Comparative Analysis of Sarima and Holt-Winters Models for Rainfall Prediction in Cajamarca, Peru

José Romero-Rojas[1], Segundo Sánchez-Tello[2], Mariela Núñez-Figueroa[2], Sandra Rodriguez-Avila[1], Edwin Valencia-Castillo[1]

[1] Universidad Nacional de Cajamarca, Perú
[2] Universidad Nacional de Jaén, Perú

ABSTRACT
The research compares two models for rainfall prediction in the Cajamarca Valley, Peru, for the period 2021-2024 using the SARIMA econometric models and the Holt-Winters Additive Exponential Smoothing. The analysis of monthly rainfall in the Cajamarca Valley reveals considerable variability in the data. The rainiest year was 2012 with a total of 823.9 mm, while 2020 was the least rainy recording only 340.2 mm. Rainfall is unpredictable for agricultural activities, as well as for human consumption. In the Cajamarca valley there are significant fluctuations over the years, with months of less rainfall generally between May and September. When performing the comparative analysis of the statistics corresponding to both models in order to evaluate their predictive capacity, it has been observed that the Holt-Winters Additive Exponential Smoothing Model stands out significantly compared to the SARIMA Model. In this sense, the Root Mean Squared of Errors (RMSE) shows values of 28,697 versus 41,546, and the Mean Absolute Error (MAE) shows values of 20,044 as opposed to 29,260, both results being lower in the SARIMA model. It is worth noting that lower values in these metrics indicate greater model precision. In addition, the R^2 (Coefficient of Determination) evaluates the effectiveness of the fit of a model, where a higher value indicates a better quality of fit. In this situation, a value of 0.657 is observed, which exceeds the 0.287 corresponding to the SARIMA model. This high R^2 value suggests that the Holt-Winters Additive Exponential Smoothing model provides superior tuning and more accurate predictions of rainfall in the Cajamarca Valley. The methodology to predict rainfall in the Cajamarca Valley can be applicable to any city in the country and other places with similar characteristics.
KEYWORDS: Rainfall Forecasting, Time Series, SARIMA Models, Holt-Winters Additive Exponential Smoothing Model, Prediction.

1. INTRODUCTION

The purpose of the research is to predict rainfall in the valley of Cajamarca, department of Cajamarca - Peru, for the analysis monthly rainfall data in mm of rainfall from the period December 2001 - May 2024 were used. The technique for the analysis has been the ARIMA Econometric Model that explains the best rainfall prediction fit and its extension to the SARIMA models, as well as the additive Winter Exponential Smoothing Model as time series analysis models useful to identify regular prediction patterns.
For statistics, the integrated autoregressive model of moving averages is a statistical model that uses past values from the series to predict future values. ARIMA models are

in themselves a set of models that take into account the existence of a series of autocorrelations, and that turns out to be a fairly stable and accurate technique to forecast the future of a time series that is the case of the historical rainfall recorded monthly in the valley of Cajamarca - Peru.

Predicting rainfall in the Cajamarca valley is of vital importance to know the distribution of rainfall (precipitation) in the course of the year, its spatial variability, the quantity, the time of fall and the probable amounts of its occurrence that determine the planning and ordering of agricultural activities, hence knowing the rainfall regime for agricultural activities is of vital importance. In addition, they serve for the planning of water storage, the knowledge of surface water runoff and water flows in rivers for urban planning and fundamentally for the permanent supply of drinking water for the city of Cajamarca.

The proposed objective has been to develop a predictive model for rainfall in the Cajamarca Valley, Peru, using time series techniques such as SARIMA and the Holt-Winters Additive Exponential Smoothing model for the period 2021-2024 to improve water resources management in the region.

The effectiveness of the SARIMA models and the Holt-Winters Additive Exponential Smoothing Model has been compared to determine the degree of accuracy of the models by comparing them with data observed in the time series.

The hypothesis formulated is to determine if there is a significant relationship between past and future rainfall in the Cajamarca Valley, in addition to testing whether the SARIMA model provides more accurate predictions of rainfall in the Cajamarca Valley during the period 2021-2024 compared to the Holt-Winters Additive Exponential Smoothing Model.

2. THEORETICAL FRAMEWORK

The theoretical framework that supports this applied research is based on the analysis of a time series of monthly rainfall data in the Cajamarca valley using ARIMA Models and Winter's Additive Exponential Smoothing. These models are presented as an efficient tool to predict future values of monthly rainfall.

A time series is an ordered sequence of observations, each of which is associated with a moment in time. Examples of time series can be found in any field of science. In economics, when we look for data to study the behavior of an economic variable and its relationship with others over time, these data are often presented in the form of time series (Gujarati, Danodar, Porter, Dawn 2010). Thus, we can think of series such as daily stock prices, monthly exports, monthly consumption, quarterly profits, etc. In Meteorology, we have time series of temperature, amount of rain falling in a region, wind speed, etc. In Marketing, monthly or weekly sales series are of great interest. In Demography, the series of Total Population, birth rates, etc. are studied. In medicine, electrocardiograms or electroencephalograms. In Astronomy, solar activity, or in Sociology, data such as the number of crimes, etc. (Hamat, Pedro, Monsalve Abelardo 2015).

Contreras, Lucero (2024) carried out the temporal and predictive spatial-temporal analysis of the surface of the water bodies of the Chaychapampa River interbasin, Velille district, Chumbivilcas, the sample is 80 data that were processed in the STATA software using statistical validation tests, which were Kolmogorov-Smirnov, Mann Kendall Test, Kruskal Wallis Test and Dickey Fuller Test, the same that were the basis for the construction of the predictive model ARIMA (p, q).

Analyzing the temporal and predictive space of the surface of water bodies, the most appropriate model was ARIMA (1,1), performing a similar simulation from 2013 to 2022 and the period 2023 to 2026.

García, Elix (2024), carried out a comparative analysis of the semi-distributed hydrological models gr4j, socont and hbv applied to the forecast of daily flows in the Jequetepeque River basin. For the analysis of the data, he used the RStudio software. Concluding that the GR4J semi-distributed hydrological model obtained a Nash-Sutcliffe value of 0.829 in the calibration stage and 0.777 in the validation; for the semi-distributed SOCONT model, a Nash-Sutcliffe value of 0.562 was obtained in the calibration stage and 0.564 in the validation stage, and for the HBV model, a Nash-Sutcliffe value of 0.819 was obtained in the calibration stage and 0.810 in the validation. With the Nash-Sutcliffe values that are greater than 0.77 for the calibration and validation stages, it is concluded that more than 77% of the daily flows can be optimally predicted by the GR4J and HBV hydrological models, while those daily flows predicted by the SOCONT model that obtained very low values for Nash-Sutcliffe, The predictions are considered not acceptable in the basin under study.

Marín, Rigoberto (2021) used univariate ARIMA models and the prediction of precipitation data in the Amojú micro-basin, Jaén, Cajamarca in the period from January 1984 to August 2017. obtained the ARIMA model (1,0,1) with a better performance to make a forecast in relation to the observed data obY ˆ=2.80385+0.964867 Y ˆ t-12-0.851158εt-12:

Luque Gladys (2022) determined the appropriate regional probabilistic model for maximum rainfall in 24 hours. Observing that the maximum rainfall in 24 hours that occurs on the coast of the Piura region has a return period of 35 years on average and in the mountains of the Piura region they have a return period of 28 years on average.

Ruales Johanna and Jaranillo Manuel (2023) conducted a study to facilitate the planning of the expansion of the electricity grid in the medium term through analysis of electricity consumption in the residential, industrial and commercial sectors, using time series SARIMA: Case study Empresa Eléctrica Riobamba S.A. (EERSA)-Ecuador. The energy consumption prediction obtained from the SARIMA models compared to the actual generation that existed in the years 2021 and 2022 was efficient.

Gutiérrez Sergio (2021) used the WRF (Weather Research and Forecasting) model in the simulation of precipitation to determine its capacity to represent and predict rainfall events in the mid-mountain city of Manizales. Its analysis allowed to observe the prediction capabilities of the model; in this context, it concludes that the WRF model presents a medium-high performance in the prediction of rainfall events.

The studies confirm that the econometric models ARIMA, SARIMA, as well as the WRF (Weather Research and Forecasting) were efficient in predicting time series with a high degree of confidence.

3. METHODS

The methodology used in the research is based on a quantitative and predictive approach through the application of ARIMA econometric models, with the aim of analyzing the

monthly time series of the behavior of rainfall in mm of rainfall (1 mm. of precipitation = 1 l./m2 of surface) in the Cajamarca Valley - Peru in the period December 2021 - May 2024. And the purpose is the prediction of future values.
The obtaining of historical rainfall data has been provided by the National Service of Meteorology and Hydrology (SENAMHI) of the city of Cajamarca.
The **ARIMA** models have been developed by (Box & Jenkin, 1976) Time series analysis: forecasting and control. These models are based on the idea that the present value of a time series depends on its past values and previous random errors.
The general formula is:

$$Y_t = \phi_0 + \sum_{i=1}^{p} \phi_i Y_{y-1} + \sum_{j=1}^{q} \phi_j \varepsilon_{t-1} + \varepsilon_{t-1}$$

Where:
Y_t : Value of the time series in time t
ϕ_0 : Model Constant
ϕ_i : Coefficient of the autoregressive component
ϕ_j : Coefficient of the moving average component
ε_t : error term or white noise at time t
The **SARIMA** (Seasonal Autoregressive Integrated Moving Average) model is an extension of the ARIMA model that incorporates seasonal components, making it suitable for handling time series that present seasonal patterns (Mora, Mauricio 2023).
The general notation for a SARIMA model is expressed as: SARIMA(p,d,q)(P,D,Q)s
p: Order of the non-seasonal autoregressive part.
d: Number of non-seasonal differentiations required to make the series stationary.
q: Order of the Non-Seasonal Moving Average Part.
P: Order of the Seasonal Autoregressive Part.
D: Number of seasonal differentiations required.
Q: Order of the Seasonal Moving Average Part.
s: Number of periods in each season (e.g. 12 for monthly data).
SARIMA is used in periodic series defined as L that are less than 12 months. That is, regular and seasonal delays are taken into account

$$zt = \delta + \phi 1zt-1 + \phi 2zt-2 + \ldots + \phi pzt-p + \Phi 1zt-s + \Phi 2zt-2s + \ldots + \Phi Pzt-Ps + \theta 1at-1 + \theta 2at-2 + \ldots + \theta qat-q + \Theta 1at-s + \Theta 2at-2s + \ldots \Theta Qat-Qs + a$$

Where:
zt = Observed value at time t

δ = Model Constant

ϕ_i = Coefficients of the non-seasonal self-correlative part (AR)

Φ_i = Seasonal self-correcting part coefficients (SAR)

θ_i = Coefficients of the seasonal moving average (SMA)

a_t =Error term or white noise in time t

The Holt-Winters predictive exponential smoothing model, also known as triple exponential smoothing, is a technique used to forecast time series that exhibit both trends and seasonality. This method is particularly useful in contexts where the data shows cyclical patterns and fluctuations over time. Parra, Raúl (2023).

The model is based on three main components:

Level (L): Represents the average value of the time series in a given period.

Trend (T): Captures the direction and magnitude of long-term change in the data.

Seasonality (S): Reflects periodic variations that occur at regular intervals, such as months or quarters.

Model Formulas

The Holt-Winterss Additive Exponential Smoothing model can be either additive or multiplicative, depending on how the seasonal data behaves. Below are the formulas for the additive model, which is the most common and is the one that has been used for our research:

Level:

$$L_t = \alpha(Y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + T_{t-1})$$

Tendency

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$$

Seasonality

$$S_t = \gamma(Y_t - L_t) + (1-\gamma)S_{t-s}$$

Where

h = is the number of periods forward that you want to forecast.

Forecast

$$\hat{Y}_{t+h} = L_t + hT_t + S_{t+h-s}$$

Parameters

$\alpha =$ Smoothing factor for the level ($0< \alpha <1$)

$\beta =$ Smoothing factor for the trend ($0< \beta <1$)

$\gamma =$ Smoothing factor for seasonality ($0< \gamma <$These parameters determine how much weight is given to recent observations versus older ones.

The software used for data analysis and predictions has been SPSS and R Studio, known for its powerful analysis capabilities.

4. RESULTS

Analyzing the data on monthly rainfall in mm. of rain in the Cajamarca valley we find the following results:

Table 1. Monthly rainfall observed in the Cajamarca Valley -- Peru. Reference: Augusto Weberbaur Meteorological Station (mm).

	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SET	OCT	NOV	DIC	TOTAL
2001												90.9	**90.9**
2002	27.0	60.0	133.1	77.2	23.0	8.8	10.7	3.4	14.6	90.3	99.9	86.1	**634.1**

2003	51.1	61.4	103.5	42.1	30.7	22.3	1.8	10.6	14.8	46.0	63.8	80.7	**528.8**
2004	36.1	102.0	56.9	44.5	42.4	2.1	13.8	29.4	19.0	63.4	92.6	123.7	**625.9**
2005	84.9	53.7	136.6	54.0	7.2	4.5	0.6	3.5	31.2	92.3	30.0	87.8	**586.3**
2006	83.2	101.6	199.3	77.6	7.7	23.9	1.8	6.1	33.6	12.7	60.4	81.7	**689.6**
2007	95.4	17.5	182.4	111.5	29.0	1.4	10.7	6.4	11.6	118.9	97.6	68.8	**751.2**
2008	80.2	133.3	118.4	99.1	22.7	15.4	2.3	11.7	34.7	96.5	72.2	S/D	**686.5**
2009	180.7	74.6	110.5	78.8	42.2	17.9	12.3	3.9	11.8	78.5	109.4	74.2	**794.8**
2010	49.5	112.9	154.0	88.4	31.6	8.6	2.6	1.3	28.9	43.4	52.5	70.8	**644.5**
2011	76.6	73.3	125.2	102.0	16.7	0.4	8.3	0.0	47.1	31.5	24.4	109.7	**615.2**
2012	154.2	134.7	126.4	72.8	51.5	0.8	0.0	2.5	19.1	83.3	120.3	58.3	**823.9**
2013	61.5	98.0	213.6	73.8	62.6	7.5	5.7	8.9	3.7	110.7	17.0	51.9	**714.9**
2014	75.7	67.3	143.2	78.8	26.9	5.0	2.0	3.9	27.7	26.5	45.7	114.9	**617.6**
2015	184.7	55.4	202.2	63.0	75.8	3.0	4.4	0.1	27.8	16.8	99.6	39.5	**772.3**
2016	82.9	85.3	121.3	56.2	7.0	1.6	2.1	1.1	25.1	60.0	16.1	63.1	**521.8**
2017	77.5	72.3	138.9	78.6	47.2	12.0	2.3	20.9	21.2	65.3	63.2	168.1	**767.5**
2018	99.0	126.4	117.3	73.3	50.1	10.8	0.5	0.0	24.4	61.8	97.4	69.4	**730.4**
2019	46.9	107.3	172.7	78.1	37.4	9.1	11.8	0.0	7.6	121.8	60.4	162.7	**815.8**
2020	38.2	31.2	S/D	S/D	S/D	S/D	27.6	0.7	10.6	33.2	58.3	140.4	**340.2**
2021	100.8	54.3	138.6	129.5	38.2	10.5	4.5	12.3	23.7	108.8	58.3	45.7	**725.2**
2022	64.5	177.9	161.2	94.0	S/D	20.2	9.4	21.6	S/D	31.2	6.3	65.0	**651.3**
2023	87.6	119.8	122.6	68.6	49.4	0.0	3.5	3.3	3.1	106.5	87.8	173.4	**825.6**
2024	56.6	98.1	69.1	79.1	42.1			1					**361.9**
MÁXIMA	**191.2**	**177.9**	**230.2**	**129.5**	**75.8**	**23.9**	**27.6**	**29.4**	**47.1**	**121.8**	**120.3**	**173.4**	**1,348.1**
MEDIA	**86.9**	**88.3**	**142.5**	**77.3**	**35.9**	**8.7**	**6.5**	**6.4**	**21.6**	**67.2**	**66.4**	**92.1**	**699.8**
MÍNIMA	**27.0**	**17.5**	**56.9**	**42.1**	**7.0**	**0.0**	**0.0**	**0.0**	**3.1**	**12.7**	**6.3**	**39.5**	**212.1**

Note.- Own elaboration. Source: SENAMHI, Cajamarca.

Table No. 1 shows the descriptive statistics of rainfall in the Cajamarca valley. Observing that rainfall is not predictable and is very varied over the years, the rainiest year has been 2001 with 908.6 mm. and the least rainy year 2020 with 340.2 mm. the less rainy months from May to September. Monthly rainfall and its variations can be seen in figure No. 1. For their maximum, average and minimum values.

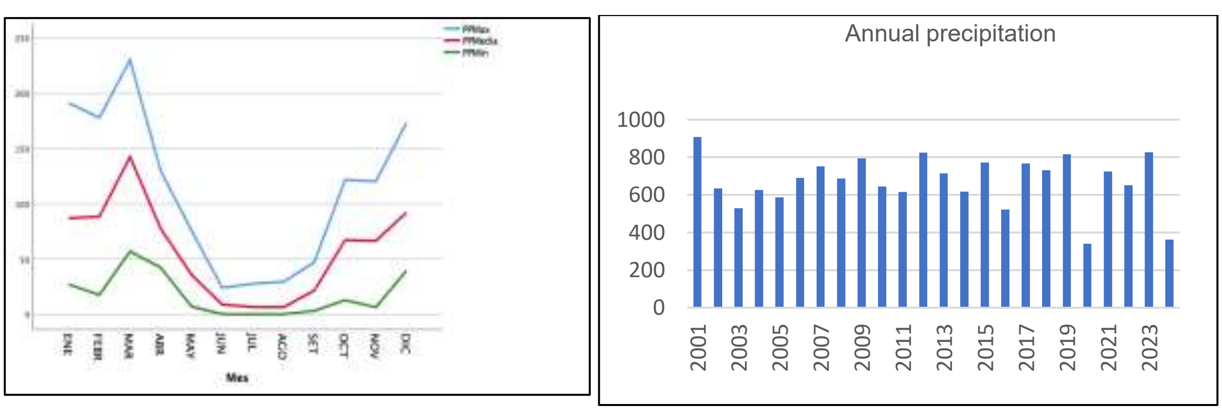


Fig. 1. Historical annual and monthly rainfall, maximum, average and minimum in mm. for the valley of Cajamarca - Peru.

Table 2. Statistics of the SARIMA model (0,0,0)(0,1,0)

Model	Model fit statistics			Ljung-Box Q		
	R2	RMSE	MAE	Statistical	DF	Sig.
Monthly rainfall mm. SARIMA Model	0.287	41.456	29.260	96.488	18	0.000

Table No. 3. SARIMA Model Parameters (0,0,0)(0,1,0)

Model			Estimate	SE	t	Sig.
Monthly precipitation mm. SARIMA Model	No transformation	Constant	0.280	2,.678	0.104	0.917
		Seasonal difference	1			

The exploratory analysis of the time series of rainfall in the Cajamarca valley irrefutably indicates that it does not present trends, but it does present seasonal patterns through the years as can be seen in figure No. 1. With this knowledge, we proceeded to look for a model that allows predicting future values with greater precession and several ARIMA models were analyzed, and their extension of the SARIMA Model.

The SARIMA (Seasonal Autoregressive Integrated Moving Average) model is an extension of the ARIMA model that incorporates seasonal components, making it suitable for handling time series that present seasonal patterns. This seasonal component allows us to model the behavior of the series based on its values in previous periods without using autoregressive or moving average components.

The Sarima model that best fits is the (0,0,0)(0,1,0) this time series model combines autoregressive, integrated and moving average components.

The SARIMA model (p,d,q)(P,D,Q) has 2 components:

Interpretation of the parameters p, d, q: (0,0,0):

$p = 0$: There is no autoregressive component. This means that the model does not use past values in the series to predict futures.

$d = 0$: No differentiation is applied to the series. This indicates that the series is already stationary and does not need transformations to stabilize its mean.

$q = 0$: There is no moving average component. This implies that the model does not use past errors in predictions.

Interpretation of the parameters P, D, Q: (0,1,0):

$P = 0$: There is no seasonal autoregressive component.

$D = 1$: Seasonal differentiation is applied to the series. This means that the differences between the values in the series in seasonal periods are being considered, which is useful for capturing seasonal patterns over a 12-month period.

$Q = 0$: There is no seasonal moving average component.

The model (0,0,0)(0,1,0) is the one that allows a better prediction of the ARIMA model, we can express it mathematically by the following equation:

$$Y_t = u + Y_{t-12} + \varepsilon_t$$

$$Y_t = 0.28 + Y_{t-12} + \varepsilon_t$$

Where:

u: Model Constant (0.28)
Yt: Represents the value of the time series at time t.
Yt−12: It is the value of the series in the same month of the previous year, which indicates that a seasonal difference (in this case, of one year) is being considered
εt: It is the term error or white noise over time that is assumed to follow a normal distribution.
The Yt−12 component clearly indicates that we are subtracting the value of the series in an earlier seasonal period (12 periods ago), which helps to capture seasonal patterns of rainfall in the Cajamarca valley. The term **εt** represents random fluctuations not explained by the model.

Table 3. Statistics of the Holt-Winters Additive Exponential Smoothing Model

Model	**Model fit statistics**			**Ljung-Box Q**		
	R2	**RMSE**	**MAE**	**Estadísticos**	**DF**	**Sig.**
Monthly precipitation mm. Holt-Winters Additive Exponential Smoothing Model	0.657	28.697	20.422	25.044	18	0.049

Table 4. Parameters of the Holt-Winters Additive Exponential Smoothing Model

Modelo			**Estimación**	**SE**	**t**	**Sig.**
Monthly precipitation mm. Holt-Winters Additive Exponential Smoothing Model	No transformation	Alfa (α) (level)	0.098	0.039	2.546	0.011
		Gamma (β) (trend)	1.639E-5	0.028	0.001	1,000
		Delta (γ) (seasonality)	1.225E-5	0.041	0.000	1.000

Tables 3 and 4 present the statistics and parameters of the Holt-Winters additive exponential smoothing model. In them we observe that they are highly significant for the statisticians and for the parameters of the model:
The parameters of the Holt-Winters additive exponential smoothing model are:
$\alpha = 0.098$ = Smoothing factor for the level.
$\beta = 0.00001639$ = Smoothing factor for the trend.
$\gamma = 0.00001225$ = Smoothing factor for seasonality
To predict using Winter's Additive Smoothed Predictive Model we calculate:

1. Level calculation (Lt)

$$L_t = 0.098 \cdot (Y_t - S_{t-s}) + (1 - 0.098) \cdot (L_{t-1} + T_{t-1})$$

2. Calculating the trend (Tt)

$$T_t = 0.00001639 \cdot (L_t - L_{t-1}) + (1 - 0.00001639) \cdot T_{t-1}$$

3. Calculating seasonality (St)

$$S_t = 0.00001225 \cdot (Y_t - L_t) + (1 - 0.00001225) \cdot S_{t-s}$$

4- Prediction

To forecast h periods forward, the following formula is used:

$$\hat{Y}_{t+h} = L_t + hT_t + S_{t+h-s}$$

Where:
Yt = Observed value over time t
Lt = Observed value over time t
Tt = Trend over time t
St = Seasonality over time t
St-1 = Seasonality of the same period of the previous year
h = Number of Forward Periods for Prediction
s = Number of periods as per year (s =12)
With the parameters (α, β and γ) allows us to predict future rainfall, for the present research the prediction has been made until December 2030, observing the seasonal trend. Figure No. 3 can be visualized compared with the values obtained with the SARIMA model and the observed values.

Table 5. Monthly rainfall: values predicted by the SARIMA model in the Cajamarca valley - Peru. Reference: Augusto Weberbaur Meteorological Station (mm).

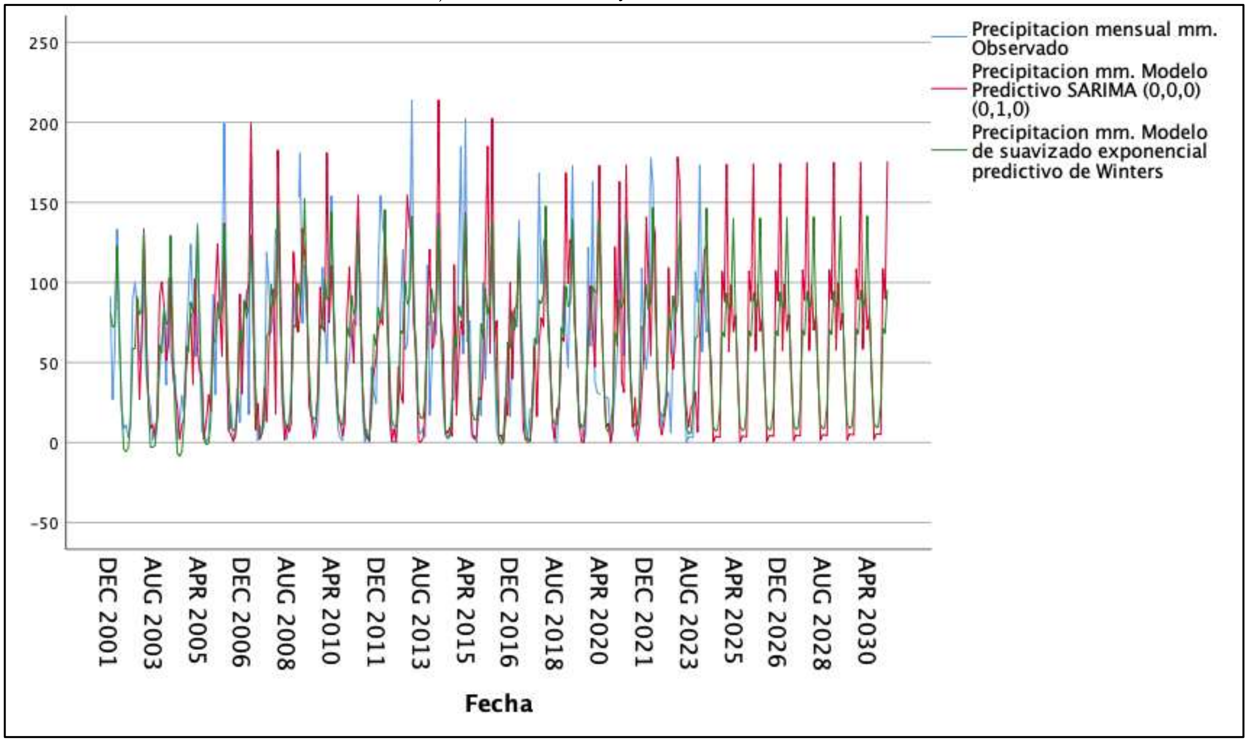
AÑO/MES	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Set.	Oct.	Nov.	Dic.
2001												-
2002	-	-	-	-	-	-	-	-	-	-	-	91.2
2003	27.3	60.3	133.4	77.5	23.3	9.1	11.0	3.7	14.9	90.6	100.2	86.4
2004	51.4	61.7	103.8	42.4	31.0	22.6	2.1	10.9	15.1	46.3	64.1	81.0
2005	36.4	102.3	57.2	44.8	42.7	2.4	14.1	29.7	19.3	63.7	92.9	124.0
2006	85.2	54.0	136.9	54.3	7.5	4.8	0.9	3.8	31.5	92.6	30.3	88.1
2007	83.5	101.9	199.6	77.9	8.0	24.2	2.1	6.4	33.9	13.0	60.7	82.0
2008	95.7	17.8	182.7	111.8	29.3	1.7	11.0	6.7	11.9	119.2	97.9	69.1
2009	80.5	133.6	118.7	99.4	23.0	15.7	2.6	12.0	35.0	96.8	72.5	69.4
2010	181.0	74.9	110.8	79.1	42.5	18.2	12.6	4.2	12.1	78.8	109.7	74.5
2011	49.8	113.2	154.3	88.7	31.9	8.9	2.9	1.6	29.2	43.7	52.8	71.1
2012	76.9	73.6	125.5	102.3	17.0	0.7	8.6	0.3	47.4	31.8	24.7	110.0
2013	154.5	135.0	126.7	73.1	51.8	1.1	0.3	2.8	19.4	83.6	120.6	58.6
2014	61.8	98.3	213.9	74.1	62.9	7.8	6.0	9.2	4.0	111.0	17.3	52.2
2015	76.0	67.6	143.5	79.1	27.2	5.3	2.3	4.2	28.0	26.8	46.0	115.2
2016	185.0	55.7	202.5	63.3	76.1	3.3	4.7	0.4	28.1	17.1	99.9	39.8
2017	83.2	85.6	121.6	56.5	7.3	1.9	2.4	1.4	25.4	60.3	16.4	63.4
2018	77.8	72.6	139.2	78.9	47.5	12.3	2.6	21.2	21.5	65.6	63.5	168.4
2019	99.3	126.7	117.6	73.6	50.4	11.1	0.8	0.3	24.7	62.1	97.7	69.7
2020	47.2	107.6	173.0	78.4	37.7	9.4	12.1	0.3	7.9	122.1	60.7	163.0
2021	38.5	31.5	173.3	78.7	38.0	9.7	27.9	1.0	10.9	33.5	58.6	140.7
2022	101.1	54.6	138.9	129.8	38.5	10.8	4.8	12.6	24.0	109.1	58.6	46.0
2023	64.8	178.2	161.5	94.3	38.8	20.5	9.7	21.9	24.3	31.5	6.6	65.3
2024	87.9	120.1	122.9	68.9	49.7	0.3	3.8	3.6	3.4	106.8	88.1	173.7

2025	56.9	98.4	69.4	79.4	42.4	0.6	4.1	3.9	3.7	107.1	88.4	174.0
2026	57.2	98.7	69.7	79.7	42.7	0.8	4.3	4.1	3.9	107.3	88.6	174.2
2027	57.4	98.9	69.9	79.9	42.9	1.1	4.6	4.4	4.2	107.6	88.9	174.5
2028	57.7	99.2	70.2	80.2	43.2	1.4	4.9	4.7	4.5	107.9	89.2	174.8
2029	58.0	99.5	70.5	80.5	43.5	1.7	5.2	5.0	4.8	108.2	89.5	175.1
2030	58.3	99.8	70.8	80.8	43.8	2.0	5.5	5.3	5.1	108.5	89.8	175.4

Table 6. Monthly rainfall: values predicted by the Holt-Winters Additive Exponential Smoothing Model in the Cajamarca valley - Peru. Reference: Augusto Weberbaur Meteorological Station (mm).

AÑO/MES	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Set.	Oct.	Nov.	Dic.
2001	-	-	-	-	-	-	-	-	-	-	-	81.0
2002	72.3	73.2	122.8	63.5	22.1	-4.3	-5.6	-3.4	11.5	58.8	58.9	90.0
2003	79.9	82.5	131.3	68.3	22.9	-2.7	-2.9	-1.9	13.6	60.7	56.3	84.1
2004	74.1	75.7	129.2	61.8	17.3	-6.6	-8.4	-5.7	12.0	59.7	57.1	87.6
2005	81.5	87.2	134.8	74.7	29.9	1.2	-1.1	-0.4	14.3	62.9	62.8	86.7
2006	77.1	83.1	135.7	81.7	38.5	9.1	7.9	7.9	22.0	70.1	61.5	88.4
2007	78.1	85.2	129.4	74.3	35.2	8.2	4.8	6.0	20.3	66.4	68.6	98.5
2008	85.9	90.7	145.8	82.8	41.6	13.4	10.9	10.6	25.0	72.9	72.2	99.3
2009	89.7	103.9	151.9	87.6	44.0	17.4	14.8	15.1	28.3	73.6	71.1	102.0
2010	89.6	91.0	144.0	84.7	42.3	14.8	11.6	11.3	24.6	72.0	66.2	91.9
2011	80.2	85.2	134.9	73.7	33.7	5.6	2.4	3.6	17.5	67.4	60.9	84.4
2012	77.2	90.1	145.3	83.2	39.4	14.2	10.2	9.8	23.3	69.9	68.2	100.4
2013	86.6	89.5	141.2	88.0	43.9	19.3	15.5	15.1	28.8	73.3	73.9	95.5
2014	81.5	86.3	135.3	75.8	33.3	6.3	3.5	3.9	18.2	66.1	59.2	85.0
2015	78.2	94.0	141.1	86.8	41.7	18.6	14.4	14.0	27.0	74.0	65.4	95.8
2016	80.6	86.2	137.0	75.2	30.6	1.8	-0.9	0.0	14.4	62.4	59.2	82.0
2017	70.5	76.6	127.0	67.9	26.2	1.8	0.1	0.9	17.2	64.6	61.6	88.9
2018	87.0	93.5	147.6	84.3	40.5	15.0	11.9	11.4	24.6	71.5	67.6	97.6
2019	85.1	86.8	139.6	82.6	39.4	12.8	9.7	10.5	23.8	69.2	71.3	97.3
2020	94.1	94.0	138.7	78.4	35.6	9.2	6.5	9.2	22.6	68.4	62.0	88.7
2021	84.1	91.1	138.3	78.1	40.4	13.7	10.7	10.7	25.2	72.0	72.6	98.3
2022	83.4	87.0	146.7	87.9	45.7	19.3	16.7	16.6	31.3	78.3	70.7	91.5
2023	79.2	85.4	139.6	77.7	34.0	9.1	5.5	5.9	20.0	65.3	66.3	95.5
2024	93.5	95.2	146.4	78.5	35.8	10.0	7.3	7.9	22.2	69.2	66.2	93.3
2025	83.6	88.9	139.8	79.5	36.8	10.3	7.7	8.3	22.5	69.5	66.5	93.6
2026	83.9	89.3	140.1	79.9	37.1	10.7	8.0	8.6	22.9	69.9	66.9	93.9
2027	84.3	89.6	140.5	80.2	37.4	11.0	8.4	8.9	23.2	70.2	67.2	94.3
2028	84.6	90.0	140.8	80.6	37.8	11.4	8.7	9.3	23.6	70.5	67.5	94.6
2029	84.9	90.3	141.2	80.9	38.1	11.7	9.0	9.6	23.9	70.9	67.9	95.0
2030	85.3	90.7	141.5	81.2	38.5	12.0	9.4	10.0	24.2	71.2	68.2	95.3

Fig. 3.- Rainfall observed and predictive by the SARIMA Model (0,0,0)(0,1,0) and the Holt-Winters Additive Exponential Smoothing Model in mm. for the Cajamarca Valley - Peru



Tables 5 and 6 . and graph 3 presents the results of the SARIMA models and the Holt-Winters Additive Exponential Smoothing Model, using the model's performance statistics: RMSE (Root Mean Square Error), MAE (Mean Absolute Error) and R^2 We find:

Modelo SARIMA: RMSE: 41.546; MAE: 29.260, R^2: 0.287

Holt-Winters Additive Exponential Smoothing Model: RMSE: 28.697, MAE: 20.044, R^2: 0.657

The RMSE**:** measures the average magnitude of errors between predictions and actual values. A lower RMSE indicates a better fit of the model.

In this case, the Holt-Winters additive exponential smoothing model has an RMSE of 28.697, which is significantly lower than the RMSE of the SARIMA model (41.546). This suggests that Winter's model performs better in terms of accuracy in forecasting rainfall in the Cajamarca Valley.

MAE: It also measures the accuracy of the model, but it is based on the average absolute error without considering the direction of error.

Again, the Holt-Winters Additive Exponential Smoothing Model shows an ESA of 20.044, which is lower than the SARIMA Model ESA (29.260). This reinforces the previous conclusion: the Holt-Winters Additive Exponential Smoothing Model is more accurate in its predictions.

R^2 (Coefficient of Determination): Indicates the proportion of the total variability in the data that is explained by the model. A higher value indicates a better ability to predict future values of rainfall in the valleys of Cajamarca.
The Holt-Winters Additive Exponential Smoothing Model has an R^2 of 0.657, which means that it explains approximately 65.7% of the variability in the data. In contrast, the SARIMA model has an R^2 of only 0.287, indicating that it only explains 28.7% of the variability. This suggests that Winter's model is much more effective at capturing trends and patterns in rainfall in the Cajamarca Valley. which is visualized in Graph Nº3 the comparison of the observed and predictive rainfall by the SARIMA Model (0,0,0)(0,1,0) and the Additive Holt-Winters Exponential Smoothing Model in mm. for the valley of Cajamarca - Peru.

5. DISCUSSION AND CONCLUSION

Based on the performance statistics of the models analyzed, we conclude:
Precision: The Holt-Winters Additive Exponential Smoothing Model is clearly superior to the SARIMA model in both RMSE and MAE, indicating that it makes more accurate predictions.
Model Fit: R^2 also supports this conclusion, showing that Winter's model explains a higher proportion of the variability in the data. Therefore, we reject the hypothesis that the SARIMA Model provides more accurate predictions of rainfall in the Cajamarca Valley during the period 2021-2024 compared to the Holt-Winters additive exponential smoothing model.
The SARIMA and Holt-Winters Additive Exponential Smoothing models have identified seasonal patterns with great variability, observing two hydrologically different seasons, a rainy period from October to April and a relatively dry period from May to September. Therefore, the results of the research will be used to make decisions on water storage in the rainy season to meet water needs in the rainfed period, both for agricultural activities and for the supply of human consumption.
Temporal variations show that there are large fluctuations from one year to the next that affect the planning and management of water resources.
The methodology to predict rainfall in the Cajamarca Valley can be applicable to any city, the country and other places with similar characteristics. Likewise, the use of the models used, such as the SARIMA Model and the Additive Holt-Winters Exponential Smoothing Model, to predict future values through time series, is applicable in other fields of science such as Economics. Engineering, medicine and others.

Acknowledgements
The authors would like to express their gratitude to all the participants and the authorities of the Universidad Nacional de Cajamarca and Universidad Nacional de Jaén, who allowed the completion of the study.

Author contributions statement

JR, SS, MN, SR, and EV designed the study. JR, SS, MN, SR, and EV collected data. JR, SS, and EV curated and analyzed the dataset. JR, SS, MN, and SR wrote the first version of the manuscript. SR supervised the project. JR, SS, MN, SR, and EV arranged funding. All authors read, reviewed and approved the final version of the manuscript.

Funding: The research was conducted independently by the researchers and no funding was received.

Data availability: The data supporting the findings of this study are available from the corresponding author, upon reasonable request.

Statements and Declarations

Ethics approval and consent to participate

All participants declared their intention to participate in the study

Competing Interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

REFERENCES

1. Box, G, & Jenkin, G (1976) Time Series Analysis: Forecasting and Control (2a. ed.). Oakland, CA: Holden-Day.

2. Diaz Eddy, Fustamante Danly, Gave Joshua, Heredia, Keico, Sedano María (2023) Pronóstico de la producción de papa en la cuenca Jequetepeque-Cajamarca en base a las variables meteorológicas utilizando técnicas de Machine Learning, repositorio.esan.edu.pe. esan.edu.pe

3. Contreras L, Huarancca J, Bustamante C (2024) Análisis espacio temporal y predictivo de la superficie de los cuerpos de agua de la intercuenca del río Chaychapampa, distrito de Velille, Chumbivilcas, Cusco. repositorio.continental.edu.pe. 143 pags continental.edu.pe

4. Condezo M (2021) repositorio.lamolina.edu.pe. Análisis de los sistemas de recursos hídricos en los proyectos especiales Chavimochic y Chinecas-Perú. lamolina.edu.pe

5. García, E (2024) Análisis comparativo de los modelos hidrológicos semidistribuidos gr4j, socont y hbv aplicados al pronóstico de caudales diarios en la cuenca del Río Jequetepeque repositorio.unc.edu.pe. 84 Pgs. unc.edu.pe

6. Gutierrez, S (2021) Desempeño del modelo WRF en la simulación de la precipitación en una ciudad de media montaña: Caso de estudio Manizales. repositorio.unal.edu.co. tésis 104-pgs. unal.edu.co

7. Gujarati D, Porter D (2010) Econometría 104-pgs. ISBN: 978-607-15-0294-0

8. Hamat P, Monsalve A (2015) Introducción al Análisis de Series de Tiempo con Aplicaciones a la Econometría y Finanzas. Escuela Venezolana de Matemáticas EMALCA 201 pags. https://www.researchgate.net/publication/295106544
9. Escuela Venezolana de Matemáticas EMALCA
10. Ordoñez L (2022) Aplicación de redes neuronales artificiales (RNA) al modelamiento de lluvia-escorrentía en la cuenca del río Chancay Lambayeque. Repositorio.uss.edu.pe. 184 pgs. uss.edu.pe
11. Lozano, G (2021) Business analytics para el pronóstico de deserciones de estudiantes en el Colegio Héroes de la Breña. Tésis 104 pgs. repositorio.ucv.edu.pe. ucv.edu.pe
12. Luque G (2022) Modelamiento probabilístico de las lluvias máximas en 24 horas en la vertiente del pacífico de la región Piura 1965-2019. Repositorio.senamhi.gob.pe. senamhi.gob.pe
13. Marín R (2021) Utilización de modelos univariantes ARIMA y la predicción de datos de precipitación en la microcuenca Amojú, Jaén, Cajamarca, repositorio.ucv.edu.pe. 86 pgas. ucv.edu.pe
14. Mora M (2023) Modelos SARIMA y Órdenes Estacionales. https://es.linkedin.com/pulse/tiempedia-modelos-sarima-y-%C3%B3rdenes-estacionales-mora-caballero-qk97e.
15. Narváez C, Ortega LD, Serrano J (2024) Proyección energética de la central eólica Huascachaca de 50 MW aplicando redes neuronales. dspace.ups.edu.ec . ups.edu.ec
16. Parra R (2023) Pronósticos (Forecast) de la demanda - Método Triple Suavizado exponencial o de Holt-Winters. https://es.linkedin.com/pulse/triple-suavido-exponencial-o-metodo-de-holt-winters-raul
17. Ruales J, Jaranillo M (2023) Planificación de la expansión de la red eléctrica a mediano plazo mediante análisis de consumo eléctrico en los sectores residencial.
18. Valderrama J, Purizaca C, Muñoz S, Tuesta V, Mejía H (2021) Importancia de las redes neuronales artificiales en la ingeniería civil: Una revisión sistemática de la literatura. scielo.org.co. scielo.org.co