlighted the Weibull distribution's superior performance considering monthly precipitation data from 1951 to 2016 obtained in Central Africa. Stagge et al., (2015) recommended modeling precipitations across Europe using the gamma distribution for drought index calculations. Moccia et al., (2021) analyze daily precipitation data from rain gauges in the contrasting climates of Lazio and Sicily, employing six probability distributions (Pareto Type II, Fréchet, Lognormal, Weibull, Gamma, and Gumbel). They found that contrary to common practice, heavy-tailed distributions (Fréchet and Pareto type II) outperform light-tailed distributions.

In this study, monthly average precipitation data are modeled using the gamma, lognormal, and Weibull distributions, which are common in this area as pointed out previously. Additionally, an alternative distribution, namely the Nakagami distribution, is employed with the aim of increasing modeling the accuracy of precipitations data of the region. To the best of the author's knowledge, only one previous study (Kassem et al., 2021) has utilized the Nakagami distribution for modeling precipitations as mentioned previously. This study contributes to the related literature by expanding its application in this field. In addition, a comprehensive dataset spanning 63 years of monthly precipitation is employed. The Nakagami distribution is a flexible distribution that is commonly used to model fading in wireless communication systems. In the case of the Nakagami distribution, the tail behavior can range from heavy-tailed to light-tailed, depending on the value of the shape parameter. When the shape parameter is small, the Nakagami distribution has heavy tails, meaning that the probability of observing extreme values is higher compared to distributions with lighter tails. Given the rarity of extreme precipitation events, the tail behavior of the Nakagami distribution may be advantageous in modeling such occurrences. This issue is also referenced in the literature review with an example (see Moccia et al., 2021). It should be noted that the Nakagami distribution is one of the most successful distributions out of 37 distributions in the study of Kassem et al. (2021). Motivated by these, the Nakagami distribution is employed in this study for modeling the precipitation of Van province. A monthly precipitation dataset has been employed in the analysis to provide valuable insights for drought analysis, as most well-known drought indices such as SPI and SPEI are generally conducted on monthly data and use pdfs in their calculations. In estimating parameters of the distributions MLE methodology is employed. As is known MLE methodology is an effective method in estimating parameters. Furthermore, precipitation values for 10, 25, 50, and 100-year return periods are provided. The rest of the study is planned as follows. In section 2, the modeling methodology is explained. Afterward, in section 3 the precipitation data are described and modeled using different distribution functions. The results are presented and discussed. In section 4, the study is finalized with some concluding remarks.

Materials and Methods

In this section, data is described and the pdfs for gamma, Weibull, lognormal, and Nakagami distributions and ML estimations of the parameters are provided.

Data Set

The data comprises the average monthly 63 years of precipitation from 1960 to 2023 of Van obtained from the General Directorate of Meteorology.

Table 1. Descriptive statistics for the precipitation data.

Descriptive statistics	
Mean	33.3197
Variance	750.8476
Skewness(standard error)	1.10274(0.088)
Kurtosis(standart error)	1.4809(0.176)

The data set can be obtained from the website of the General Directorate of Meteorology. Missing observations in data are handled by using linear interpolation based on the linear relationship between the known values. The geographical coordinates of Van, Türkiye, are latitude 38.503490 and longitude 43.396450. The descriptive statistics for the data are provided in Table 1.

Method

Gamma Distribution

A random variable following the Gamma distribution with the shape parameter α and the scale γ parameter has the following pdf

$$f(x;\alpha,\gamma)=\frac{1}{\gamma^{\alpha}\Gamma(\alpha)}x^{\alpha-1}e^{\frac{-x}{\gamma}},x>0;\gamma,\alpha>0. \quad \text{(Eq.1)}$$

The Lognormal Distribution

If a random variable X is log-normally distributed, then $Y = ln(X)$ normally distributed. Consequently, the pdf of the lognormal distribution is

$$f(x;\mu,\sigma)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left\{\frac{-(\log(x)-\mu)^2}{2\sigma^2}\right\},x>0 \quad \text{(Eq.2)}$$

if the natural logarithm of X is normally distributed with mean μ and variance σ^2 (see the lognormal distribution from Evans et al., (2001).

Weibull Distribution

When X is a random variable following the Weibull distribution with shape parameter β and the scale parameter θ the pdf of the Weibull distribution is given as,

$$f(x;\theta,\beta)=\frac{\beta}{\theta^{\beta}}x^{\beta-1}e^{-(\frac{x}{\theta})^{\beta}},x>0;\theta,\beta>0. \quad \text{(Eq.3)}$$

Nakagami Distribution

The Nakagami distribution has the following density function with the shape parameter λ and the scale ω parameter when X has the Nakagami distribution

$$f(x;\lambda,\omega)=2(\frac{\lambda}{\omega})^{\lambda}\frac{1}{\Gamma(\lambda)}x^{(2\lambda-1)}e^{\frac{-\lambda}{\omega}x^2},x>0;\lambda,\omega>0. \quad \text{(Eq.4)}$$

Parameter Estimation

In this work for obtaining the parameters of the considered pdf's the MLE method is utilized. The MLE is a parameter estimation method that finds the parameter values for a model by maximizing the likelihood function, representing the probability of observed data. MLE provides the most efficient estimates especially when the sample size is sufficiently large. The estimators of parameters of the considered distributions are not presented here since they were obtained previously in the literature. However, the ML estimations for the lognormal, Weibull, Gamma, and Nakagami distributions can be referenced from the previous studies of Evans et al., (2001), Cohen, (1965), Choi & Wette, (1969), and Schwartz et al., (2013), respectively.

Return Period

The return period is the average time between the occurrence of a specific event, such as a natural disaster or extreme weather event, of a certain magnitude or greater. If an event of magnitude x_T occurs once in T years, where x represents the precipitations, the probability (P) of the variable exceeding or equaling x in any given year is expressed as follows

$$P(x\geq x_T)=\frac{1}{T} \quad \text{(Eq.5)}$$

Results and Discussion

In this section, first, the data is modeled using gamma, lognormal, Weibull, and Nakagami distribution functions. Afterward, the performances of these pdfs for describing data are compared using various well-known criteria including, root mean squared error (RMSE), coefficient of determination (R^2), and Kolmogorov-Sminirnov test. Achieving lower values for RMSE and KS, and higher values for the KS (p-value) and R^2 criteria, indicates a better fit. The application is implemented using Matlab R2021 software. The estimated parameters of the distributions are listed in Table 2, arranged in the order specified within the parentheses. The first column corresponds to the first parameter, and the second column corresponds to the estimation of the second parameter in parentheses. The goodness of fit performances of the Gamma, lognormal, Weibull, and Nakagami distributions are provided in Table 3. Subsequently, the most suitable pdf is determined, and return periods for 10, 25, 50, and 100 years are calculated for the precipitation data.

Table 2. ML estimations of the parameters for the corresponding pdfs

Distribution	Parameters	
Gamma (α,γ)	1.09435	31.01217
Nakagami (λ,ω)	0.42299	1894.6069
Weibull (β,Θ)	35.2136	1.11853
Lognormal(μ,σ)	3.0025	1.31627

Table 3. Fitting performances of the considered distributions

Distribution	R^2	KS p*	KS	RMSE
Gamma	0.980	6.25e-04	0.072	0.040
Lognormal	0.910	4.27e-13	0.138	0.079
Weibull	0.987	0.019	0.055	0.032
Nakagami	0.992	0.247	0.036	0.017

**p-value*

According to Table 3, the Weibull and Nakagami distributions stand out in modeling precipitations of Van province. However, Nakagami distribution is more successful for all of the criteria. It can be said that the Nakagami distribution describes the precipitation data of Van province better for all of the criteria considered. For this reason, the Nakagami distribution is chosen for modeling. Furthermore, in Figure 1 the fitted densities for gamma, lognormal, Weibull, and Nakagami distributions are given. According to the figures the gamma, lognormal, and Weibull distributions overfitted data at the peaks consequently, the Nakagami distribution described data better. Furthermore, although it's not obvious from the figure, the Nakagami distribution effectively modeled rare precipitation amounts, with values reaching up to 180 mm at the tail end of the distribution resulting in better fitting performance. For pointing out this situation tail probabilities are calculated between the closest rivals

(Nakagami and Weibull distributions). The tail probability can be calculated using $P(X > x) = 1 - F(x)$ formulation. Here $F(x)$ is the cumulative distribution function. The tail probabilities for the precipitations higher than 150 mm are 0.0063581 and 0.011 for the Weibull and Nakagami distributions, respectively. This result proves that the Nakagami distribution's ability to model rare events at the tail of the distribution is higher.

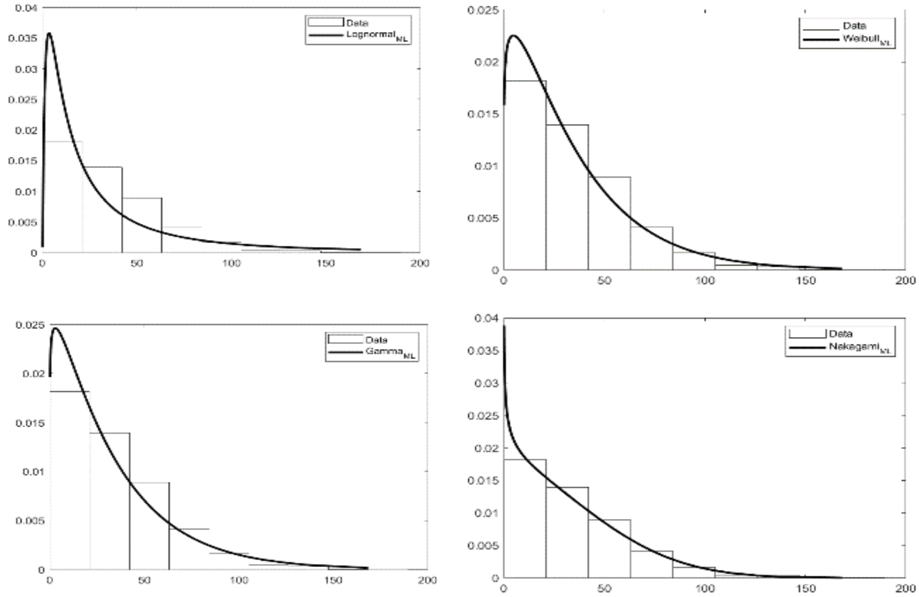

Fig. 1. Fitted density plots for Gamma, Weibull, lognormal, and Nakagami distributions

Return period information serves as a valuable guide for policymakers and planners, facilitating informed decision-making within specific timeframes. By incorporating the expected maximum monthly rainfall derived from the return period analysis, policymakers can make well-informed decisions more easily. In Figure 2, the precipitation's return levels of 1 to 100 years are illustrated.

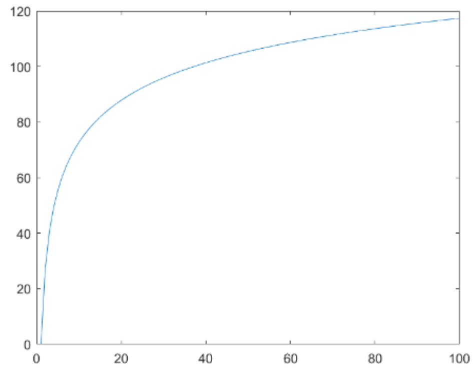

Fig. 2. Precipitation return level estimations

For Van station, the precipitation amounts corresponding to 10, 25, 50, and 100-year return periods are 72.7843, 92.3600, 105.4244, and 117.3812, respectively when modeled with the Nakagami distribution. For example, 72.7843 corresponds to the precipitation value that would be expected, on average, once every 10 years.

Given that previous studies have evaluated different sets of distributions, comparing these results directly is challenging. Additionally, variations in the data and study duration across previous works further complicate such comparisons. The studies employing the Nakagami distribution are very limited. Kassem et al. (2021) showed that Nakagami is one of the distributions that stand out in modeling precipitations out of the 37 distributions they considered. In this context, the Weibull distribution is the second-best model for precipitation for this study, and the literature also highlights the Weibull distribution as a significant choice in related studies. In a study conducted in Italy by Moccia et al., (2022) the lognormal distribution emerged as the best-fitting model for monthly precipitations, followed by the Weibull distribution. Hinis and Geyikli (2023) found that for the 3-month data, the Weibull and Pearson III distributions were the best fits, while the Weibull and Logistic distributions best fit the 12-month data. The findings of the study may be applicable to other regions with comparable characteristics with similar climatic and geographical conditions. However, unique local climatic conditions or unusual precipitation patterns might affect how well the model fits. Further studies in comparable areas would help to validate the applicability of this distribution.

Conclusion

Precipitation characteristics are shaped by the unique combination of a site's geography and surrounding environment. Each region exhibits its unique features, leading to the consideration of alternative distributions to accurately model its distinct patterns. In Van province, the economic structure is largely based on agriculture and animal husbandry, which is affected by the variability in the precipitation regime. In addition, the pearl mullet population, which is an endemic species to the Lake Van basin is affected by the variability in the precipitations as well. In this study, 63 years of monthly average precipitations in Van, Türkiye, are modeled using various statistical distributions. According to the results, the Nakagami distribution is selected as the best fit among all.

It is important to highlight that the Nakagami distribution is chosen in the analysis for its flexibility allowing for the representation of rare precipitation occurings. Furthermore, precipitation values for 10, 25, 50, and 100-year return periods have been determined with the aim of providing insights for policymakers and planners.

References

Alam, M. A., Emura, K., Farnham, C., Yuan, J. (2018). Best-Fit Probability Distributions and Return Periods for Maximum Monthly Rainfall in Bangladesh. *Climate*, *6*(1). doi.org/10.3390/cli6010009

Angelidis, P., Maris, F., Kotsovinos, N., Hrissanthou, V. (2012). Computation of Drought Index SPI with Alternative Distribution Functions. *Water Resources Management*, *26*(9), 2453–2473. doi.org/ 10.1007/s11269-012-0026-0

Choi, S. C., Wette, R. (1969). Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias. *Technometrics*, *11*(4), 683–690. doi.org/10.1080/00401706.1969.10490731

Cohen, A. C. (1965). Maximum Likelihood Estimation in the Weibull Distribution Based On Complete and On Censored Samples. *Technometrics*, *7*(4), 579–588. doi.org/10.1080/00401706.1965.10490300

Evans, M., Hastings, N., Peacock, B. (2001). *Statistical distributions*. IOP Publishing.

Galoie, M., Zenz, G., Eslamian, S. (2013). Application of L-moments for IDF determination in an Austrian basin. *International Journal of Hydrology Science and Technology*, *3*(1), 30–48. doi.org/10.1504/ IJHST.2013.055231

Ghiaei, F., Kankal, M., Anilan, T., Yuksek, O. (2018). Regional intensity–duration–frequency analysis in the Eastern Black Sea Basin, Türkiye, by using L-moments and regression analysis. *Theoretical and Applied Climatology*, *131*(1), 245–257. doi.org/ 10.1007/s00704-016-1953-0

Guenang, G. M., Komkoua, M. A. J., Pokam, M. W., Tanessong, R. S., Tchakoutio, S. A., Vondou, A., Tamoffo, A. T., Djiotang, L., Yepdo, Z., Mkankam, K. F. (2019). Sensitivity of SPI to Distribution Functions and Correlation Between its Values at Different Time Scales in Central Africa. *Earth Systems and Environment*, *3*(2), 203–214. doi.org/10.1007/s41748-019-00102-3

Hinis, M. A., Geyikli, M. S. (2023). Accuracy Evaluation of Standardized Precipitation Index (SPI) Estimation under Conventional Assumption in Yeşilırmak, Kızılırmak, and Konya Closed Basins, Turkey. *Advances in Meteorology*, 2023(1), 5142965. doi.org/10.1155/2023/5142965

Kassem, Y., Gökçekuş, H., Gökçekuş, R. (2021). Identification of the most suitable probability distribution models for monthly and annual rainfall series in Güzelyurt Region, Northern Cyprus. *Desalination and Water Treatment*, 215, 427–451. doi.org/10.5004/dwt.2021.26904

Moccia, B., Mineo, C., Ridolfi, E., Russo, F., Napolitano, F. (2021). Probability distributions of daily rainfall extremes in Lazio and Sicily, Italy, and design rainfall inferences. *Journal of Hydrology: Regional Studies*, *33*, 100771. doi.org/10.1016/j.ejrh.2020.100771

Moccia, B., Mineo, C., Ridolfi, E., Russo, F., Napolitano, F. (2022). SPI-Based Drought Classification in Italy: Influence of Different Probability Distribution Functions. *Water*, 14(22). doi.org/10.3390/ w14223668

Ozonur, D., Pobocikova, I., & de Souza, A. (2021). Statistical analysis of monthly rainfall in Central West Brazil using probability distributions. *Modeling Earth Systems and Environment*, *7*(3), 1979–1989. doi.org/10.1007/s40808-020-00954-z

Schwartz, J., Godwin, R. T., Giles, D. E. (2013). Improved maximum-likelihood estimation of the shape parameter in the Nakagami distribution. *Journal of Statistical Computation and Simulation*, *83*(3), 434–445. doi.org/10.1080/00949655.2011.615316

Seckin, N., Haktanir, T., Yurtal, R. (2011). Flood frequency analysis of Turkey using L-moments method. *Hydrological Processes*, *25*(22), 3499–3505. doi.org/10.1002/hyp.8077

Spinoni, J., Barbosa, P., Bucchignani, E. Et al., (2020) . uture global meteorological drought hot spots: a study based on CORDEX data, *J. Clim.*, 33, 3635-3661, 10.1175/JCLI-D-19-0084.1

Stagge, J. H., Tallaksen, L. M., Gudmundsson, L., Van Loon, A. F., Stahl, K. (2015). Candidate Distributions for Climatological Drought Indices (SPI and SPEI). *International Journal of Climatology*, *35*(13), 4027–4040. doi.org/10.1002/joc.4267

Tosunoglu, F., Gurbuz, F. (2019). Mapping spatial variability of annual rainfall under different return periods in Turkey: The application of various distribution functions and model selection techniques. *Meteorological Applications*, *26*(4), 671–681. doi.org/10.1002/met.1793

Tramblay Y, Koutroulis A, Samaniego L, et al. (2020). Challenges for drought assessment in the Mediterranean region under future climate scenarios. *Earth-Sci Rev*. 2020;210:103348.