Migration Letters

Volume: 21, No: 7 (2024), pp. 446-456

ISSN: 1741-8984 (Print) ISSN: 1741-8992 (Online)

www.migrationletters.com

Regional Frequency Analysis With L-Moments For The Determination Of Drought Event Maps In High Andean Basins In Peru

Melania Zapana Quispe[1*], Eduardo Chavarri-Velarde[2], Yénica Pachac Huerta[3], Robinson Peña Murillo[4], Yony Laqui Vilca[5], Waldo Lavado-Casimiro[6]

ABSTRACT

Droughts are extreme climatic phenomena that are difficult to quantify spatially and temporally, due to the effects of climate change impacting socio-economic development. In this context, the research aims to estimate and map the frequency of theoretical meteorological droughts at different return periods through the method of regional frequency analysis (ARF) based on L-moments (LM), using 97 virtual stations from the gridded PISCO product based on annual mean precipitation (PMA). Two homogeneous rainfall regions were identified using a combination of the Ward method and the LM approach. For ARF, the generalized normal distribution (GNO) was selected due to its best fit with the Z^{DIST} statistic, which allowed for the determination of the regional growth curve (quantiles). Finally, exponential predictive equations were obtained at a regional scale to relate LM and PMA, enabling the generation of meteorological drought maps.

***Keywords**: return period, PISCO, precipitation, probability, high Andean basins.*[1]

1. Introduction

Meteorological droughts result from precipitation deficits compared to what is considered "normal" (Núñez et al., 2011) and are a natural hazard (Wilhite, 2000). Drought episodes can cause significant social, economic, and environmental impacts, especially in arid and semi-arid regions. Historically, these events have affected large populations (35% of the world's population), often leading to fatalities (50% of mortality), while 7% of global economic losses have been attributed to their occurrence (Below et al., 2007). These losses are likely underestimated as indirect impacts are much more complex to quantify than direct consequences (Núñez et al., 2011), because they depend not only on the physical and temporal characteristics of the event (Wilhite et al., 2007). These impacts are more severe when communities are less prepared to face them, highlighting the importance of meteorological drought analysis to understand the probability of occurrence at different severity and duration levels (Acuña et al., 2011).

Precipitation is one of the most challenging meteorological variables to include in mathematical or statistical models, primarily due to its discontinuity and high randomness, making reliable prediction models essential for this variable (García-Marín et al., 2015).

1 Universidad Nacional Agraria La Molina; ORCID, 0000-0002-3412-4285;
2 Universidad Nacional Agraria La Molina; ORCID, 0000-0002-8445-8996;
3 Universidad Nacional Agraria La Molina, Universidad Nacional Santiago Antunez de Mayolo; ORCID, 0000-0002-1577-0548;
4 Universidad Nacional Agraria La Molina; ORCID, 0000-0001-6196-4039;
5 Universidad Nacional Mayor de San Marcos; ORCID, 0000-0002-8567-4543;
6 Universidad Nacional Agraria La Molina, Servicio Nacional de Meteorología e Hidrología del Perú; ORCID, 0000-0002-0051-0743;
*Corresponding autor: 20201343@lamolina.edu.pe

Regional frequency analysis (RFA) is a widely used method for analyzing extreme precipitation changes, which can be influenced in its results by identifying homogeneous regions and selecting regional distributions (Hosking y Wallis, 1997). Data from sites within a homogeneous region can be aggregated to improve accuracy in estimating the probability-quantile relationship at all sites (Wallis et al., 2007).

There are different methods to fit probability distributions to precipitation record samples, among which stand out the method of moments, the maximum likelihood method, the probability-weighted moments method, and the L-moments method (Rivano, 2004). L-moments are characterized by providing more accurate calculations than other methods in small sample sizes, with outliers in the data highly influencing the value of the standard deviation, making L-moments less sensitive to such values (Hosking y Wallis, 1997).

There are different methods to fit probability distributions to samples of precipitation records, among which the method of moments, maximum likelihood method, probability-weighted moments method, and L-moments method stand out (Rivano, 2004). L-moments are known for providing more precise calculations than other methods in small sample sizes, with outliers in the data that highly influence the value of the standard deviation, and L-moments are less sensitive to such values (Hosking y Wallis, 1997).

In recent years, the methodology of L-moment-based Regional Frequency Analysis (ARF-LM) has been validated in different parts of the worl, such as the United States, Mexico, Turkey, Germany, and New Zealand, as well as various European countries (Núñez et al., 2011) Canada (Alila, 1999), Korea (Lee y Maeng, 2003), and Morocco (Moujahid et al., 2018). Regarding Peru, the first initiative (Acuña et al., 2011) was applied to eleven hydrographic units (Olmos, Motupe, La Leche, Chancay Lambayeque, Zaña, Chaman, Jequetepeque, Chicama, Moche, Virú, and Huamansaña), located in the Pacific hydrographic region, for droughts at 60% of the PMA.

However, an analysis of meteorological droughts is limited by the absence of a suitable source of precipitation records for probability estimation. Therefore, monthly gridded data from the PISCO v2.0 product (Peruvian Interpolation of the SENAMHI's Climatological and Hydrological Stations) was used. The research objective was to determine maps of meteorological drought events in the high Andean basins of Peru using ARF-LM.

2. Materials and methods

2.1. Description of the study area

The study area is located in southern Peru, in the departments of Apurímac and Cusco, encompassing the high Andean basins of Peru, which belong to the Atlantic watershed (Figure 1). It extends between longitudes 71° 00´ - 73° 30´ and latitudes 11º 54´ - 15º 36´. It has an area of 61,533.0 km^2, an average elevation of 3,435.0 meters above sea level, a rainy-semi-cold climate with rainy summers averaging 700 mm annually, and dry winters with moderate frosts and average annual temperatures of 7 °C.

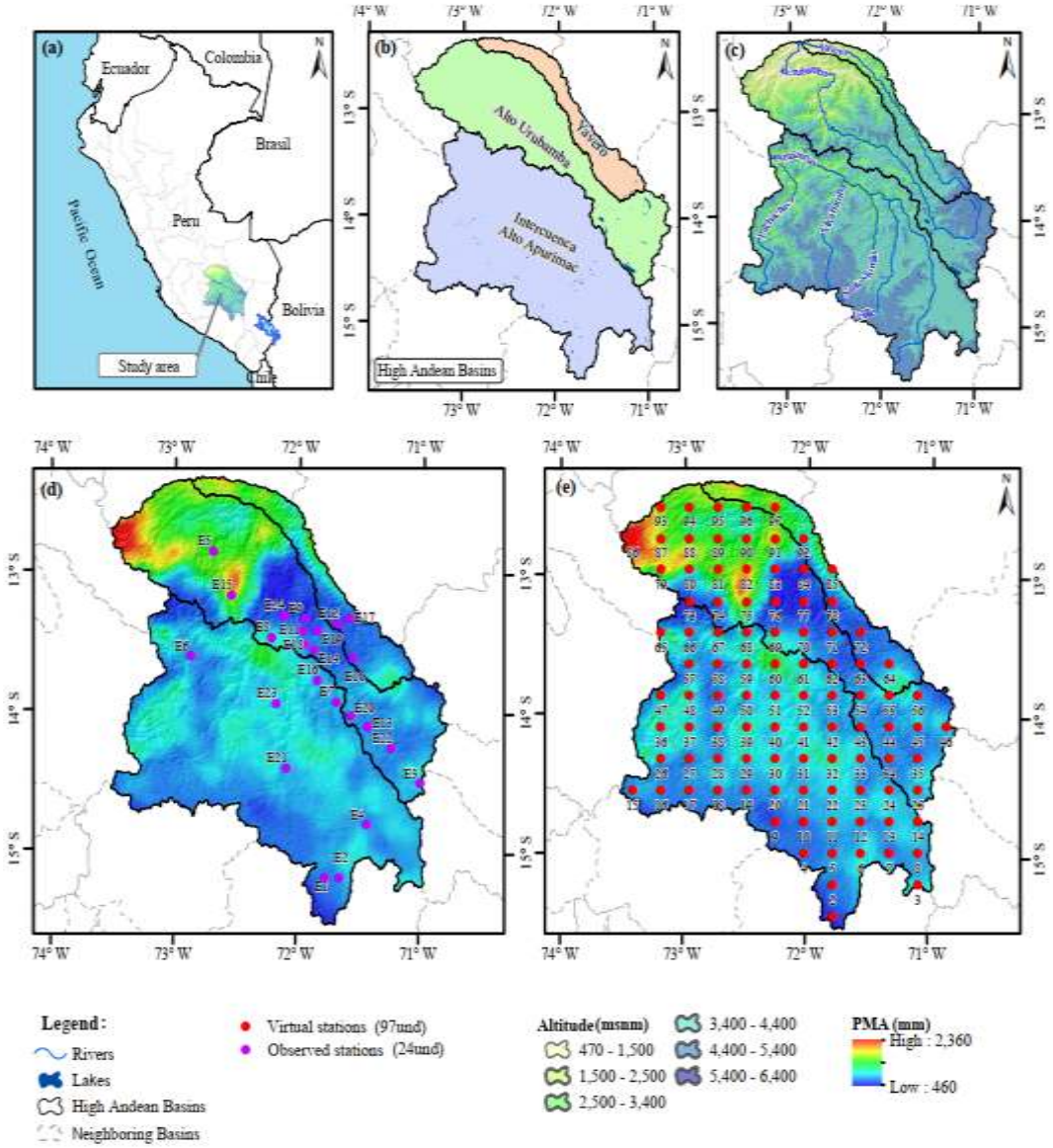

Figure 1. (a) Location of the study area in Peru; (b) Andean highland basins under study, (c) Altitude of the study area; (d) Observed stations (SENAMHI); (e) Virtual stations of PISCO product to measure the annual mean precipitation (PMA).

2.2. Data source

Ninety-seven virtual monthly precipitation stations were considered in the study area, based on the gridded product PISCO, developed by the National Meteorology and Hydrology Service of Peru (SENAMHI). PISCO combines ground station data with climatologist, reanalysis, and satellite products for rainfall estimation, to obtain nationwide gridded data at a resolution of 0.05° (~5 km). The analysis covers a record of 36 years from the period 1981–2016 (Aybar et al., 2020) and is available at:

https://iridl.ldeo.columbia.edu/SOURCES/.SENAMHI/.HSR/?Set-Language=es.

2.3. Methodology

2.3.1. Data preparation

Quality control checks were conducted based on statistical tests of precipitation data from the virtual stations in the study area. In the initial stage of the study, an exploratory data analysis (AED) was performed to detect outliers using box plots (Hubert y Vandervieren, 2008). Subsequently, the assumptions were verified (hypothesis testing of virtual stations to analyze the stationarity of the series) for the ARF-LM, considering the non-parametric Mann-Kendall test (Yue et al., 2002).

2.3.2. Homogeneous zone identification

To identify homogeneous regions using ARF-LM, a preliminary grouping of homogeneous regions was conducted using the Ward method (Miyamoto et al., 2015), aiming to determine regions of stations that synthesize the regional rainfall behavior of each zone. Subsequently, the final grouping of stations was carried out using LM through station filtering employing the discordance measure (Di) and the heterogeneity measure (Hj; j=1, 2, 3, n), where it is considered acceptably homogeneous if Hj<1; possibly heterogeneous if 1<Hj<2; and definitely heterogeneous when Hj>2 (Hosking y Wallis, 1997).

2.3.3. Determination of probability distribution function

Five distributions with 3 parameters were evaluated: Generalized Logistic (GLO), General Extreme Value (GEV), Generalized Pareto (GPA), Log normal (LN3), Generalized Normal (GNO), and Pearson Type III (PE3). From these distributions, the one with the best fit was selected based on regional LM ratios diagrams, representing L-skewness (τ3) and L-kurtosis (τ4) on a plane. Subsequently, the Z-statistic ($\left|Z^{DIST}\right| \leq 1.64$) was used at a 90% confidence level (Hosking y Wallis, 1997).

2.3.4. Determination of regional and local quantiles

Quantiles are defined as the amount of precipitation associated with a specific probability function (Maeda et al., 2013). Station-level quantiles were calculated by multiplying the regional quantiles by the scale factor associated with each station (Eq. 1), as per Hosking & Wallis, (1997):

$$\widehat{Q}_i(F) = \hat{u}_i * \hat{q}(F) \qquad (1)$$

$\widehat{Q}_i(F)$: is the quantile function for station; $\hat{u}_i$: is the mean value for station I; $\hat{q}(F)$: is the regional growth curve. Uncertainty was estimated using the root mean square error (RMSE) precision measure and 95% error bounds (Núñez et al., 2011).

2.3.5. Mapping the return period of drought events

In this stage, a raster image of the PMA based on the gridded PISCO product data was used to calculate the spatial variability of LM in the study area. Using the PMA raster and the prediction function (Eq. 2), the coefficient of variation (L-Cv, τ), L-skewness (Sk, τ3), and L-kurtosis (Ku, τ4) values were generated for each cell of the PMA raster.

The selected function to describe the relationship between the LM and PMA is as follows:

$$\text{L-Momento ratio} = \propto e^{-\beta(PMA)} + \delta \qquad (2)$$

α is the scale factor, β is the decay factor, δ is the calculated LM limit value, and the LM ratios corresponding to the linear moments (L-Cv, Sk, and Ku) according to Wallis et al., (2007).

The maps of drought return periods are generated from the distribution parameters for each raster cell (L-Cv, Sk, and Ku), the non-exceedance probability (F), and using the cumulative distribution function for the selected regional probability distribution (Hosking y Wallis, 1997). Drought threshold maps to represent precipitation deficits at 20%, 40%, and 60% were generated from the PMA at 80%, 60%, and 40% of return periods.

3. Results

Based on the gridded PISCO data, no qualitative outliers were detected for station precipitation. It was observed that precipitation values were within the confidence limits (upper and lower). Furthermore, it was determined that out of the 97 stations, 9% exhibited significant trends, which were not considered, and 91% of the stations were accepted for analysis.

Two homogeneous regions were considered using a combination of the Ward method and the LM method. The measure of discordance for homogeneous region 1 indicates that the regional mean annual precipitation (PMAR) was 759.14 mm, with a range of discordance from 0.08 - 2.56, and in homogeneous region 2, the PMAR was 1,252.51 mm, with a range of discordance from 0.262 - 2.00. The characteristics of the LM (Table 1) in each homogeneous region indicated the dispersion of the LM ratios of each station with respect to the regional LM ratios (Figure 2).

Table 1. Main regional L-moment characteristics of homogeneous regions.

	N° E	l_1	τ	τ_3	τ_4	H_1	H_2	H_3
Region 1	56	1	0.10	0.04	0.13	0.94	-2.31	-3.32
Region 2	14	1	0.06	-0.03	0.12	0.21	-1.80	-2.69

N° E: number of stations, H: heterogeneity. l_1: sample mean, τ: coefficient of variation, τ_3: L-skewness coefficient, τ_4: L-kurtosis coefficient.

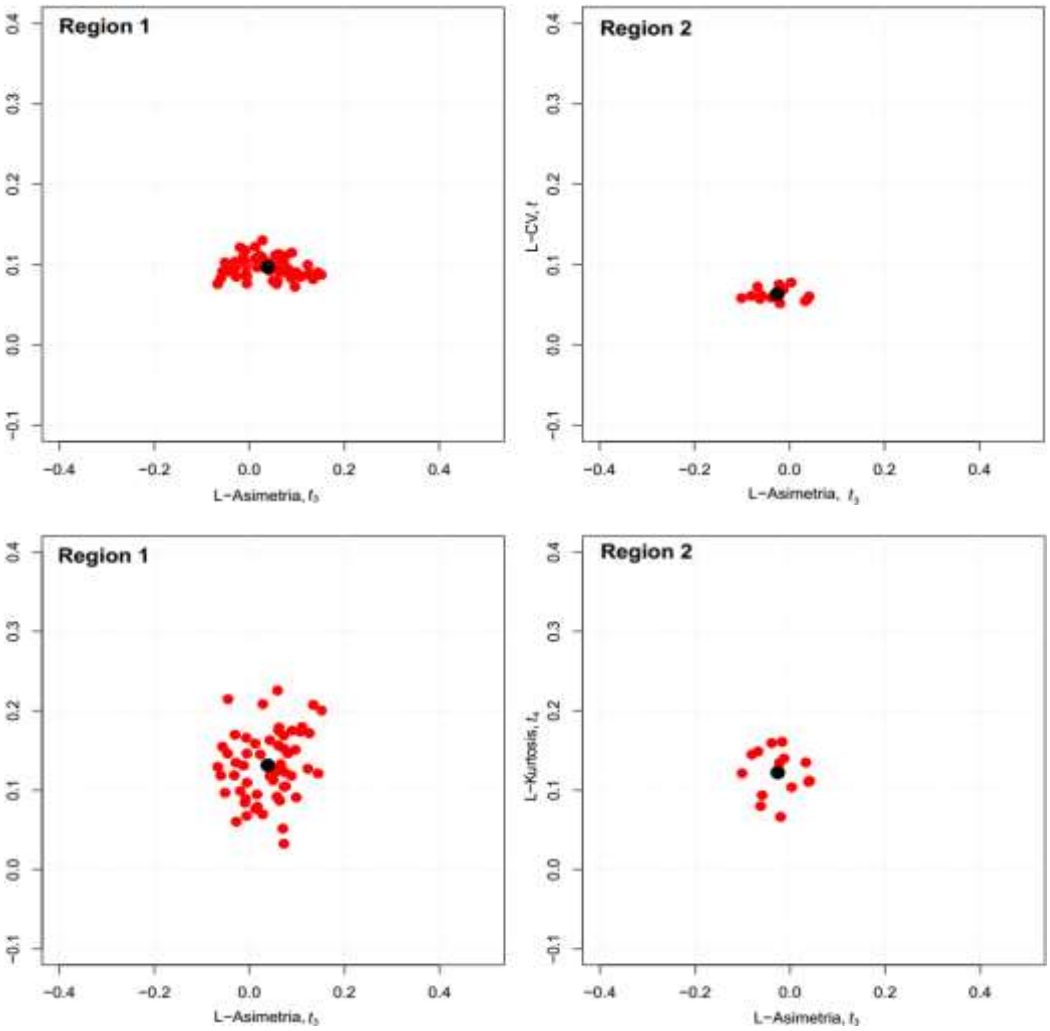

Figure 2. L-moment ratios of stations for each homogeneous region. Mean of regional L-moment ratios (point in bold). Mean of each station (point in red).

The measure of heterogeneity for the 2 homogeneous regions according to the statistics (Table 1), statistic H1 was considered the most stringent for evaluating homogeneity; the regions showed values below 1, considered homogeneous (Figure 3). The formation of homogeneous regions considered 70 stations for analysis.

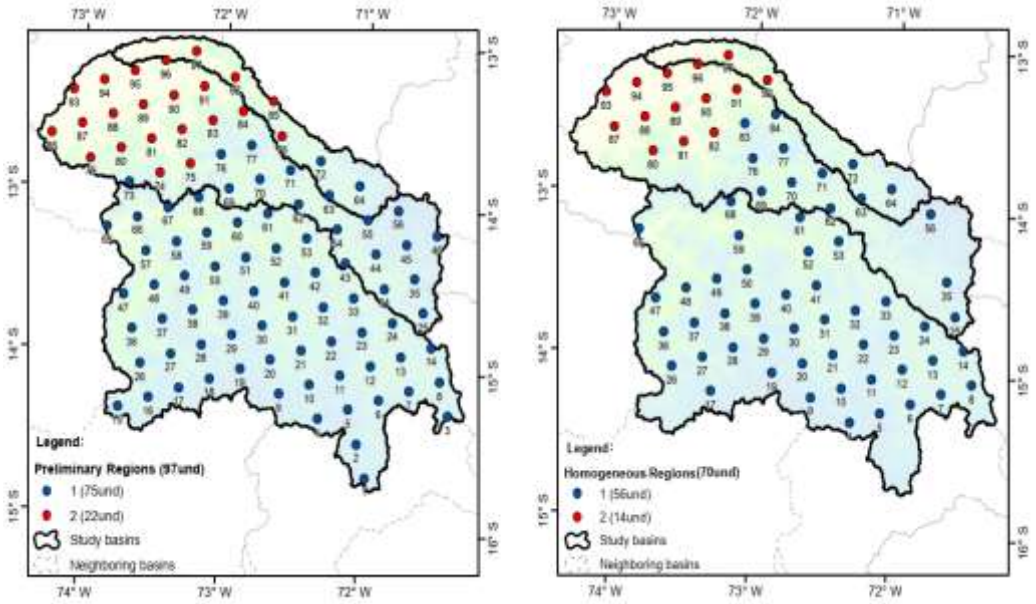

Figure 3. Location of stations for each homogeneous region (RH). Blue points RH 1 and red points RH 2.

According to the diagram of LM ratios per station with the mean of regional LM ratios for homogeneous regions 1 and 2 with respect to the 5 theoretical distributions, the best fit is presented for PE3 and GNO, respectively (Figure 4). Likewise, using the Z^{DIST} statistic in homogeneous regions 1 and 2 showed a better fit to PE3 and GNO, respectively (Table 2).

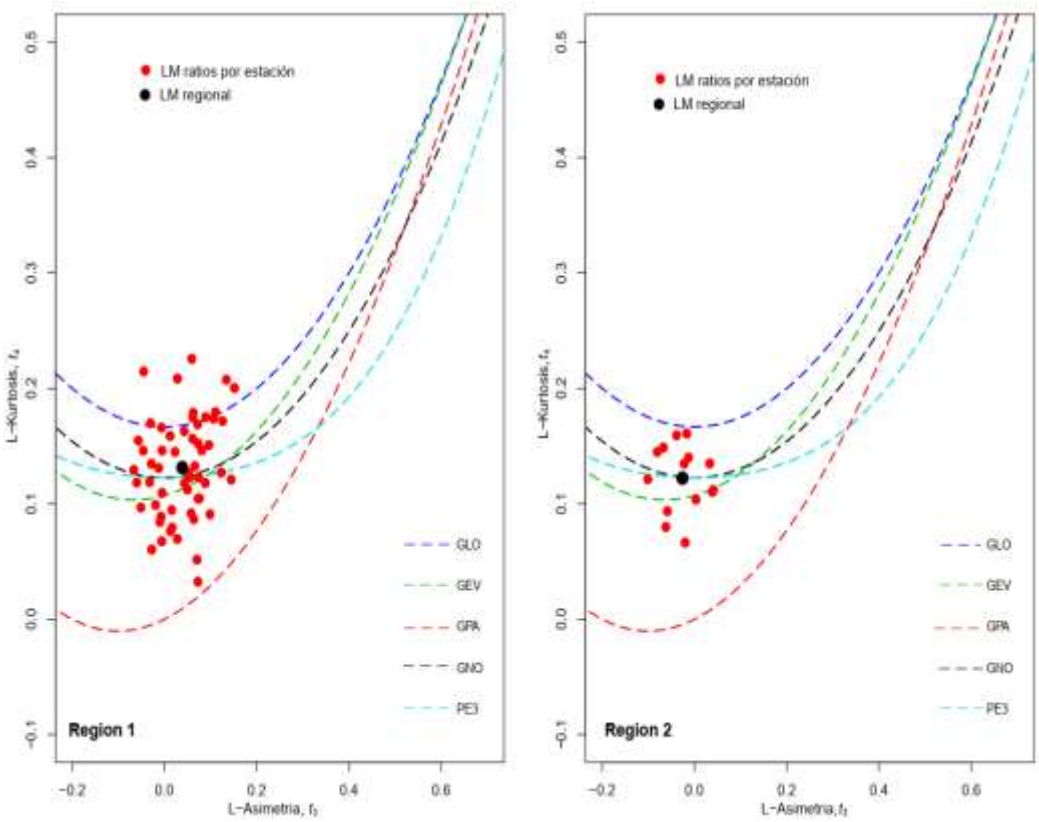

Figure 4. Diagram of L-moment (LM) ratios per station with LM regional ratios vs candidate distributions for homogeneous regions. Generalised Logistic (GLO), Generalised Extreme Values (GEV), Generalised Pareto (GPA), Log-Normal (LN3), Generalised Normal (GNO), and Pearson Type III (PE3).

Table 2. Z^{DIST} statistic for various distributions

Statistic	Region 1	Region 2
Z^{GLO}	4.99	3.03
Z^{GEV}	-2.44	-1.09
Z^{GNO}	-0.96	0.12
Z^{PE3}	-1.06	0.09
Z^{GPA}	-16.42	-8.32

Z^{DIST}: Goodness-of-fit statistic.

Subsequently, the quantiles of the regional growth curve for frequencies of 0.01, 0.02, 0.04, 0.05, 0.07, 0.1, 0.2, and 0.5 determined that the most robust distribution is the GNO for homogeneous regions 1 and 2 for return periods of 100, 50, 25, 20, 15, 10, 5, and 2 years (Table 3).

Table 3. Parameters and regional quantiles for the best-fitting distribution

		Regional quantiles for non-exceedance probabilities								
	Distribution	**F =**	**0.01**	**0.02**	**0.04**	**0.05**	**0.07**	**0.1**	**0.2**	**0.5**
		TR =	**100**	**50**	**25**	**20**	**15**	**10**	**5**	**2**
Region 1	**GNO**	Quantile	0.63	0.67	0.71	0.73	0.75	0.79	0.85	0.99
		RMSE	0.08	0.07	0.06	0.05	0.05	0.04	0.03	0.01
		LEI	0.52	0.57	0.63	0.65	0.68	0.72	0.81	0.98
		LES	0.76	0.78	0.81	0.82	0.83	0.85	0.90	1.00
Region 2	**GNO**	Quantile	0.73	0.76	0.80	0.81	0.83	0.85	0.91	1.00
		RMSE	0.05	0.04	0.03	0.03	0.03	0.02	0.01	0.01
		LEI	0.65	0.70	0.75	0.76	0.78	0.82	0.88	1.00
		LES	0.81	0.83	0.85	0.86	0.87	0.89	0.93	1.01

F: Non-exceedance probability, TR: Return period, RMSE: Root mean square error, LEI: Lower error limit, LES: Upper error limit.

For the mapping of the TR of drought events, curves were constructed based on the regional curves associating LM ratios and PMA adjusted to an exponential function where LM decreases as precipitation increases (Eq. 3, Eq. 4, and Eq. 5). The generated exponential functions were used to calculate the spatial distribution of the parameters of the GNO distribution, and through map algebra, the spatial distribution of LM in the study area was obtained, where high L-Cv regions are areas more likely to experience drought events compared to those with low L-Cv values. Finally, the maps were evaluated at different TRs (Figure 5), for deficits of 20%, 30%, and 40% of the PMA, drought occurs more prominently at return periods of 18 to 25 years, 16 to 50 years, and 76 to 100 years respectively. Although the latter also prevailed in the 30% deficit.

$$L\text{-}Cv = 0.24e^{-0.0020(PMA)} + 0.4 \quad (3)$$

$$L\text{-}Sk = 0.30e^{-0.0032(PMA)} + 0.4 \quad (4)$$

$$L\text{-}Ku = 0.24e^{-0.0041(PMA)} + 0.4 \quad (5)$$

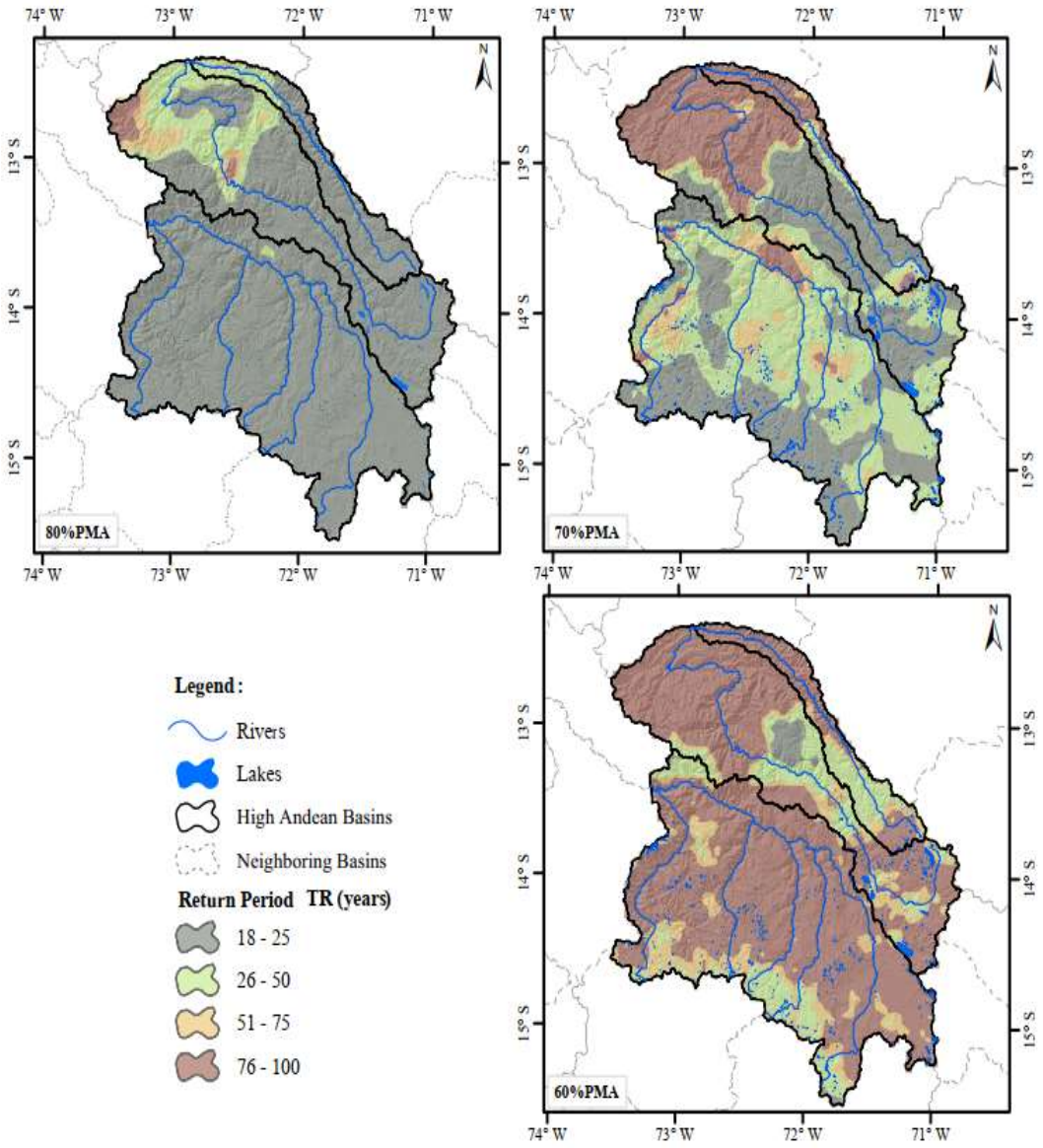

Figure 5. Drought maps at different equivalent return periods: (a) 80% of the mean annual precipitation (PMA) with a 20% precipitation deficit; (b) 70% of PMA with a 30% precipitation deficit; (c) 60% of PMA with a 40% precipitation deficit.

4. Discussion

4.1. Regional frequency analysis based on L-moment

The assessment of data quality and stationarity (Mann-Kendall) was conducted for each station (local level), which provides greater robustness compared to regional-level evaluation (Núñez et al., 2011). The maximum discordance for homogeneous region 1 and homogeneous region 2 was lower than the maximum allowed (3.00; Hosking & Wallis, 1997) for regions with more than 15 stations. Regarding the H1 statistic for homogeneous region 1 and homogeneous region 2, these were lower than the maximum allowed (1.00; Hosking & Wallis, 1997).

The distribution that bests fit for homogeneous regions 1 and 2 is the GNO. This distribution also appeared in the Lake Titicaca watershed, with the same climatic behavior as homogeneous region 1 (Fernández & Lavado, 2017).

The quantiles of the regional growth curve for frequencies 0.01, 0.02, 0.04, 0.05, 0.07, 0.1, 0.2, 0.5, were higher in the driest region than in the wettest region, indicating more frequent extremes. Numerous studies have demonstrated this trend (Wallis et al., 2007, (Núñez et al., 2011), (Kaluba et al., 2017)

The RMSE for homogeneous regions 1 and 2 for a return period of 100 years falls within the permitted range (0.05 – 0.160; Hu et al., 2019), indicating that the quantile estimates of the PMA are reliable and can be used in the study area.

The mapping of meteorological drought at 40% deficit of PMA occurred in the southwest of the study area with a TR of 18 - 26 years, a result also observed in central-northern Chile (Núñez et al., 2011), a location with similar climatic behavior.

5. Conclusions

In this study, the ARF-LM methodology was developed to estimate the spatial distribution of drought frequency in the study area, located in a transition zone between semi-arid and humid areas, relying solely on the use of gridded data from the PISCO product, valid for sites with and without rainfall records.

The general objective was to characterize meteorological drought using ARF-LM. The statistics demonstrated high reliability, enabling the identification of homogeneous regions adjusted by various probability distributions. The GNO distribution was considered the most acceptable, facilitating the determination of the regional growth curve for the study area. Finally, exponential predictive equations at a regional scale were obtained to relate LM and PMA, allowing for the generation of meteorological drought maps at different return periods.

6. Acknowledgments

Special thanks to the National Meteorology and Hydrology Service of Peru (SENAMHI) for providing the information that made this research possible.

Conflict of Interest: The authors declare no conflicts of interest

7. References

Acuña, J., Felipe, O., Ordoñez, J., y Arboleda, F. (2011). Análisis regional de frecuencia de precipitación anual para la determinación de mapas de sequías. 12.

Alila, Y. (1999). A hierarchical approach for the regionalization of precipitation annual maxima in Canada. Journal of Geophysical Research: Atmospheres, 104(D24), 31645-31655. https://doi.org/10.1029/1999JD900764

Aybar, C., Fernández, C., Huerta, A., Lavado, W., Vega, F., y Felipe-Obando, O. (2020). Construction of a high-resolution gridded rainfall dataset for Peru from 1981 to the present day. Hydrological Sciences Journal, 65(5), 770-785. https://doi.org/10.1080/02626667.2019.1649411

Below, R., Grover-Kopec, E., y Dilley, M. (2007). Documenting Drought-Related Disasters: A Global Reassessment. The Journal of Environment & Development, 16(3), 328-344. https://doi.org/10.1177/1070496507306222

Fernández, C. A., y Lavado, W. S. (2017). Regional maximum rainfall analysis using L-moments at the Titicaca Lake drainage, Peru. Theoretical and Applied Climatology, 129(3-4), 1295-1307. https://doi.org/10.1007/s00704-016-1845-3

García-Marín, A. P., Estévez, J., Medina-Cobo, M. T., y Ayuso-Muñoz, J. L. (2015). Delimiting homogeneous regions using the multifractal properties of validated rainfall data series. Journal of Hydrology, 529, 106-119. https://doi.org/10.1016/j.jhydrol.2015.07.021

Hosking, J. R. M., y Wallis, J. R. (1997). Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge University Press, New York, United States of America.

Hu, C., Xia, J., She, D., Xu, C., Zhang, L., Song, Z., y Zhao, L. (2019). A modified regional L-moment method for regional extreme precipitation frequency analysis in the Songliao River Basin of China. Atmospheric Research, 230, 104629. https://doi.org/10.1016/j.atmosres.2019.104629

Hubert, M., y Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. Computational Statistics & Data Analysis, 52(12), 5186-5201. https://doi.org/10.1016/j.csda.2007.11.008

Kaluba, P., Verbist, K. M. J., Cornelis, W. M., y Van Ranst, E. (2017). Spatial mapping of drought in Zambia using regional frequency analysis. Hydrological Sciences Journal, 62(11), 1825-1839. https://doi.org/10.1080/02626667.2017.1343475

Lee, S. H., y Maeng, S. J. (2003). Frequency analysis of extreme rainfall using L-moment. Irrigation and Drainage, 52(3), 219-230. https://doi.org/10.1002/ird.90

Maeda, E. E., Arevalo Torres, J., y Carmona-Moreno, C. (2013). Characterisation of global precipitation frequency through the L-moments approach: Characterisation of global precipitation frequency. Area, 45(1), 98-108. https://doi.org/10.1111/j.1475-4762.2012.01127.x

Miyamoto, S., Abe, R., Endo, Y., y Takeshita, J. (2015). Ward method of hierarchical clustering for non-Euclidean similarity measures. 2015 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR), 60-63. https://doi.org/10.1109/SOCPAR.2015.7492784

Moujahid, M., Stour, L., Agoumi, A., y Saidi, A. (2018). Regional approach for the analysis of annual maximum daily precipitation in northern Morocco. Weather and Climate Extremes, 21, 43-51. https://doi.org/10.1016/j.wace.2018.05.005

Núñez, J. H., Verbist, K., Wallis, J. R., Schaefer, M. G., Morales, L., y Cornelis, W. M. (2011). Regional frequency analysis for mapping drought events in north-central Chile. Journal of Hydrology, 405(3-4), 352-366. https://doi.org/10.1016/j.jhydrol.2011.05.035

Rivano, F. (2004). Análisis de eventos extremos de precipitación y su efecto en el diseño de drenaje superficial de tierras agrícolas del sur de Chile. Universidad Austral de Chile.

Wallis, J. R., Schaefer, M. G., Barker, B. L., y Taylor, G. H. (2007). Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington State. Hydrology and Earth System Sciences, 11(1), 415-442. https://doi.org/10.5194/hess-11-415-2007

Wilhite, D. A. (2000). Chapter 1 Drought as a Natural Hazard: Concepts and Definitions. 22.

Wilhite, D. A., Svoboda, M. D., y Hayes, M. J. (2007). Understanding the complex impacts of drought: A key to enhancing drought mitigation and preparedness. Water Resources Management, 21(5), 763-774. https://doi.org/10.1007/s11269-006-9076-5

Yue, S., Pilon, P., Phinney, B., y Cavadias, G. (2002). The influence of autocorrelation on the ability to detect trend in hydrological series. Hydrological Processes, 16(9), 1807-1829. https://doi.org/10.1002/hyp.1095