

**UNIVERSIDAD DE PANAMÁ**  
**VICERRECTORÍA DE INVESTIGACIÓN Y POSTGRADO**  
**PROGRAMA DE MAESTRÍA EN ESTADÍSTICA APLICADA**

INFORME DE PROYECTO DE INTERVENCIÓN

**MODELO ESTADÍSTICO PARA ESTIMAR LA PRECIPITACIÓN  
PLUVIAL SOBRE LA CUENCA DEL RÍO CHIRIQUÍ, PERIODO  
1972-2016**

PRESENTADO COMO UNO DE LOS REQUISITOS PARA OBTENER EL GRADO  
DE MAESTRO EN ESTADÍSTICA APLICADA

**JULIO JAVIER SALDAÑA RIVERA**

PANAMÁ, REPÚBLICA DE PANAMÁ

2018

## DEDICATORIA

A mis queridos padres, Javier e Isabel, y especialmente a mi esposa e hijo, Rosaura Esther y Jordi Javier; que en todo momento me alentaron y apoyaron para la culminación de este trabajo.

*Julio Javier*

## AGRADECIMIENTO

... A Dios, porque sin él nada de esto hubiese sido posible.

Quiero plasmar mi más sincero agradecimiento a la Vicerrectoría de Investigación y Postgrado, por permitirme ingresar al programa de estudio superior de la Universidad de Panamá (UP).

Al cuerpo docente de la Facultad de Ciencias Naturales, Exactas y Tecnología de la Universidad de Panamá (UP), por la formación que me entregaron durante los años de estudio; y quienes, a su vez me motivaron en la investigación y desarrollo de este trabajo permitiéndome alcanzar mi meta para convertirme en el profesional que soy hoy.

A mi familia, en especial a mi padre, por el esfuerzo que hizo durante su vida para que yo me convirtiera en un hombre de bien; a mi madre, por su cariño y dedicación para que siguiera adelante en el desarrollo de mis estudios; y muy especialmente a mi esposa, por su apoyo incondicional y dedicación para que culminara satisfactoriamente mis estudios.

A mis compañeros de grupo del máster y a todos aquellos que de una u otra forma me tendieron su mano.

...Mil gracias.

## ÍNDICE

DEDICATORIA .....	2
AGRADECIMIENTO .....	3
1. RESUMEN - ABSTRACT.....	13
1.1. RESUMEN.....	13
1.2. ABSTRACT.....	14
2. MARCO INTRODUCTORIO .....	15
2.1. Planteamiento del problema y objetivos.....	17
2.2. El Problema:.....	18
2.3. Objetivo General:.....	18
2.4. Objetivos específicos:.....	18
2.5. Justificación: .....	19
3. MARCO REFERENCIAL .....	20
3.1. Fundamentación teórica .....	20
3.1.1. ¿Qué se entiende por precipitación pluvial y cuenca? .....	20
3.1.2. ¿Qué se entiende por modelo estadístico? .....	22
3.1.3. ¿De qué trata la Metodología Box - Jenkins?.....	23
3.1.4. ¿Qué una serie temporal? .....	25
3.2. Revisión literaria.....	28
3.2.1. Cuenca del río Chiriquí: .....	28

3.2.2. Metodología Box - Jenkins: .....	29
4. ASPECTOS METODOLÓGICOS .....	30
4.1. Diseño de la investigación .....	31
4.2. Metodología de la investigación .....	33
4.2.1. Contexto, población - muestra .....	33
4.2.2. Criterios de selección: .....	35
4.2.3. Procedimiento utilizado: .....	36
4.3. Hipótesis de Trabajo: .....	37
4.4. Variables: .....	38
4.4.1. Análisis estadístico: .....	39
4.4.2. Instrumento: .....	54
5. ANÁLISIS DE LOS RESULTADOS Y DISCUSIÓN .....	54
5.1. Análisis descriptivo .....	55
5.2. Precisión en la estimación de la precipitación .....	56
5.3. Análisis inferencial .....	57
5.3.1. Análisis de consistencia de los registros pluviométricos.....	57
5.3.2. Modelo ARIMA con la metodología de Box-Jenkins .....	63
5.3.3. Comparación de los datos pronosticados versus los reales del año 2016107	
6. CONCLUSIONES Y RECOMENDACIONES .....	109
6.1. En relación a las preguntas-objetivos del trabajo .....	110

6.2. Perspectivas a futuro.....	113
7. REFERENCIAS BIBLIOGRÁFICAS.....	114
ANEXO 1: Estación Finca Lérída (108-001) .....	119
ANEXO 2: Estación Los Naranjos (108-017) .....	120
ANEXO 3: Estación Veladero Gualaca (108-014) .....	121
ANEXO 4: Estación David (108-023).....	122
ANEXO 5: Cuenca 108 - río Chiriquí .....	123
ANEXO 6: Tabla de Datos: Estación Finca Lérída (108-001) zona alta .....	124
ANEXO 7: Tabla de Datos: Estación David (108-023) zona baja .....	125
ANEXO 8: Tabla de Datos: Estación Los Naranjos (108-017) zona alta .....	126
ANEXO 9: Tabla de Datos: Estación Veladero Gualaca (108-014) zona baja .....	127
ANEXO 10: Modelos AR y MA teóricos .....	128
ANEXO 11: Valores críticos de Distribución Chi-Cuadrado $\chi^2$ .....	130

## ÍNDICE DE TABLAS

TABLA 1: ESTACIONES METEOROLÓGICAS DE ESTUDIO .....	35
TABLA 2: IDENTIFICACIÓN DE MODELOS .....	48
TABLA 3: ESTADÍSTICOS DE ERROR.....	53
TABLA 4: RESUMEN ESTADÍSTICO ANUAL DE LA PRECIPITACIÓN PLUVIAL ESTACIÓN FINCA LÉRIDA..	55
TABLA 5: RESUMEN ESTADÍSTICO ANUAL DE LA PRECIPITACIÓN PLUVIAL-ESTACIÓN DAVID .....	56
TABLA 6: PRECISIÓN EN LA ESTIMACIÓN DE LA PRECIPITACIÓN: ESTACIÓN FINCA LÉRIDA.....	57
TABLA 7: PRECISIÓN EN LA ESTIMACIÓN DE LA PRECIPITACIÓN: ESTACIÓN DAVID.....	57
TABLA 8: PRUEBA DE HOMOGENEIDAD ANUAL: 108-001 .....	58
TABLA 9: PRUEBA DE HOMOGENEIDAD ANUAL: 108-023.....	60
TABLA 10: PRUEBA DE LEVENE (HOMOGENEIDAD EN VARIANZA) ESTACIÓN FINCA LÉRIDA.....	65
TABLA 11: PRUEBA DE LEVENE (HOMOGENEIDAD EN VARIANZA) ESTACIÓN DAVID.....	66
TABLA 12: CONTRASTES DE RAÍCES UNITARIAS, ESTACIÓN FINCA LÉRIDA .....	80
TABLA 13: CONTRASTES DE RAÍCES UNITARIAS, ESTACIÓN DAVID .....	81
TABLA 14: ESTIMACIÓN DE COEFICIENTES Y PARÁMETROS DE CORRELACIÓN DEL MODELO 1.....	87
TABLA 15: ESTIMACIÓN DE COEFICIENTES Y PARÁMETROS DE CORRELACIÓN DEL MODELO 2.....	88
TABLA 16: ESTIMACIÓN DE COEFICIENTES Y PARÁMETROS DE CORRELACIÓN DEL MODELO 1.....	89
TABLA 17: ESTIMACIÓN DE COEFICIENTES Y PARÁMETROS DE CORRELACIÓN DEL MODELO 2.....	90
TABLA 18: PARÁMETROS ESTIMADOS DEL MODELO 1: <i>ARIMA</i> (2, 1, 1)(2, 1, 2) <sub>12</sub> - ESTACIÓN FINCA LÉRIDA .....	92
TABLA 19: ESTIMADOS DEL MODELO 2: <i>ARIMA</i> (2, 1, 2)(2, 1, 1) <sub>12</sub> - ESTACIÓN FINCA LÉRIDA ..	95
TABLA 20: PARÁMETROS DE CORRELACIÓN Y ESTIMADOS DEL MODELO 1: <i>ARIMA</i> (2, 1, 1)(2, 1, 1) <sub>12</sub> - ESTACIÓN DAVID.....	98

<b>TABLA 21: PARÁMETROS DE CORRELACIÓN Y ESTIMADOS DEL MODELO 2:</b>	
<b><i>ARIMA</i>(2, 1, 1)(3, 1, 1)<sub>12</sub> - ESTACIÓN DAVID.....</b>	<b>101</b>
<b>TABLA 22: ESTADÍSTICAS DE ERROR Y ANÁLISIS RESIDUAL DE LOS MODELOS PROPUESTOS - ESTACIÓN FINCA LÉRIDA .....</b>	<b>104</b>
<b>TABLA 23: ESTADÍSTICAS DE ERROR Y ANÁLISIS RESIDUAL DE LOS MODELOS PROPUESTOS - ESTACIÓN DAVID .....</b>	<b>104</b>
<b>TABLA 24: PRONÓSTICOS DEL MODELO <i>ARIMA</i>(2, 1, 2)(2, 1, 1)<sub>12</sub> EN EL PRIMER CUATRIMESTRE DEL 2016 PARA LA ESTACIÓN FINCA LÉRIDA .....</b>	<b>105</b>
<b>TABLA 25: PRONÓSTICOS DEL MODELO <i>ARIMA</i>(2, 1, 1)(2, 1, 1)<sub>12</sub> EN EL PRIMER CUATRIMESTRE DEL 2016 PARA LA ESTACIÓN DAVID .....</b>	<b>106</b>
<b>TABLA 26: TOTAL DE PRECIPITACIÓN PLUVIAL DEL PRIMER CUATRIMESTRE PARA LAS ESTACIONES METEOROLÓGICAS EN ESTUDIO, AÑO: 2016 .....</b>	<b>107</b>
<b>TABLA 27: COMPARACIÓN DE PRONÓSTICOS DEL MODELO <i>ARIMA</i>(2, 1, 1)(2, 1, 1)<sub>12</sub> VERSUS DATOS REALES DEL PRIMER CUATRIMESTRE, AÑO: 2016 - ESTACIÓN DAVID .....</b>	<b>108</b>

## ÍNDICE DE GRÁFICAS

GRÁFICA 1: CONSISTENCIA DE LOS REGISTROS PLUVIOMÉTRICOS ESTACIÓN FINCA LÉRIDA - LOS NARANJOS (ZONA ALTA) .....	61
GRÁFICA 2: CONSISTENCIA DE LOS REGISTROS PLUVIOMÉTRICOS ESTACIÓN DAVID - VELADERO GUALACA (ZONA BAJA) .....	62
GRÁFICA 3: SERIE DE TIEMPO ESTACIÓN FINCA LÉRIDA .....	63
GRÁFICA 4: SERIE DE TIEMPO ESTACIÓN DAVID .....	64
GRÁFICA 5: GRÁFICO CAJA Y BIGOTES ESTACIÓN FINCA LÉRIDA.....	65
GRÁFICA 6: GRÁFICO CAJA Y BIGOTES ESTACIÓN DAVID.....	67
GRÁFICA 7: TRANSFORMACIÓN RAÍZ CUADRADA - ESTACIÓN FINCA LÉRIDA .....	68
GRÁFICA 8: TRANSFORMACIÓN LOGARITMO NATURAL - ESTACIÓN FINCA LÉRIDA .....	68
GRÁFICA 9: TRANSFORMACIÓN RAÍZ CUADRADA - ESTACIÓN DAVID .....	69
GRÁFICA 10: TRANSFORMACIÓN $x^{0,15}$ - ESTACIÓN DAVID .....	69
GRÁFICA 11:GRÁFICA LINEAL DE LA SERIE ESTACIONARIA EN VARIANZA, ESTACIÓN FINCA LÉRIDA .	71
GRÁFICA 12: GRÁFICA LINEAL DE LA SERIE ESTACIONARIA EN VARIANZA, ESTACIÓN DAVID.....	71
GRÁFICA 13: FUNCIÓN DE AUTOCORRELACIÓN SIMPLE DE LA SERIE TRANSFORMADA - ESTACIÓN FINCA LÉRIDA .....	72
GRÁFICA 14: FUNCIÓN DE AUTOCORRELACIÓN PARCIAL DE LA SERIE TRANSFORMADA - ESTACIÓN FINCA LÉRIDA .....	72
GRÁFICA 15: FUNCIÓN DE AUTOCORRELACIÓN SIMPLE - ESTACIÓN DAVID .....	73
GRÁFICA 16: FUNCIÓN DE AUTOCORRELACIÓN PARCIAL DE LA SERIE - ESTACIÓN DAVID.....	73
GRÁFICA 17: DIFERENCIADO ESTACIONAL DE LA SERIE TRANSFORMADA - ESTACIÓN FINCA LÉRIDA.	74
GRÁFICA 18: DIFERENCIADO ESTACIONAL DE LA SERIE TRANSFORMADA - ESTACIÓN DAVID.....	75
GRÁFICA 19: FAS DE LA SERIE TRANSFORMADA Y DIFERENCIADA - ESTACIÓN FINCA LÉRIDA .....	76

GRÁFICA 20: FAP DE LA SERIE TRANSFORMADA Y DIFERENCIADA - ESTACIÓN FINCA LÉRIDA .....	76
GRÁFICA 21: FAS DE LA SERIE TRANSFORMADA Y DIFERENCIADA - ESTACIÓN DAVID .....	77
GRÁFICA 22: FAP DE LA SERIE TRANSFORMADA Y DIFERENCIADA - ESTACIÓN DAVID .....	77
GRÁFICA 23: DIFERENCIADO DE LA PARTE ESTACIONAL Y REGULAR DE LA SERIE TRANSFORMADA - ESTACIÓN FINCA LÉRIDA .....	78
GRÁFICA 24: DIFERENCIADO DE LA PARTE ESTACIONAL Y REGULAR DE LA SERIE TRANSFORMADA - ESTACIÓN DAVID .....	79
GRÁFICA 25: FAS DE LA SERIE TRANSFORMADA Y DIFERENCIADA EN LA PARTE ESTACIONAL Y REGULAR SOBRE LOS 36 PRIMEROS RETARDOS - ESTACIÓN FINCA LÉRIDA .....	83
GRÁFICA 26: FAP DE LA SERIE TRANSFORMADA Y DIFERENCIADA EN LA PARTE ESTACIONAL Y REGULAR SOBRE LOS 36 PRIMEROS RETARDOS - ESTACIÓN FINCA LÉRIDA .....	84
GRÁFICA 27: FAS DE LA SERIE TRANSFORMADA Y DIFERENCIADA EN LA PARTE ESTACIONAL Y REGULAR SOBRE LOS 36 PRIMEROS RETARDOS - ESTACIÓN DAVID.....	85
GRÁFICA 28: FAP DE LA SERIE TRANSFORMADA Y DIFERENCIADA EN LA PARTE ESTACIONAL Y REGULAR SOBRE LOS 36 PRIMEROS RETARDOS - ESTACIÓN DAVID.....	85
GRÁFICA 29: HISTOGRAMA DE RESIDUALES DEL MODELO 1: $ARIMA(2, 1, 1)(2, 1, 2)_{12}$ - ESTACIÓN FINCA LÉRIDA .....	91
GRÁFICA 30: NORMALIDAD DE RESIDUALES DEL MODELO 1: $ARIMA(2, 1, 1)(2, 1, 2)_{12}$ - ESTACIÓN FINCA LÉRIDA .....	92
GRÁFICA 31: AUTOCORRELOGRAMA DE LOS RESIDUALES DEL MODELO 1: $ARIMA(2, 1, 1)(2, 1, 2)_{12}$ - ESTACIÓN FINCA LÉRIDA .....	93
GRÁFICA 32: HISTOGRAMA DE RESIDUALES DEL MODELO 2: $ARIMA(2, 1, 2)(2, 1, 1)_{12}$ - ESTACIÓN FINCA LÉRIDA .....	94

GRÁFICA 33: NORMALIDAD DE RESIDUALES DEL MODELO 2: $ARIMA(2, 1, 2)(2, 1, 1)_{12}$ - ESTACIÓN FINCA LÉRIDA .....	95
GRÁFICA 34: AUTOCORRELOGRAMA DE LOS RESIDUALES DEL MODELO 2: $ARIMA(2, 1, 2)(2, 1, 1)_{12}$ - ESTACIÓN FINCA LÉRIDA .....	96
GRÁFICA 35: HISTOGRAMA DE RESIDUALES DEL MODELO 1: $ARIMA(2, 1, 1)(2, 1, 1)_{12}$ - ESTACIÓN DAVID .....	97
GRÁFICA 36: NORMALIDAD DE RESIDUALES DEL MODELO 1: $ARIMA(2, 1, 1)(2, 1, 1)_{12}$ - ESTACIÓN DAVID .....	98
GRÁFICA 37: AUTOCORRELOGRAMA DE LOS RESIDUALES DEL MODELO 1: $ARIMA(2, 1, 1)(2, 1, 1)_{12}$ - ESTACIÓN DAVID.....	99
GRÁFICA 38: HISTOGRAMA DE RESIDUALES DEL MODELO 2: $ARIMA(2, 1, 1)(3, 1, 1)_{12}$ - ESTACIÓN DAVID .....	100
GRÁFICA 39: NORMALIDAD DE RESIDUALES DEL MODELO 2: $ARIMA(2, 1, 1)(3, 1, 1)_{12}$ - ESTACIÓN DAVID .....	101
GRÁFICA 40: AUTOCORRELOGRAMA DE LOS RESIDUALES DEL MODELO 2: $ARIMA(2, 1, 1)(3, 1, 1)_{12}$ - ESTACIÓN DAVID.....	102
GRÁFICA 41: PRONÓSTICOS DEL MODELO $ARIMA(2, 1, 2)(2, 1, 1)_{12}$ EN EL PRIMER CUATRIMESTRE DEL 2016 PARA LA ESTACIÓN FINCA LÉRIDA .....	105
GRÁFICA 42: PRONÓSTICOS DEL MODELO $ARIMA(2, 1, 1)(2, 1, 1)_{12}$ EN EL PRIMER CUATRIMESTRE DEL 2016 PARA LA ESTACIÓN DAVID .....	106
GRÁFICA 43: COMPARACIÓN DE PRONÓSTICOS DEL MODELO $ARIMA(2, 1, 1)(2, 1, 1)_{12}$ VERSUS DATOS REALES DEL PRIMER CUATRIMESTRE, AÑO: 2016 - ESTACIÓN DAVID .....	108

## ÍNDICE DE ILUSTRACIONES

ILUSTRACIÓN 1: MAPA DE CUENCAS HIDROGRÁFICAS EN LA REPÚBLICA DE PANAMÁ .....	22
ILUSTRACIÓN 2: DIVISIÓN POLÍTICA ADMINISTRATIVA DE LA CUENCA DEL RÍO CHIRIQUÍ POR DISTRITO Y CORREGIMIENTO, AÑO 2003 .....	34
ILUSTRACIÓN 3: RANGO DE ACEPTACIÓN DEL 95% PARA LA PRUEBA DE RACHAS .....	43
ILUSTRACIÓN 4: ZONA DE RECHAZO DE LA DISTRIBUCIÓN CHI - CUADRADO .....	52

## 1. RESUMEN - ABSTRACT

### 1.1. RESUMEN

En esta investigación se logra crear un modelo estadístico que se ajusten a los registros históricos mensuales de la precipitación pluvial sobre la zona alta y baja de la cuenca del río Chiriquí con la aplicación de la metodología Box - Jenkins; por consiguiente, la investigación se ajusta a un estudio de tipo longitudinal no experimental, y se le considera de tipo longitudinal de tendencia con un enfoque retrospectivo a lo largo de los meses en la cuenca del río Chiriquí ya ocurridos. Es por ello que, el estudio se concentra en las estaciones meteorológicas que poseen los registros completos de manera mensual con 44 registros anuales de la zona alta y baja de la cuenca del río Chiriquí, ubicada en la provincia de Chiriquí, República de Panamá. Además, se realizaron dos tipos de análisis estadísticos, que corresponden al análisis descriptivo y el análisis inferencial de los datos; con éste último, se desarrolla una clara percepción científica del comportamiento, como la tendencia y estacionalidad que predomina en la precipitación pluvial sobre la cuenca del río Chiriquí, desde su pasado, presente y futuro, logrando así, beneficiar y prevenir posibles desastres al sector agricultor, ganadero y energético, al igual que empresas hidroeléctricas, meteorológicas y otras instituciones.

La mejor estructura que logra modelar la variable precipitación pluvial fue el modelo autorregresivo integral de media móvil estacional multiplicativo (SARIMA).

El modelo  $ARIMA(2, 1, 2)(2, 1, 1)_{12}$  para la estación Finca Lérica y  $ARIMA(2, 1, 1)(2, 1, 1)_{12}$  para la estación David; fueron utilizados para predecir la precipitación pluvial del primer cuatrimestre del año 2016, en cada una de las diferentes

estaciones ubicadas en la cuenca del río Chiriquí; obteniendo un porcentaje de error medio absoluto de **34,87%** para la estación Finca Lérída y **29,09%** para la estación David.

## 1.2. ABSTRACT

In this research, it is possible to create a statistical model that fits the monthly historical records of rainfall over the upper and lower zone of the Chiriqui river basin with the application of the Box - Jenkins methodology; therefore, the research is adjusted to a non-experimental longitudinal type study, and is considered to be longitudinal trend type with a retrospective approach over the months in the Chiriqui river basin already occurred. That is why the study will focus on meteorological stations that have complete records on a monthly basis with 44 annual records of the upper and lower zone of the Chiriqui river basin, located in the province of Chiriqui, Republic of Panama. In addition, two types of statistical analysis were carried out, corresponding to the descriptive and inferential analysis of the data; with the latter, a clear scientific perception of behavior is developed, such as the tendency and seasonality that predominates in rainfall over the Chiriqui river basin, from its past, present and future, thus achieving, benefiting and preventing possible disasters to the agricultural sector, livestock and energy, as well as hydroelectric, meteorological and other institutions.

The best structure that manages to model the rainfall variable was the Seasonal Autoregressive Integrated Moving Average (SARIMA).

The  $ARIMA(2, 1, 2)(2, 1, 1)_{12}$  model for the Finca Lerida station and  $ARIMA(2, 1, 1)(2, 1, 1)_{12}$  model for the David station; were used to predict the rainfall of the first quarter of 2016, in each of the different stations located in the Chiriqui river

basin; obtaining an absolute average error percentage of **34,87%** for the Finca Lerida station and **29,09%** for the David station.

## 2. MARCO INTRODUCTORIO

La posibilidad de preparar una intervención oportuna garantiza a las instituciones meteorológicas, agricultores y protección civil, la reducción de las consecuencias, en ocasiones letales, derivadas a los desbordamientos de los ríos y otros daños, al igual que la escasez de las lluvias. Por consiguiente, la predicción de series de tiempo en el mundo moderno se presenta como una gran alternativa que nos ayuda a diagnosticar la posible conducta presentada en los rendimientos de las series temporales, sirviendo de referencia para tomar una acertada decisión a la hora de enfrentarse a diferentes alternativas. Esta implementación de los modelos heteroscedásticos han sido más aplicados en el mercado financiero actual, ya que les ayudan a identificar con mayor certeza el posible comportamiento de un activo a través del tiempo (Marie, 2007).

La modelación ARIMA (Autorregresive Integrated Moving Average), desarrollada por George Box y Gwilym Jenkins en la década de los setenta del siglo XIX, con el propósito de obtener mejores pronósticos en el control de la contaminación de la bahía de San Francisco, constituyó una verdadera revolución en el análisis de series de tiempo, el libro “Time Series Analysis, Forecasting and Control” publicado en 1976, se convirtió rápidamente en un clásico, y sus procedimientos se utilizan ampliamente desde entonces en diferentes ramas de la ciencia, conociéndose como metodología Box-Jenkins (Molinero, 2004).

Como nos indica Mellado (2004),

La metodología Box-Jenkins consiste en extraer los movimientos predecibles de los datos observados. La serie de tiempo se descompone en varios componentes, algunas veces llamados “filtros”, precisamente porque la filosofía del método consiste en detectar los distintos componentes usando los filtros correspondientes, hasta obtener residuales no predecibles cuyo comportamiento tiene poca influencia en el resultado final (pp. 251-252).

Esta metodología ha sido utilizada en la salud pública con frecuencia creciente en todo el mundo debido a sus bondades para modelar adecuadamente el comportamiento de los eventos de salud. En un modelo ARIMA se considera que el comportamiento de la variable, en cualquier momento del tiempo, está influenciado por las observaciones de la propia variable (recientes o remotas), incorporadas a un modelo mediante los términos autorregresivos (AR) y los errores o influencia de los elementos aleatorios (recientes o remotos) que se representan con los términos de medias móviles (MA). Un modelo finalmente puede incluir cualquiera de esos parámetros, aislados o en diferentes combinaciones, además se pueden incluir como parámetros de los modelos las diferenciaciones realizadas para estabilizar la serie y esto garantiza el análisis más integral. Los modelos ARIMA son prácticamente los más exactos para la predicción disponibles hasta el momento y son especialmente adecuados para las series de periodicidad cortas (horas, días, semanas), las más utilizadas en la vigilancia y de las que se puede obtener una historia anterior relativamente larga (Marie, 2007, p. 2).

En general, las series de tiempo además del componente estacional, presentan otros componentes, como tendencia, aleatoriedad, por mencionar algunos, de los que, por

otro lado, el componente estacional no es independiente. Se trata de proponer un modelo *ARIMA* que recoja no solo las relaciones entre periodos (años); sino también las relaciones dentro de los periodos (intraanuales) y la interacción entre ambas estructuras.

Los modelos *ARIMA* estacionales multiplicativos,  $SARIMA(p, d, q)(P, D, Q)_S$  son flexibles en el sentido de que especifican estacionalidades estocásticas, tendencias estocásticas y además, recogen la posible interacción entre ambos componentes (González, 2007). Sin embargo, son poco los estudios que con el uso de las series de tiempo se apliquen a la meteorología, hidrología, entre otros; es decir al sector climático; y así, como el sector financiero y salud le ayuda a identificar y prever con mayor certeza los activos a través del tiempo, lo mismo podría aplicarse al sector meteorológico; permitiendo así, prever posibles desastres naturales.

La prolongada variabilidad de las precipitaciones pluviales a lo largo del tiempo se ve afectada directamente a las hidroeléctricas que repercute a la demanda de energía de manera dramática, pérdidas en el sector agrícola; y a su vez, inundaciones y desbordamientos de ríos causando desastres naturales, entre otras.

La nueva visión sobre el futuro del cambio climático resalta la necesidad de identificar y establecer unos objetivos renovados y frescos que permitan a las instalaciones meteorológicas, empresas de sector energético, agricultores, entre otros, adaptarse a los desafíos del nuevo entorno que se vive.

### **2.1. Planteamiento del problema y objetivos**

A través de los últimos años se han presentado una serie de sucesos meteorológicos que ponen de manifiesto el posible desequilibrio de la cuenca el río Chiriquí, que se ve

muy afectada por los fenómenos atmosféricos, produciendo múltiples cambios y altas precipitaciones pluviales atípicas que la convierten en un lugar de alto riesgo de inundación, por lo que se hace necesario el análisis de los datos de precipitación pluvial, utilizando modelos estadísticos para modelos autorregresivos de media móvil estacional multiplicativo (*SARIMA*) con la metodología Box - Jenkins, para obtener las posibles predicciones de variabilidad de precipitaciones en los meses venideros y así poder prevenir desastres, bajas en la producción agropecuaria y fluctuaciones en los afluentes del río Chiriquí, producto de la intensidad de precipitación pluvial sobre la cuenca.

## **2.2. El Problema:**

¿La metodología Box - Jenkins para series de tiempo, puede utilizarse como herramienta para estimar la precipitación pluvial de las estaciones pluviométricas ubicados en la cuenca del río Chiriquí a lo largo del tiempo?

## **2.3. Objetivo General:**

Crear un modelo estadístico que mejor se ajusten a los registros históricos mensuales de la precipitación pluvial sobre la zona alta y baja de la cuenca del río Chiriquí, utilizando la metodología Box - Jenkins.

## **2.4. Objetivos específicos:**

- ✓ Analizar descriptivamente el comportamiento de la precipitación pluvial en la cuenca del Río Chiriquí.
- ✓ Verificar la validez de los datos pluviométricos de las estaciones utilizadas.
- ✓ Aplicar la metodología Box - Jenkins a la serie de tiempo en estudio para cada estación meteorológica.

- ✓ Encontrar el modelo de pronóstico más adecuado para la estación meteorológica.
- ✓ Pronosticar la precipitación pluvial del primer cuatrimestre del año 2016, para las estaciones pluviométricas con un 95% de confianza.
- ✓ Comparar los datos pronosticados versus los datos reales del primer cuatrimestre de 2016.

### **2.5. Justificación:**

Hoy día, toda institución encargada de la meteorología necesita conocer y entender el comportamiento y cambios que sufre la precipitación pluvial en todas las cuencas existentes a nivel nacional que alimentan a nuestros ríos y afluentes.

Los estudios de precipitaciones pluviales permiten a las instituciones meteorológicas tener una mejor comprensión y análisis del comportamiento de la precipitación pluvial del área en estudio; y a la vez, ayuda a encontrar modelos que se ajusten a la realidad, ya que es de gran relevancia para las empresas hidroeléctricas, agricultores, ganaderos, entre otros el conocimiento de los futuros eventos de las precipitaciones pluviales, ya que al ellos estar informados de un posible aumento o disminución de la precipitación pluvial, podrán prepararse y así no afectar su producción para el beneficio propio y de la nación.

Es necesario que las instituciones meteorológicas utilicen más de un modelo estadístico y que realicen comparaciones para verificar y comprobar que los resultados que se obtengan sean los más confiables y los que se ajustan más a la realidad.

### **3. MARCO REFERENCIAL**

#### **3.1. Fundamentación teórica**

En la búsqueda de la delimitación de esta investigación, se da un pequeño pantallazo acerca de los conceptos que se procuran desarrollar y analizar en función del problema planteado. Por tanto, de los referentes investigados se puede mencionar como relevante las siguientes conceptualizaciones:

##### **3.1.1. ¿Qué se entiende por precipitación pluvial y cuenca?**

Según la Gerencia de Hidrometeorología de ETESA: “La migración estacional de las masas de aire tropical del Pacífico y sub-tropical del Atlántico que acompañan al sol en su curso anual constituyen el control dominante sobre los patrones de precipitación en Panamá. Estas migraciones en combinación con la orografía local, establecen regímenes de precipitación bien definidos.”

La distribución estacional del régimen pluviométrico de la cuenca del Río Chiriquí en la vertiente del Pacífico, se caracteriza por una estación lluviosa extendida, con una intensidad entre moderada a fuerte, acompañada de actividad eléctrica que ocurre especialmente en horas de la tarde; donde este factor, es determinado por la migración anual llamada Zona de Convergencia Intertropical (ZCIT), que es la zona de confluencia de los vientos alisios de ambos hemisferios, en dirección al Norte (mes de junio) y al Sur (mes de octubre).

De acuerdo a la UNESCO (2008),

La época lluviosa empieza a finales de abril o principios de mayo y persiste hasta mediados o fines de noviembre, siendo los meses de septiembre y octubre los más lluviosos; en algunas áreas de la cordillera la estación tiene una duración mayor. Dentro de esta temporada se presenta frecuentemente un período seco conocido como Veranillo, entre julio y agosto. Entre diciembre y finales de abril se establece, en esta vertiente, la estación seca con ausencia casi total de lluvia. Algunas veces, en este período ocurren temporales y lluvias copiosas, ocasionadas por incursiones de frentes fríos intensos que logran alcanzar nuestras latitudes y que son empujados por avances vigorosos de masas enormes de aire polar, procedentes de las regiones árticas heladas (p. 27).

El territorio de la República de Panamá tiene un área superficial continental e insular de 75 524 km<sup>2</sup>, y se ha dividido en 52 cuencas hidrográficas, las que resultaron de la clasificación elaborada durante el desarrollo del “Proyecto Hidrometeorológico Centroamericano (PHCA)” que culminó en el año 1972.

De estas cuencas, 18 están en la vertiente del mar Caribe y le corresponden números impares, comenzando desde el número 87 hasta la 121; y 34 pertenecen a la vertiente del océano Pacífico y le corresponden números pares, iniciando con la cuenca número 100 hasta la 166. En este proyecto participaron Guatemala, Honduras, Nicaragua, El Salvador, Costa Rica y Panamá. La cuenca N°1 se inició en la República de Guatemala y la última cuenca finalizó en l república de Panamá con el N°166.<sup>1</sup>

---

<sup>1</sup> Obtenido de: <http://www.hidromet.com.pa/cuencas.php>

Dentro de la provincia de Chiriquí tenemos la cuenca del Río Chiriquí, cuenca 108 que cuenta con una extensión de 1 905 km<sup>2</sup>, presentando una elevación máxima de 3 474 msnm y su cauce principal tiene una longitud de 130 km. Sus principales afluentes son los ríos David, Cochea, Caldera, Gualaca y el río Chiriquí.

**Ilustración 1: Mapa de cuencas hidrográficas en la República de Panamá**



Fuente: Gerencia de Hidrometeorología de la Empresa de Transmisión Eléctrica S.A. (ETESA)<sup>2</sup>

### 3.1.2. ¿Qué se entiende por modelo estadístico?

De acuerdo a Ocaña (2011), un modelo estadístico es una ecuación matemática que reproduce los fenómenos que observamos de la forma más exacta posible. Para ello tiene en cuenta los datos suministrados y la influencia que el azar tiene en estas observaciones.

<sup>2</sup> Obtenido de: [http://www.hidromet.com.pa/Mapas/Cuencas\\_Hidrograficas\\_Panama.pdf](http://www.hidromet.com.pa/Mapas/Cuencas_Hidrograficas_Panama.pdf)

Es importante entender que un modelo es una representación formal de un sistema real, con el que se pretende aumentar su comprensión, hacer predicciones y ayudar a su control. Los modelos pueden ser físicos (descritos por variables medibles), análogos (diagrama de flujo) y simbólicos (matemáticos, lingüísticos, esquemáticos). Los modelos matemáticos o cuantitativos son descritos por un conjunto de símbolos y relaciones lógico-matemáticas.

### 3.1.3. ¿De qué trata la Metodología Box - Jenkins?

Es la técnica estadística desarrollada por Box y Jenkins para ajustar modelos *ARIMA* a datos de series de tiempo. Los modelos *ARIMA* son modelos paramétricos que tratan de obtener la representación de la serie en términos de la interrelación temporal de sus elementos.

No obstante, para Mellado (2004):

El método de Box-Jenkins consiste en extraer los movimientos predecibles de los datos observados. La serie de tiempo se descompone en varios componentes, algunas veces llamados “filtros”, precisamente porque la filosofía del método consiste en detectar los distintos componentes usando los filtros correspondientes, hasta obtener residuales no predecibles cuyo comportamiento tiene poca influencia en el resultado final (pp. 251-252).

Una serie temporal es el registro histórico de algunas actividades, con mediciones hechas en intervalos iguales de tiempo, manteniendo una consistencia en la actividad y el método de medición.

La idea básica, detrás de los modelos de pronóstico de series de tiempo, es encontrar una fórmula matemática que genere de forma aproximada los patrones históricos de las series de tiempo.

A comienzo de los años 70, G.E.P. Box, profesor de Estadística de la Universidad de Wisconsin, y G.M. Jenkins, profesor de Ingeniería de Sistemas de la Universidad de Lancaster, introdujeron una pequeña revolución en el enfoque del análisis de series temporales, en su libro *Time series analysis. Forecasting and control* (Box y Jenkins, 1976) con el propósito de establecer mejores mecanismos de pronóstico y control para el modelaje de series temporales reales. El libro se convirtió rápidamente en un clásico, y sus procedimientos se utilizan ampliamente desde entonces en diferentes ramas de la ciencia, conociéndose como modelos *ARIMA* y también como Metodología Box-Jenkins". (Moliner, 2004)

Las técnicas de Box - Jenkins aplican métodos autorregresivos y de media móvil a los problemas de pronóstico de series de tiempo. Este tipo de modelos que caracterizan las series como sumas o diferencias, ponderadas o no, de variables aleatorias o de las series resultantes, fue propuesto por Yule y Slutsky en la década de los 20. Fueron la base de los procesos de medias móviles y autorregresivos que han tenido un desarrollo espectacular tras la publicación en 1970 del libro de Box - Jenkins sobre modelos **ARIMA**.

Las alternativas del modelo se identifican según los órdenes de los parámetros que se estiman, distinguiéndose:

- Autorregresivos **AR(p)**

- Media móvil  $MA(q)$
- Autorregresivo de media móvil  $ARMA(p, q)$
- Autorregresivo integrado de media móvil  $ARIMA(p, d, q)$

Autorregresivo integrado de media móvil estacional multiplicativo  $ARIMA(p, d, q)(P, D, Q)_S$ .

La gran aportación de Box-Jenkins fue proponer una metodología de modelización de las series, es decir, un conjunto de reglas para construir modelos que reproduzcan satisfactoriamente el comportamiento de series temporales reales.

#### Características de la metodología Box-Jenkins:

- Tiene solamente en cuenta la pauta de serie de tiempo en el pasado.
- Ignora la información de variables causales.
- Procedimiento técnicamente sofisticado de predicción de una variable.
- Utiliza la observación más reciente como valor inicial.
- Permite examinar el modelo más adecuado
- Analiza errores recientes de pronósticos para seleccionar el ajuste apropiado para periodos futuros.
- Extrae mucha información de la serie de tiempo, más que cualquier otro método.

#### 3.1.4. ¿Qué es una serie temporal?

Una serie de tiempo es una sucesión de variables aleatorias  $Y_t$  ( $t = 1, 2, 3, \dots, n$ ) que miden el estado de alguna actividad en particular a través del tiempo y cada valor

observado puede considerarse una muestra aleatoria de tamaño uno de la variable; además, el análisis de la serie temporal se enmarca en los *procesos estocásticos*.

El instrumento fundamental a la hora de analizar las propiedades de una serie temporal en términos de la interrelación temporal de sus observaciones es el denominado “*coeficiente de autocorrelación*” que mide la correlación, es decir, el grado de asociación lineal que existe entre observaciones,  $Y_a$  y  $Y_b$ , separadas por  $k$  retardos si  $a - b = k$ .

La función de autocorrelación de un proceso estocástico estacionario es una función de  $k$  que recoge el conjunto de los coeficientes de autocorrelación del proceso y se denota por  $\rho_k$ ;  $k = 0, 1, 2, 3, \dots$ . La función de autocorrelación se suele representar gráficamente por medio de un gráfico de barras denominado *correlograma*.

Para poder aplicar la metodología Box - Jenkins se debe convertir este proceso estocástico en un proceso estacionario. Un proceso estocástico es estacionario si:

- La media de  $Y_t$  es constante
- La varianza de  $Y_t$  es constante
- La correlación entre  $Y_t$  y  $Y_{t+k}$  depende únicamente del número de retardos  $k$ .

Para la estimación de los parámetros del modelo se utilizan las funciones de autocorrelogramas simple y parcial:

- *La autocorrelación simple* de orden  $k$  ( $\rho_k$ ): Es la representación de los coeficientes de autocorrelación en función de los retardos, es decir, es la correlación entre variables separadas por  $k$  retardos.

- La autocorrelación parcial de orden  $k$  ( $\alpha k$ ): Es la correlación parcial entre variables separadas por  $k$  retardos eliminando el efecto de las  $k - 1$  variables intermedias.

Para ajustar el modelo el proceso estocástico más sencillo es el denominado *Ruido Blanco* ( $a_t, t = 0, \pm 1, \pm 2, \dots$ ) que es una secuencia de variables aleatorias de media cero, varianzas constantes, covarianzas nulas, correlación de los  $k$  ( $\rho k$ ) es igual a cero y  $a_t$  sigue una distribución normal (González, 2007):

$$E(a_t) = 0, \forall t \quad V(a_t) = \sigma^2, \forall t \quad Cov(a_t a_s) = 0, \forall t \neq s$$

Así, un proceso ruido blanco,  $a_t \sim RB(0, \sigma^2)$ , es estacionario si la varianzas  $\sigma^2$  es finita con función de autocovarianzas,  $\gamma_k$ , (FASV):

$$\gamma_k = \sigma^2, k = 0 \quad \text{y} \quad \gamma_k = 0, k > 0$$

Y función de autocorrelación (FAS):

$$\gamma_k = 1, k = 0 \quad \text{y} \quad \gamma_k = 0, k > 0$$

#### 3.1.4.1. Modelo autorregresivo integrado de media móvil estacional multiplicativo

En general, las series de tiempo además del componente estacional, presentan otros componentes, como tendencia, aleatoriedad, por mencionar algunos, de los que, por otro lado, el componente estacional no es independiente. Se trata de proponer un modelo *ARIMA* que recoja no solo las relaciones entre periodos (años); sino también las relaciones dentro de los periodos (intraanuales) y la interacción entre ambas estructuras. Este modelo completo para la serie observada es:

$$\Phi_P(L^s)\phi_p(L)\Delta^d\Delta_s^D Y_t = \theta_q(L)\Theta_Q(L^s)a_t$$

donde  $\phi_p(L)$  y  $\theta_q(L)$  son los polinomios autorregresivos y medias móviles de la parte regular y  $d$  es el orden de integración de la parte regular; mientras que  $\Phi_P(L^s)$  y  $\Theta_Q(L^s)$  son los polinomios autorregresivos y medias móviles de la parte estacional  $s$  y  $D$  son el orden de integración de la parte estacional.

Los modelos *ARIMA* estacionales multiplicativos,  $ARIMA(p, d, q)(P, D, Q)_s$  son flexibles en el sentido de que especifican estacionalidades estocásticas, tendencias estocásticas y además, recogen la posible interacción entre ambos componentes.

Esta clase de modelos se basa en la hipótesis central de que la relación de dependencia estacional es la misma para todos los periodos. Este supuesto no se tiene por qué cumplir siempre; no obstante, estos modelos son capaces de representar muchos fenómenos estacionales que se encuentra en la práctica de una forma muy simple. Uno de los criterios que se ha de seguir al construir un modelo  $ARIMA(p, d, q)(P, D, Q)_s$  es el de parsimonia, es decir, buscar el modelo más sencillo y con el menor número de parámetros posible que reproduzca las características de la serie (González, 2007, p. 110).

## 3.2. Revisión literaria

### 3.2.1. Cuenca del río Chiriquí:

Se han realizado diferentes estudios en relación a la precipitación pluvial sobre la cuenca del río Chiriquí. Algunos de estos estudios son acerca de la Caracterización de la Hidrología en las sub-cuenca del río David (Mora, 2003) y Caracterización Hidrológica en

el área de Boquete cuenca alta del río Caldera (Ortíz, 2006); además, de la Actualización del diagnóstico de los Recursos Naturales Existentes en la cuenca del río Caldera (De Batista, 1999) y Análisis del Recurso Hídrico en la cuenca del río David (Olivares, 1994); entre otros.

En el estudio de Batista (1999) y Olivares (1994) se aborda el tema desde la perspectiva de los recursos naturales e hídricos existentes en la cuenca del río Caldera y cuenca del río David, donde involucra de una manera superficial la precipitación pluvial sobre la zona estudiada; sin embargo, los estudios de Mora (2003) y el de Ortiz (2006) hacen un análisis más detallado de la precipitación pluvial sobre la sub-cuenca del río David y la cuenca alta del río Caldera, y a la vez estudian otras variables que involucran a la hidrología que poseen esas sub-cuencas de la cuenca del río Chiriquí.

### **3.2.2. Metodología Box - Jenkins:**

Mucho de los estudios realizado con la metodología Box - Jenkins, son utilizados para la Economía (desempleo, PIB, otros) (UAM, 2004), Finanzas (rentabilidad de acciones, precio del petróleo, otros) (Gracia S., Vázquez C., Reyes O., Sáenz S., & Limón L., 2009), Ingeniería (demanda y precios de energía eléctrica, control de calidad)(Pulido, Roldán, López, & Gutiérrez, 2002) y (Murillos, Trejos, & Carvajal, 2003), entre otras como el sector salud (Marie, 2007) y (Martinez, Silva, & Fabbro, 2011).

Hasta la fecha, son pocos los trabajos que se encuentran en la literatura de proyecciones meteorológicas que utilizan modelos de series de tiempo con el enfoque de la metodología de Box - Jenkins. Sin embargo, algunas de esas literaturas que podemos citar son: Análisis de serie de tiempo de caudales del río El Tala periodo 1937-1960

(Verón, Herrera, & Rodríguez, n.d.); Modelo matemático para el pronóstico del clima con fines agrícolas en la cuenca del río Paria (Medinaceli, 2007) y Modelación de los escurrimientos del río Yaqui (Cruz M. & García P., 2008).

En el estudio de Verón y otros (n.d.), trata de analizar la serie temporal de los caudales del Río El Tala, en Argentina, entre 1936 y 1960 con el objetivo de entender y describir los procesos aleatorios que generan las observaciones, predecir valores futuros de la variable y el control óptimo de esta clase de procesos.

En cuanto al trabajo de Medinaceli (2007), trata de desarrollar un método que sirva como herramienta para conocer y cuantificar los valores futuros para cada uno de los factores climatológicos en la cuenca del río Paria, en Bolivia, basado en la metodología Box - Jenkins.

Sin embargo, en el trabajo de Cruz M. & García P. (2008), su objetivo se plantea en el obtener un modelo confiable para la predicción de las esorrentías del río Yaqui en el estado de Sonora, México; y para ellos realizan varios modelos entre ellos el ARIMA, regresión dinámica y funciones de probabilidad, clasificados como modelos de caja negra por Baratti et al. (2003; citado de Cruz M. & García P., 2008)

#### **4. ASPECTOS METODOLÓGICOS**

En este trabajo se estudia la cuenca del río Chiriquí, debido a que posee mayor cantidad de registros históricos (series más completas) de datos pluviométricos en toda la provincia de Chiriquí.

Cabe señalar que las dos series en estudio contienen registros de precipitación pluvial acumulada mensual por 44 años, que comienza en enero de 1972 y termina en diciembre de 2015 toda esta información fue facilitada por la Gerencia de Hidrometeorología de la Empresa de Transmisión Eléctrica S.A.(2009), teniendo toda esta información disponible es posible realizar los análisis de series de tiempo a los datos de precipitación pluvial.

#### 4.1. Diseño de la investigación

Se considera un análisis de los datos de tipo cuantitativo, haciendo uso de la estadística descriptiva, e inferencial con ayuda de diferentes métodos estadísticos que permita una mayor comprensión e interpretación de lo que arrojen los datos con respecto a la realidad, haciendo hincapié en el contraste de las evidencias tangibles u observables, para generar un mejor análisis e interpretación de lo que se encuentre.

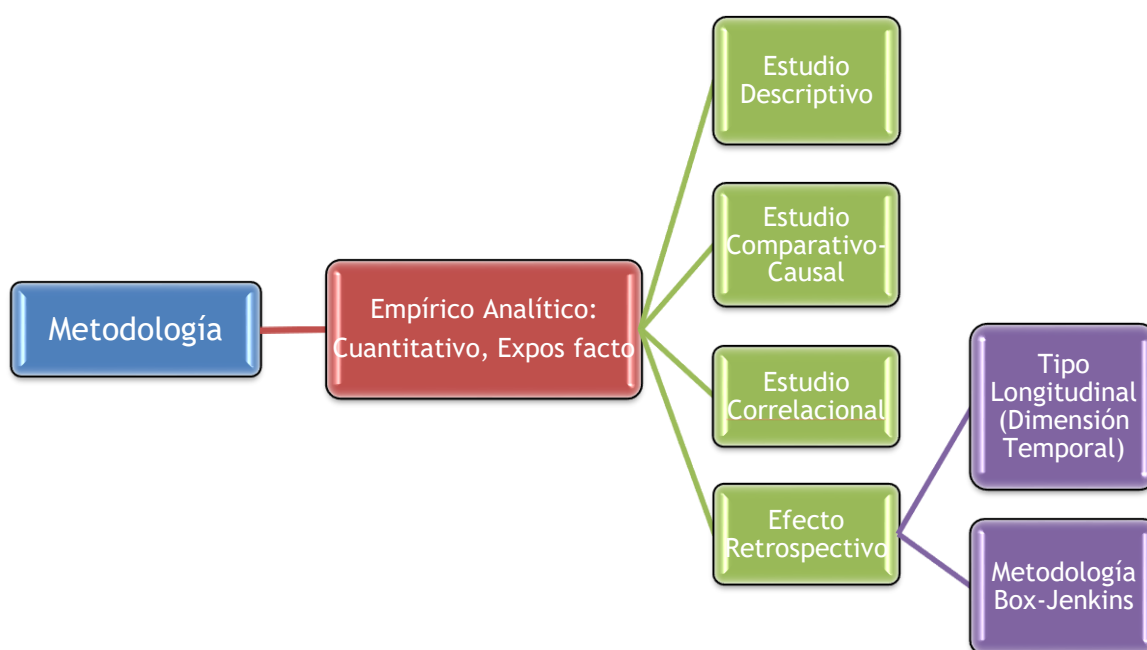
El diseño metodológico a utilizar en la investigación es de tipo:

- ***Empírico analítico:***

En el caso de la *metodología empírico analítica*, se asume un estudio de tipo cuantitativo; no obstante, la modalidad a utilizar sería de tipo *expost-facto* o no experimental, debido a que se observa el fenómeno de la precipitación pluvial tal y como se da en una serie de tiempo, en la cuenca del río Chiriquí, para después analizarlos.

Como señala Kerlinger (1979, p. 116; citado de Hernández-Sampieri, Fernández-Collado, & Baptista-Lucio, 2006, p. 205) “La investigación no experimental o *expos-facto* es cualquier investigación en la que resulta imposible manipular variables o asignar aleatoriamente a los sujetos o a las condiciones”.

O bien, como menciona Mateo y Albert (2004; 2016; citado de Navío, 2013, p. 29), la investigación no experimental no pretende la manipulación de la o de las variables independientes. Se observan situaciones ya existentes sin pretender el control de las variables.



En este tipo de investigación la variable independiente (serie temporal) ya ha ocurrido y no puede ser manipulada, no se tiene control directo sobre las variables ni influir sobre ella porque ya sucedieron al igual que sus efectos o variables dependientes (precipitación pluvial). Así pues, de las clasificaciones de los estudios no experimentales, sólo se consideran los siguientes:

- **Estudios descriptivos:** se realiza una investigación cuantitativa a las variables en estudio que trata de describir con precisión la distribución de probabilidad

con sus parámetros de centralización y parámetros dispersión, al igual que el comportamiento de la variable (precipitación pluvial).

- **Estudios comparativo-causales:** En este punto, se busca la relación entre ambas variables, pero sin la manipulación de estas. Es aquí donde se comparan los resultados arrojados de los pronósticos de la precipitación pluvial del primer cuatrimestre del año 2016 y los registros reales de la precipitación pluvial del primer cuatrimestre del año 2016.
- **Los estudios correlacionales,** es aquí donde se busca verificar la homogeneidad y consistencia de los registros pluviométricos entre las estaciones meteorológicas cercanas, a través de los métodos estadísticos, Pruebas de Rachas y Métodos de dobles acumuladas.
- **Efecto retrospectivo,** este tipo de diseño no experimental de acuerdo con la dimensión temporal, se le considera de tipo longitudinal de tendencia con un enfoque retrospectivo, debido a que se encarga de estudiar cómo evoluciona o cambia la acumulación de precipitación pluvial a lo largo de los meses en la cuenca del río Chiriquí ya ocurridos; y es aquí, donde se aplica la metodología Box - Jenkins para analizar, modelar y pronosticar la precipitación pluvial sobre la cuenta del río Chiriquí.

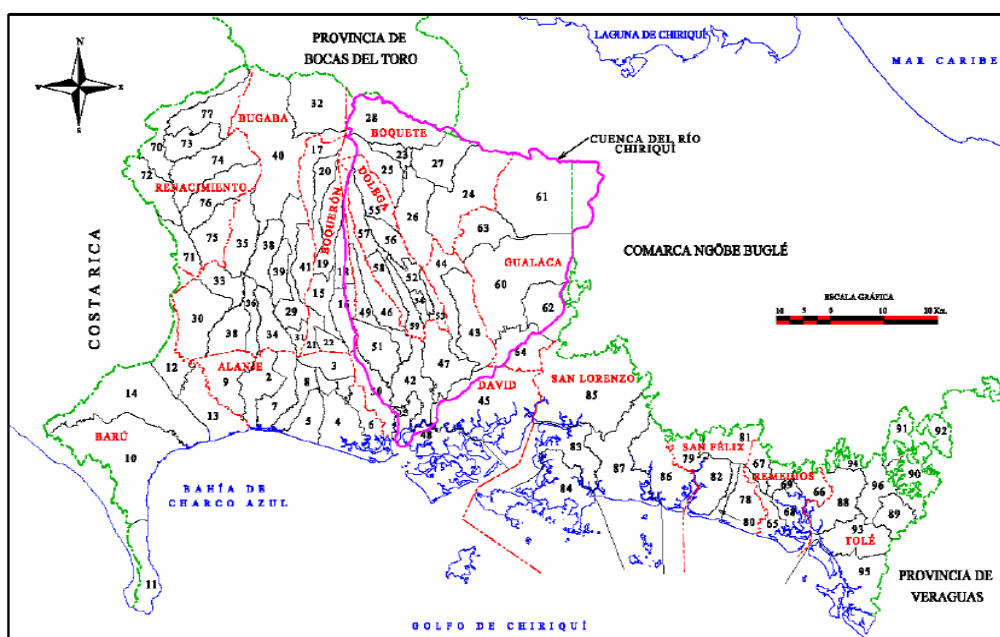
## 4.2. Metodología de la investigación

### 4.2.1. Contexto, población - muestra

Se considera la cuenca del río Chiriquí, que forma parte de la cuenca 108 según la clasificación de cuencas de la República de Panamá. Esta se encuentra localizada en la vertiente del Pacífico y al occidente de la República de Panamá, en la provincia de

Chiriquí. Sus coordenadas geográficas son: 8° 15' y 8° 53' de latitud norte, y los 82° 10' y 82° 33' de longitud oeste. Cuenta con una extensión de 1 995,9 km<sup>2</sup>, presentando una elevación máxima de 3 474 msnm y la longitud del río principal es de 130 km.

**Ilustración 2: División política administrativa de la cuenca del río Chiriquí por distrito y corregimiento, año 2003**



**Fuente:** Diseño de la red de monitoreo de la calidad del agua en las cuencas hidrográficas del río Caimito y el río Chiriquí.<sup>3</sup>

La cuenca hidrográfica del río Chiriquí, en el aspecto político administrativo, limita al norte con la provincia de Bocas del Toro y parte de la comarca Ngöbe-Buglé; al sur con el Océano Pacífico; al este con la comarca Ngöbe-Buglé, y el distrito de San Lorenzo y parte del distrito de David; al oeste con los distritos de Bugaba, Alanje y parte de Boquerón.

<sup>3</sup> Disponible en:  
[www.anam.gob.pa/...calidad/diagnostico\\_ambiental\\_%20de\\_la\\_cuenca\\_del\\_rio\\_chiriqui.pdf](http://www.anam.gob.pa/...calidad/diagnostico_ambiental_%20de_la_cuenca_del_rio_chiriqui.pdf)

**a. Población:**

En esta zona existen 19 estaciones meteorológicas. Cuatro (4) de las estaciones han sido instaladas desde el 1990 y de las quince (15) estaciones restantes han estado recolectando datos meteorológicos con más de 40 registros anuales; sólo cuatro (4) de las quince estaciones, poseen los registros completos de manera mensual, de las cuales son la estación: Finca Lérida, Los Naranjos, Veladero Gualaca y David.

**b. Muestra:**

La muestra en este tipo de investigación corresponde a una muestra dirigida de sujetos (no probabilísticas), la cual serían las estaciones meteorológicas y serán estratificadas en dos tipos de zonas (alta y baja) que será regida por la elevación en las que se encuentre.

**4.2.2. Criterios de selección:**

De las 19 estaciones meteorológicas solo se tomaron las estaciones con los 44 registros mensuales completos, para minimizar diferentes errores de los análisis estadísticos a lo largo de la investigación; siendo únicamente, dos estaciones en estudio: Finca Lérida, y David.

***Tabla 1: Estaciones meteorológicas de estudio***

Código	Estación	Tipo de estación	Elevación (msnm)	Zona
108-001	Finca Lérida	PV	1 700	Alta
108-023	David	A	27	Baja

Para la zona alta, estarán comprendidas con una elevación mayor de 500 msnm; y para la zona baja esta será menor a 500 msnm. Se estratificó de esta manera, por la variabilidad de las precipitaciones sobre la cuenca del río Chiriquí para obtener un mejor análisis de la información.

#### 4.2.3. Procedimiento utilizado:

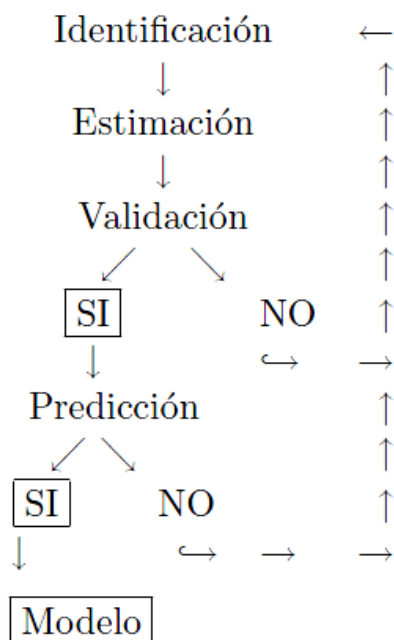
La metodología Box-Jenkins propone una metodología de modelización de las series, es decir, un conjunto de reglas para construir modelos que reproduzcan satisfactoriamente el comportamiento de serie temporales reales. Las etapas son las siguientes:

- 1) Identificar el modelo tentativo a desarrollar, se determina la forma general del modelo que supuestamente es más adecuado para los datos observados.
- 2) Estimar parámetros en el modelo tentativo a desarrollar denominado parsimonia o principio de parametrización escueta. Se trata de proponer un modelo capaz de representar la serie con el mínimo de parámetros posibles y únicamente acudir a una ampliación del mismo en caso de que sea estrictamente necesario para describir el comportamiento de la serie.
- 3) Validación o Diagnóstico la adecuación general del modelo, es decir, la transformación estacionaria para elegir el proceso estacionario  $AR(p)$ ,  $MA(q)$ ,  $ARMA(p, q)$ ,  $ARIMA(p, d, q)$ ,  $ARIMA(p, d, q)(P, D, Q)_S$ .
- 4) Predicción o Pronóstico de realizaciones futuras.

Antes de usar el modelo para pronosticar, el analista debe verificar si el modelo es adecuado o no. En esta etapa se comprueba que la especificación propuesta, realmente

recoja el patrón sistemático de comportamiento de los datos en el sentido de que no quede información en los residuos del modelo que se pueda explotar para mejorar la formulación propuesta. Si la especificación propuesta supera la validación se concluye. Pero en caso contrario se debe volver a la primera etapa y proponer una especificación alternativa. Este proceso iterativo<sup>4</sup>, de prueba y error hasta llegar al modelo final, constituye uno de los aspectos más novedosos de la metodología Box - Jenkins.

**Diagrama o esquema de la metodología Box - Jenkins:**



Fuente de González (2007, p. 51)

**4.3. Hipótesis de Trabajo:**

Existe un modelo estadístico que se ajusten a los registros históricos mensuales de la precipitación pluvial sobre la zona alta y baja de la cuenca del río Chiriquí.

<sup>4</sup> Acción de repetir una serie de pasos un cierto número de veces, para la resolución de problemas.

#### 4.4. Variables:

Definición Conceptual	Definición Operacional
<p><b>Precipitación Pluvial:</b></p> <ul style="list-style-type: none"> <li>• Para la UNESCO (1982), en su obra, Guía metodológica para la elaboración del balance hídrico de América del Sur: “Se denomina precipitación al agua que proviene de la humedad atmosférica y cae a la superficie terrestre, principalmente en estado líquido, (lluvia) o sólido (nieve o granizo)”</li> <li>• Para la Gerencia de Hidrometeorología de la Empresa de Transmisión Eléctrica S.A. (ETESA): “Precipitación es el término general que define el agua procedente de la atmósfera que cae sobre la superficie del globo terrestre, en forma de lluvia, nieve o granizo... En los trópicos, la precipitación atmosférica consiste casi en su totalidad de lluvia y constituye el elemento climático más variable de todos”</li> </ul>	<p>El instrumento para medir la precipitación se llama pluviómetro. Su componente es un recipiente que acumula agua, hasta que es vaciada y medida, ya sea en forma manual o automática.</p> <p>Con la medición de la precipitación se calcula la tasa de acumulación de lluvia, por unidad de área horizontal. Una acumulación de 1 mm corresponde a la altura que alcanza el volumen de agua en un área de un metro cuadrado.</p> <p>Sin embargo, al ser un estudio de tipo no experimental no se utiliza el pluviómetro, sino que se verifica y valida los datos pluviométricos que se utilizarán con estadísticas inferenciales.</p>
<p><b>Serie temporal</b></p> <ul style="list-style-type: none"> <li>• Las series de tiempo son secuencias de muestras de una variable tomadas típicamente en instantes sucesivos y espaciados uniformemente (López S. &amp; Martínez A., 2013, p. 2).</li> </ul>	<p>Se analiza la serie temporal mensualmente desde enero del año 1972 hasta diciembre de 2016.</p>

- |  |  |
|--|--|
| <ul style="list-style-type: none"> <li>• Es una secuencia ordenada de observaciones cada una de las cuales está asociada a un momento de tiempo (González, 2007, p. 1).</li> </ul> |  |
|--|--|

#### 4.4.1. Análisis estadístico:

En este trabajo se realiza dos tipos de análisis estadísticos, que corresponden al análisis descriptivo y el análisis inferencial de los datos.

La estadística descriptiva permite observar y conocer, a través de los datos de las estaciones meteorológicas, los patrones y el comportamiento en el que ha estado evolucionando la precipitación pluvial sobre la cuenca del río Chiriquí a medida que transcurre el tiempo.

La estadística Inferencial desarrollará una clara percepción científica del comportamiento, como la tendencia y estacionalidad que predomina en la precipitación pluvial sobre la cuenca del río Chiriquí, desde su pasado, presente y futuro.

##### 4.4.1.1. Análisis descriptivo:

La elaboración de tablas, cuadros y gráficas, permitirá presentar la información, suministrada en una forma más clara y organizada para una mejor interpretación de los resultados. Este análisis inicial también nos proporcionará una idea de la forma que tienen los datos: su posible distribución de probabilidad con sus parámetros de centralización; media, mediana y moda; así como sus parámetros de dispersión; varianza, desviación típica y coeficiente de variación.

##### 4.4.1.1.1. Precisión en la estimación de la precipitación pluvial:

Las precipitaciones son las únicas fuentes de humedad que llegan a la superficie

terrestre, la exactitud en su medición y su evaluación es determinante en el resultado de esta investigación. Sin embargo, a pesar de su importancia no se puede estimar con un alto grado de exactitud, ya que su evaluación está afectada por dos tipos de errores: error en la medida puntual y error en la evaluación espacial de la precipitación.

En el conocimiento de la precisión en la estimación de los datos obtenidos de las estaciones pluviométricas, se utilizará el coeficiente de variación aplicado por Oster (1979) en Colombia y en Chiriquí (1981), cuyo cociente no debe exceder del 0,25 para zona montañosa y 0,20 para zonas bajas.

El coeficiente de variación (*C.V.*), es el cociente entre la desviación típica ( $\sigma$ ) y la media o promedio ( $\bar{y}$ ) de la distribución:

$$C.V. = \frac{\sigma}{\bar{y}}$$

Es una medida de dispersión útil para comparar dispersiones a escalas distintas, pues es una medida invariante ante cambios de escala. Sirve para comparar variables que están a distintas escalas, pero que se mantienen correlacionadas estadísticamente y sustantivamente con un factor en común. El estudio de este cociente nos permite conocer las estaciones meteorológicas confiables de las que no lo son.

#### 4.4.1.2. Análisis inferencial:

Los datos serán validados con diferentes “*test (prueba) de significancia estadística*” con la que se prueba científicamente, tanto la validez del modelo como la consistencia de la información y el control de calidad de los datos con las pruebas de hipótesis para la homogeneidad de las series en estudio.

El método más relevante que se utilizará en este trabajo investigativo es la *metodología Box - Jenkins*; que permitirá modelar la precipitación pluvial sobre la cuenca del Río Chiriquí, y a su vez pronosticar los meses venideros y compararlos con los datos (meses) reales del mismo año.

#### **4.4.1.2.1. Análisis de consistencia de los registros pluviométricos:**

La Organización Meteorológica Mundial (OMM) recomienda, que los elementos climáticos deben tener un récord de 30 años para obtener normas representativas del lugar.

Para realizar la prueba de consistencia, las series tienen que estar ordenadas cronológicamente; estas pruebas se realizan con las pruebas de aleatoriedad, para comprobar que la serie es consistente y que no tiene tendencia, ni oscilación. Las pruebas de aleatoriedad más utilizadas son:

- Prueba de Rachas
- Prueba de Helmert
- Prueba de Abbe
- Dobles Acumuladas

De las cuales se utilizará la prueba de Rachas para cada una de las estaciones en estudio y se rectificará la consistencia de los datos de las estaciones aplicando el método de Dobles Acumuladas o Curva de Doble Masa, planteado por UNESCO (1982), Mora (2003), Ortíz (2006) y UNESCO (2008), tanto para las zonas alta y baja de la cuenca en estudio.

La homogeneidad o consistencia de una serie se puede modificar debido a varias razones, entre los más comunes están; cambio de instrumentos, modificación del emplazamiento de la estación, cambio del entorno de la estación, cambio de las horas de observación.

#### 4.4.1.2.2. La Prueba de Rachas

Comprende las siguientes etapas:

- Estimación de la mediana ( $Me$ ) de la serie.
- Cálculo de los desvíos de cada elemento respecto a la mediana. Se asigna a cada valor de la serie anual el signo correspondiente; (+) si el valor observado de la serie está por encima de la mediana y (–) si está por debajo.

Se llama Rachas a cada grupo de “+” o “–” y se simboliza con la letra  $r$ . El número total de racha señala si la muestra es o no aleatoria.

- ❖ La escasez de rachas puede deberse a tendencias temporales o a cierta falta de independencia
- ❖ La abundancia de rachas puede deberse a fluctuaciones cíclicas (no sería un evento aleatorio)

La Hipótesis de la prueba sería:

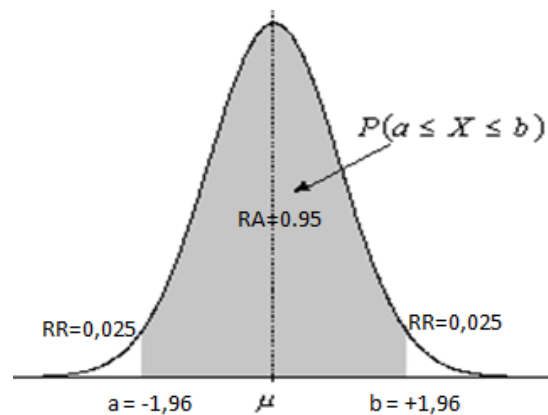
$H_0 : Me_1 - Me_2 = 0$ ; Los signos más o menos se distribuyen al azar.

$H_1 : Me_1 - Me_2 \neq 0$  Los signos más o menos no se distribuyen al azar.

- Prueba Estadística: En vista que la hipótesis concierne a la aleatoriedad de una serie de observaciones se escoge la prueba de rachas de una muestra.

- Nivel de Significación: Como la distribución es simétrica se toma como puntos críticos los de las dos colas, esto es la cola inferior  $\frac{\alpha}{2} = 0,025$  y la  $1 - \frac{\alpha}{2} = 0,975$  (siendo  $\alpha$  el nivel de significancia de  $\alpha = 0,05$ ) que se corresponde con  $-1,96$  y  $+1,96$  como los puntos críticos para el rango de aceptación (RA) con un 95% de confianza.

**Ilustración 3: Rango de aceptación del 95% para la prueba de Rachas**



- Cálculo de Significación (p-valor): Para muestras pequeñas ( $n_1$  ó  $n_2 \leq 20$ ) se utiliza la tabla de Rachas y para muestras mayores se utiliza la siguiente fórmula:

$$z = \frac{r - \mu_r}{\sigma_r} = \frac{r - \left( \frac{2n_1n_2}{n_1 + n_2} + 1 \right)}{\sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)(n_1 + n_2 - 1)}}}$$

#### 4.4.1.2.3. Método de Dobles Acumuladas o curva de doble masa

Este método se basa en la representación de ejes coordenados de parejas de puntos, definidos por las acumulaciones sucesivas de dos series de valores (correspondientes a dos estaciones cercanas en un radio de 30 km) en un mismo período temporal.

La Organización Meteorológica Mundial (OMM) dice que: “Una estación cercana es aquella que su radio no exceda una distancia de 30 km en superficies planas”. Se refiere a que la orografía del terreno no afecte, o no se interrumpa entre una estación y la otra (Castro F., 2008).

Se considerará homogéneo o consistente los valores de las dos series (estaciones), si la curva resultante es una línea recta. La pendiente de la recta representa la proporcionalidad entre las dos series de valores.

Si la serie de puntos encaja perfectamente en una recta que pasa por el origen, indicará que los valores de las estaciones en todos y cada uno de los años son aceptables. En el caso de que los datos encajen perfectamente en una recta que no pase por el origen, existirá proporcionalidad entre las dos estaciones todos los años, excepto el primero.

Cuando los puntos se ajusten a dos rectas paralelas, indicará una proporcionalidad entre las dos estaciones, aunque existan años que estén medidos por defecto o por exceso o que sean muy irregulares en una de las estaciones.

#### **4.4.1.3. Metodología Box - Jenkins**

El análisis de la serie temporal se realizará haciendo uso de los modelos ARIMA aplicando la metodología Box - Jenkins. Para obtener el modelo de series temporales se seguirán los siguientes pasos (Casado de Lucas, 2006):

- 1) Ingreso de datos y gráfica de la serie de tiempo
- 2) Transformaciones
- 3) Eliminación de la tendencia

- 4) Identificación del modelo
- 5) Estimación de los coeficientes del modelo
- 6) Contraste de validez del modelo
- 7) Análisis estadístico del error
- 8) Selección del modelo y predicción

#### 4.4.1.3.1. Gráfica de la Serie de tiempo

Se incorporan los datos disponibles (en este caso mensualmente) y con el graficador STATISTICA v12 se representa la serie histórica a fin de decidir la estacionalidad de la serie.

#### 4.4.1.3.2. Transformaciones de la serie (Estacionariedad en varianza)

Una serie es estacionaria en varianza, cuando pueda mantenerse el supuesto de que existe una única varianza para toda la serie temporal, es decir, cuando la variabilidad de la serie en torno a su media se mantenga constante a lo largo del tiempo. Si la serie es heterocedástica, o sea, que no tiene varianza constante, habrá que transformar la serie. El objetivo es conseguir alguna función que transforme la serie de forma que  $Y_t$  tenga varianza constante.

Con tomar el logaritmo natural es suficiente en nuestro caso, aunque existen algunas transformaciones más complejas, como las de Box - Cox (sugeridas por Box y Jenkins en su libro: Time Series Analysis. Forecasting and Control):

$$Y_t^{(\lambda)} = \begin{cases} \frac{Y_t^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln(Y_t) & \lambda = 0 \end{cases}$$

Donde  $\lambda$  es el parámetro de transformación. Es interesante señalar que, usualmente, las transformaciones Box-Cox no sólo estabilizan la varianza sino que también mejoran la aproximación a la distribución normal del proceso  $Y_t$  y es imprescindible en caso de que la varianza no sea constante (González, 2007, p. 48).

#### 4.4.1.3.3. Eliminación de la tendencia (Estacionariedad en media)

Se ha de identificar si la serie es estacionaria en media, es decir, si oscila en torno a un nivel constante o no.

Características de una serie estacionaria:

- Una serie es estacionaria en media cuando fluctúa alrededor de una media constante; en otras palabras, la serie estacionaria en media debe no presentar tendencia o varios tramos con medias diferentes. Para esto, se verifica con el gráfico original y las transformaciones correspondientes para observar si se cumple o no la condición de estacionariedad.
- La función de autocorrelación teórica de un proceso estacionario en media decae rápidamente hacia cero: Observando la serie, los correlogramas de las funciones de autocorrelación simple (FAS) y de autocorrelación parcial (FAP) que corresponden a los procesos autorregresivos y de medias móviles de las componentes regular y estacional, indicará si existe o no tendencia:
  - ✓ Si tiene tendencia: se toman las diferencias regulares hasta que desaparezcan. Normalmente el orden de la diferencia será 1, y raramente será mayor a 3.

- ✓ Si es estacional: se toman las diferencias estacionales hasta que desaparezca el patrón que se repite. En la práctica es muy raro tener que aplicar más de una diferencia estacional.
- Los contrastes de raíces unitarias proporcionan unos contrastes estadísticos que permiten, a partir del conjunto de información, hacer inferencia sobre la existencia o no de una raíz unitaria en una serie, es decir, sobre la no estacionariedad de la serie. Si se rechaza la hipótesis nula de existencia de raíz unitaria en  $\Delta^{d-1}Y_t$ , no se diferenciará más la serie; de lo contrario, se realiza una diferenciación más de orden 1. Para esta prueba se utilizará el contraste de Dickey-Fuller aumentado (ADF).

#### 4.4.1.3.4. Identificación del modelo

Una vez que el gráfico de la nueva serie (una transformación de la original) indica que es estacionaria, podemos intentar deducir su estructura. Para identificar los órdenes  $p$  y  $q$ , se compararán las funciones de autocorrelación muestrales con las FAS y FAP teóricas de los modelos *ARMA* cuyas características son:

- a) El decrecimiento de la altura de las barras (son los coeficientes de la función) debe ser exponencial. También suelen aparecer una alternancia de los signos o una forma sinusoidal; la alternancia depende de los signos de los parámetros, y la forma sinusoidal es debido a que hay raíces complejas en la ecuación característica del proceso. Si no hay decrecimiento exponencial, suele deberse a que la serie es integrada; para éstas el decrecimiento es lineal, no exponencial.
- b) Una de las dos funciones de autocorrelación va a mostrar una estructura como la descrita en a), mientras que la otra mostrará algunas barras que salen de la

banda de significatividad. La posición de la última de estas barras, que suele coincidir con el número de barras que sobresalen, nos dicen el orden del proceso. Según sea la FAS la que presenta la estructura de a) o sea la FAP, se dice que el proceso es *AR* o *MA*, respectivamente.

- c) Hay situaciones que no son muy claras, en estos casos puede tratarse de un modelo con las dos partes, la *AR* y la *MA*. Estos son los modelos *ARMA(p, q)*. Sus funciones FAS y FAP son combinaciones de ambas partes, por lo que son difíciles de identificar a simple vista. (ver Tabla 2)

El modelo *ARMA(p, q)* tiene media cero, varianza constante y finita y una función de autocovarianzas infinita. La función de autocorrelación es infinita decreciendo rápidamente hacia cero, pero sin truncarse (anularse).

**Tabla 2: Identificación de modelos**

Modelo	FAS	FAP
<i>MA (q)</i>	Se anula	Decrecimiento rápido
<i>AR (p)</i>	Decrecimiento rápido No se anula	Se anula
<i>ARMA (p, q)</i>	Decrecimiento rápido No se anula	Decrecimiento rápido No se anula

#### 4.4.1.3.5. Estimación de los coeficientes del modelo

En este punto, se observan las gráficas de los acápites anteriores para así tener una idea de qué modelo subyace a la serie, o al menos de cuáles serían los primeros candidatos que se deben probar. Para comprobar analíticamente (no visualmente), se intenta ajustar un modelo  $ARIMA(p, d, q)(P, D, Q)_S$  tomando en cuenta los coeficientes de correlación de los parámetros y la importancia que tienen todos sus términos, luego

de ver y analizar todo lo expresado anteriormente se decide por uno o varios modelos a validar y se realizan todos los pasos presentados y detallados anteriormente.

Frecuentemente se prueban varios modelos  $ARIMA(p, d, q)(P, D, Q)_S$  hasta que se encuentre el modelo que sea más convincente; siendo:

- El valor de « $d$ » es el número de diferencias regulares que se debe tomar en «1»
- El valor de « $D$ » es el número de diferencias estacionales para hacer la serie estacionaria,
- « $p$ » es el orden de la parte  $AR$ , autorregresivo de la parte regular,
- « $q$ » el de la parte  $MA$ , media móvil de la parte regular,
- « $P$ » es el orden de la parte  $AR$ , autorregresivo de la parte estacional y
- « $Q$ » el de la parte  $MA$ , media móvil de la parte estacional.

#### 4.4.1.3.6. Contrastes de validez del modelo

Se usan procedimientos para evaluar el o los modelos preseleccionados; para ello hay que tener en cuenta que:

- a) La FAS y la FAP de un Ruido Blanco (serie de datos procedentes de variables sin ninguna estructura común, es decir, independientes entre sí) tienen todos los coeficientes nulos en teoría, o no significativos en nuestras gráficas. Respecto a la banda de significatividad, se advierte en el libro “Análisis de series temporales”, de Daniel Peña, que: “como....estos límites sobrestiman la varianza en los retardos pequeños, un valor próximo a los límites de confianza  $\pm \frac{2}{\sqrt{n}}$  en los retardos iniciales debe considerarse un indicio claro de que el modelo es inadecuado” ( $n$  es la longitud de la serie).

- b) Además, las líneas de significatividad que aparecen en las gráficas de las funciones de autocorrelación suelen determinar una zona de confianza del 95%, es decir, el coeficiente que sobresale se puede considerar no nulo con un 95% de seguridad. Esto implica que puede haber un 5% de barras que sobresalgan sin motivo real (Wikipedia, 2017). Se debe pensar esto cuando hay alguna barra aislada que no cuadra con el resto de la estructura de barras. Debido a los estimadores utilizados para esas barras/coeficientes, y para garantizar que las estimaciones se basan en un número suficiente de datos, el programa elige el número de barras que representa.

Cuando se cumplan las condiciones que se exigen en los siguientes análisis, se validará el modelo de la siguiente manera:

- c) Las condiciones en el primero son que, la distribución de los residuos debe ser normal: los residuos observados y esperados bajo la hipótesis de distribución normal deben ser parecidos. Esta suposición se comprueba con el gráfico de probabilidad normal, que permite comparar gráficamente la función de distribución observada en la muestra tipificada, con la función de distribución normal. Si la distribución de los residuos fuera normal, dichos valores deberán ser aproximadamente iguales y en consecuencia, los puntos del gráfico estarán situados sobre la recta que pasa por el origen con pendiente igual a 1.
- d) Para cada coeficiente del modelo propuesto, el valor absoluto del estadístico «t» debe ser menor que 1,96 y mayor que -1,96, como puntos críticos; a la vez que el p-valor de ese contraste (que mide si se puede considerar cero o no) sea mayor que 0,05. Los coeficientes que no cumplan estas dos condiciones no son

necesarios en el modelo, porque no hay evidencia de que deban ser no nulos. Esto permite la búsqueda de un modelo lo más sencillo posible.

- e) Otro recurso (gráfico) es el de observar, tras ajustar el modelo *ARIMA*, las funciones FAS y FAP de los residuos, lo que permitirá una información análoga a la del contraste de aleatoriedad de Ljung-Box (sugerido por la metodología Box - Jenkins), pero de forma gráfica.
- f) El otro análisis nos proporciona varios contrastes de hipótesis permitiendo una significancia estadística, esto quiere decir que es muy probable que algo sea cierto y no por azar:
- $H_0 : \rho_k = 0$ ; no existe autocorrelación serial en los residuales. Los datos son al azar (es decir, la correlación en la población de donde se toma la muestra es aproximadamente cero, de modo que cualquier correlación observada en el resultado de los datos de la aleatoriedad del proceso de toma de muestras).
  - $H_1 : \rho_k \neq 0$ ; existe autocorrelación serial en los residuales. Los datos no son aleatorios.

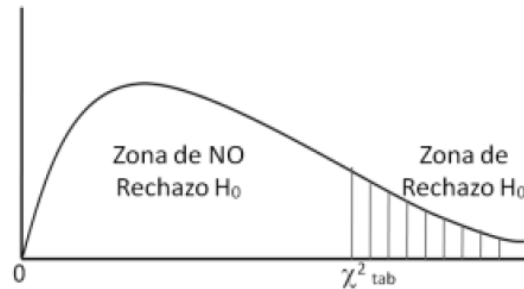
El estadístico más utilizado para contrastar esta hipótesis es el propuesto por Ljung - Box (1978), aplicado a los residuos después de ajustar el modelo:

$$Q_{(M)} = n(n + 2) \sum_{k=1}^M \frac{\rho_k^2}{n - k}$$

Donde  $n$  es el tamaño de la muestra,  $\rho_k$  es la autocorrelación en el retraso de la muestra  $k$ , y  $M$  son el número de retardo (rezagos) que se está probando. También, el  $Q_{(M)}$  se puede obtener del FAS residual del modelo que genera el programa STATISTICA.

Por nivel de significación  $\alpha$ , la región crítica para el rechazo de la hipótesis de aleatoriedad es:

**Ilustración 4: Zona de rechazo de la distribución Chi - cuadrado**



Fuente: (Pérez-Tejada, 2008, p. 500)

“Si  $Q_{(M)} < \chi_{\alpha,h}^2$ , no se rechaza la  $H_0$ , aceptándose así el  $H_0$ ”.

Donde  $\chi_{\alpha,h}^2$  es el  $\alpha$  – *cuantil* de ( $\alpha = 0,05$ ) la distribución chi-cuadrado<sup>5</sup> con grados de libertad,  $h = M - d - p$ . Siendo  $d$ , la cantidad de diferenciados y  $p$ , los parámetros utilizados en el modelo.

La prueba de Ljung-Box se utiliza comúnmente en autorregresivo integrado de media móvil (*ARIMA*) de modelado. Esta prueba permite una mayor evidencia de que los residuos son Ruido Blanco.

El programa STATISTICA v12 permite verificar cada una de las condiciones, detalladas anteriormente, para el contraste de validez del modelo.

#### 4.4.1.3.7. Análisis estadístico del error

Las diferencias entre los valores observados y los estimados por el modelo son una fuente de interés para valorarlo. Algunas de las estadísticas de error utilizados son:

<sup>5</sup> Ver **Anexo 11**: Distribución chi – cuadrado  $\chi^2$

**Tabla 3: Estadísticos de error**

Estadísticas de error	Indicadores	Propósito
Error cuadrado medio	ECM	Técnica para evaluar una técnica de pronóstico. Cada error o residual se eleva al cuadrado; luego estos valores se suman y se divide entre el número de observaciones. Este enfoque penaliza los errores mayores de pronósticos, ya que eleva cada uno al cuadrado. Esto es importante pues en ocasiones pudiera ser preferible una técnica que produzca errores moderados a otra que por lo regular tenga errores pequeños, pero que ocasionalmente arroja algunos en extremo grandes.
Raíz cuadrada del error medio	RCEM	La medida más frecuente adoptada es la RCEM. Es preferible el RCEM, al error medio cuadrático ya que el RCEM tiene las mismas unidades de medición que las series reportadas, y de esta manera es más cuadrática; en otras palabras, la pérdida asociada con un error aumenta en proporción con el cuadrado del error.
Porcentaje de error medio absoluto	PEMA	Este enfoque es útil cuando el tamaño o magnitud de la variable de pronóstico es importante en la evaluación de la precisión del pronóstico. El PEMA proporciona una indicación de que tan grandes son los errores de pronóstico comparados con los valores reales de la serie.

#### 4.4.1.3.8. Selección del modelo y Predicción

No hay un único modelo correcto (que explica suficientemente bien los datos), de la misma manera que no hay una única función que pase por varios puntos que tracemos en un sistema de ejes X-Y, algunos criterios que se pueden seguir para elegir un modelo satisfactorio son:

- ✓ Se debe verificar las condiciones gráficas y numéricas necesarias.
- ✓ Sus residuos después de ajustar el modelo deben ser Ruido Blanco.
- ✓ Entre dos modelos que sólo difieran en el p-valor del contraste de Box-Pierce, se prefiere el que tenga un valor mayor, puesto que da más garantías de que los residuos que proporciona el modelo son ruido. Entre las que verifican las mismas condiciones numéricas más o menos, queremos tomar el más sencillo posible, es decir, el que necesite menos parámetros. Los parámetros que necesita un modelo  $ARMA(p, q)$  son « $p + q$ ».
- ✓ Cuando el orden de una parte,  $AR$  ó  $MA$ , es muy alto, podemos probar a añadir algún término de la otra parte y restarle orden a ésta.

Al final, la elección del modelo final debe ser un equilibrio entre simplicidad y eficacia, porque siempre sería posible añadir parámetros indefinidamente para obtener mejoras; y así usarlo como modelo para predecir el primer cuatrimestre del año 2016.

#### 4.4.2. Instrumento:

Para el procesamiento de la información:

- Paquete de Microsoft Office 2016: Word y Excel

Para el análisis estadístico:

- ✓ HIDROBAS versión 3.0
- ✓ STATISTICA versión 12
- ✓ EasyReg Internacional versión 2015

## 5. ANÁLISIS DE LOS RESULTADOS Y DISCUSIÓN

El propósito de este apartado es analizar y discutir los resultados de la estadística descriptiva e inferencial descrita en el apartado anterior, permitiendo observar y

conocer, a través de los datos de las estaciones meteorológicas, los patrones y el comportamiento en el que ha estado evolucionando la precipitación pluvial sobre la cuenca del río Chiriquí a medida que transcurre el tiempo. Además de interpretar y comprender una clara percepción científica del comportamiento, como la tendencia y estacionalidad que predomina en la precipitación pluvial sobre la cuenca del río Chiriquí, desde su pasado, presente y futuro.

### 5.1. Análisis descriptivo

Se presentan diferentes tablas y gráficas, que muestran la distribución de probabilidad con sus parámetros de centralización; media, mediana y moda; así como sus parámetros de dispersión; varianza, desviación típica y coeficiente de variación, para cada una de las estaciones en estudio con un registro anual.

**Tabla 4: Resumen estadístico anual de la precipitación pluvial estación Finca Lériida**

Máx.	4 224,3 mm
Mín.	1 842,1 mm
Mediana	2 705,4 mm
Promedio	2 780,1 mm
Desviación Estándar	553,0 mm
Recuento	44

La Tabla 4 fue elaborado en base a la tabla de datos presentado en el Anexo 6.

Las estadísticas descriptivas que se muestran, nos indican que el valor máximo de precipitación se registró en el año 2010 con un registro de **4 224,3 mm**, en tanto que el mínimo ocurrió en el año 1977, con un registro de **1 842,1 mm**; además, es importante señalar que más del 50% de las precipitaciones pluviales oscilaron con **2 705,4 mm** y el

promedio histórico estuvo por el orden de **2 780,1 mm**, con una desviación estándar de **553,0 mm** para los 44 años de registros completos que corresponde de 1972-2015.

La Tabla 5 presenta el resumen estadístico anual de la estación David (108-023) zona baja; y la misma, fue elaborado en base a la tabla de datos presentado en el Anexo 7.

**Tabla 5: Resumen estadístico anual de la precipitación pluvial-estación David**

Máx.	3 623,7 mm
Mín.	1 868,3 mm
Mediana	2 598,9 mm
Promedio	2 585,6 mm
Desviación Estándar	370,4 mm
Recuento	44

Las estadísticas descriptivas que se muestran, nos indican que el valor máximo de precipitación se registró en el año 2008 con un registro de **3 623,7 mm**, en tanto que el mínimo ocurrió en el año 1976, con un registro de **1 868,3 mm**, además es importante señalar que más del 50% de las precipitaciones pluviales fue de **2 598,9 mm**, el promedio histórico estuvo por el orden de **2 585,6 mm**, con una desviación estándar de **370,4 mm**, para los 44 años de registros completos que corresponde de 1972-2015.

## **5.2. Precisión en la estimación de la precipitación**

El estudio del coeficiente de variación (*C.V.*) es un indicador que nos permite comparar las dispersiones de dos estaciones meteorológicas distintas, siempre que sus medias sean positivas. Se calcula para cada una de ellas y los valores que se obtienen se comparan entre sí. El cociente no debe exceder del **25%** para zona montañosa y **20%** para zonas bajas, con lo cual se verifica para las estaciones meteorológicas en estudio.

**Tabla 6: Precisión en la estimación de la precipitación: estación Finca Lérica**

	Anual
Promedio	2 780,1 mm
Desviación Estándar	553,0 mm
<b>Coeficiente de Variación</b>	<b>20 %</b>

**Tabla 7: Precisión en la estimación de la precipitación: estación David**

	Anual
Promedio	2 585,6 mm
Desviación Estándar	370,4 mm
<b>Coeficiente de Variación</b>	<b>14%</b>

En la Tabla 6 y Tabla 7 se muestra el coeficiente de variación para cada una de las estaciones; indicando un valor de 20% para la estación Finca Lérica y un valor de 14% para la estación David, lo cual se corrobora la confiabilidad de  $C.V. \leq 25\%$  y  $C.V. \leq 20\%$ , para las estaciones ubicadas en la zona alta y baja, respectivamente.

### 5.3. Análisis inferencial

#### 5.3.1. Análisis de consistencia de los registros pluviométricos

Dentro de la inferencia estadística, un contraste de hipótesis (también denominado test de hipótesis o prueba de significación) es un procedimiento para juzgar si una propiedad que se supone se cumple. Este contraste fue iniciado por Ronald Fisher, y fundamentado posteriormente por Jerzy Neyman y Karl Pearson.

### 5.3.1.1. Prueba de Rachas

Para la homogeneidad de los registros se utiliza la prueba de Rachas haciendo uso del programa “Microsoft Office Excel” para cada una de las estaciones. Y se presentan los resultados de las medianas que determinan las muestras  $n_1$  (+),  $n_2$  (-) y los grupos de rachas ( $r$ ) y posteriormente, calcular el estadístico de prueba ( $z$ ) y así rechazar o no el planteamiento de la hipótesis en un nivel de significancia que corresponde del -1,96 al +1,96 como los puntos críticos, para el rango de aceptación con una confianza del 95%.

<i>Tabla 8: Prueba de homogeneidad anual: 108-001</i>				
Zona alta: estación Finca Lérica				
Nº	Año	PP (mm)	Rachas	
1	1972	2272,0	-	1
2	1973	3500,0	+	2
3	1974	2596,9	-	3
4	1975	3279,9	+	4
5	1976	2374,9	-	5
6	1977	1842,1	-	
7	1978	2465,4	-	
8	1979	3266,2	+	6
9	1980	2868,4	+	
10	1981	2966,3	+	
11	1982	2217,5	-	7
12	1983	2052,4	-	
13	1984	2988,3	+	8
14	1985	1993,8	-	9
15	1986	2612,0	-	
16	1987	2453,1	-	
17	1988	3412,6	+	10
18	1989	2456,0	-	11
19	1990	2533,8	-	
20	1991	2470,2	-	
21	1992	1862,6	-	
22	1993	2838,6	+	12

Nº	Año	PP (mm)	Rachas	
23	1994	2356,6	-	13
24	1995	3084,6	+	14
25	1996	3229,9	+	
26	1997	2649,7	-	15
27	1998	3030,4	+	16
28	1999	3873,2	+	
29	2000	2707,6	+	
30	2001	2247,0	-	17
31	2002	2795,6	+	18
32	2003	3247,1	+	
33	2004	2994,0	+	
34	2005	3343,1	+	
35	2006	2756,9	+	
36	2007	3574,2	+	
37	2008	3720,0	+	19
38	2009	2379,6	-	
39	2010	4224,3	+	20
40	2011	3207,1	+	
41	2012	2665,0	-	21
42	2013	1986,3	-	
43	2014	2225,8	-	
44	2015	2703,1	-	
$n_1$	$n_2$	$Me$	$r$	$z$
22	22	2705,4	21	-0,61

Con los valores de las muestras de  $n_1 = 22$  y  $n_2=22$ , se observa que se debe proceder a calcular el estadístico de prueba, siendo este de  $z = -0,61$ .

**Significancia estadística:**

- Planteamiento de la hipótesis:
  - $H_0 : Me_1 - Me_2 = 0$
  - $H_1 : Me_1 - Me_2 \neq 0$
- Estadístico de prueba:  $z = -0,61$
- Región de Aceptación es,  $-1,96 \leq z_{0,05} \leq 1,96$
- Se evidencia que el estadístico de prueba  $z$  se encuentra en la región de aceptación; concluyendo que no se rechaza la hipótesis nula ( $H_0$ ) y considerando la homogeneidad de los valores, con una confiabilidad del 95% para la estación Finca Lérica.

Con respecto a la estación David (108-001), los valores de las muestras son  $n_1 = 14$  y  $n_2 = 30$ , el cual se observa en la Tabla 9. Uno de los valores de la muestra es mayor a 20, por tanto se proceder a calcular el estadístico de prueba, siendo este de  $z = -1,09$ .

**Significancia estadística:**

- Planteamiento de la hipótesis:
  - $H_0 : Me_1 - Me_2 = 0$
  - $H_1 : Me_1 - Me_2 \neq 0$
- Estadístico de prueba:  $z = -1,09$
- Región de Aceptación es,  $-1,96 \leq z_{0,05} \leq 1,96$

Se evidencia que el estadístico de prueba  $z$  se encuentra en la región de aceptación; concluyendo que no se rechaza la hipótesis nula ( $H_0$ ) y considerando la homogeneidad de los valores, con una confiabilidad del 95% para la estación David.

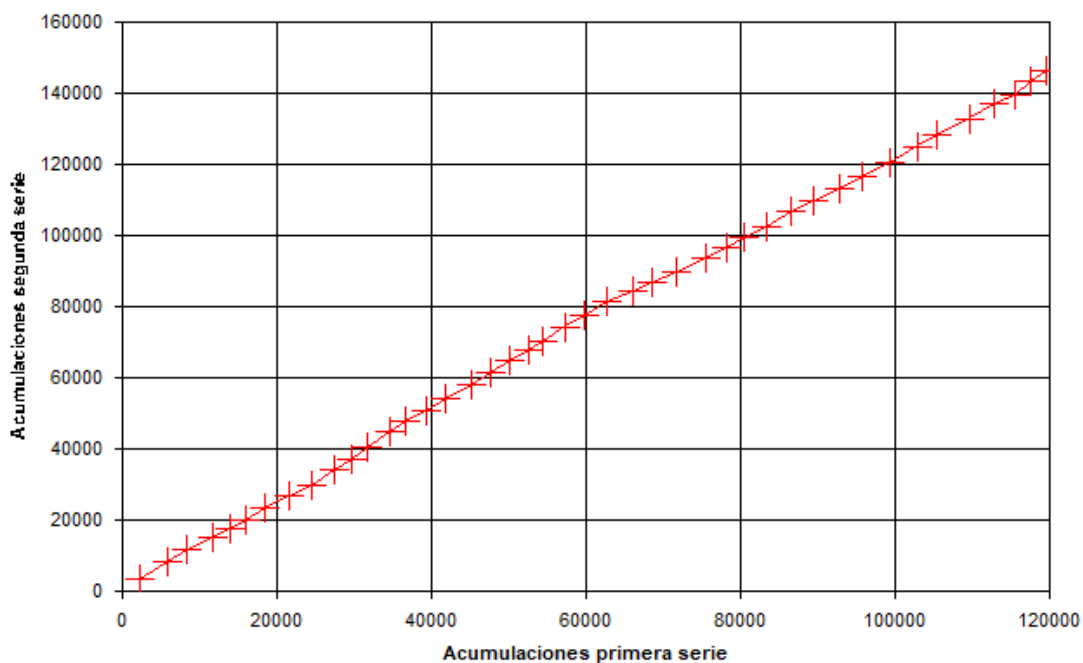
Tabla 9: Prueba de Homogeneidad Anual: 108-023				
Zona Baja: estación David				
Nº	Año	PP (mm)	Rachas	
1	1972	2316,5	-	1
2	1973	2994,3	+	2
3	1974	2699,1	-	3
4	1975	2652,9	-	
5	1976	2005,1	-	
6	1977	2444,0	-	
7	1978	2348,9	-	
8	1979	2657,6	-	
9	1980	2267,4	-	
10	1981	2602,7	-	
11	1982	2243,8	-	
12	1983	2649,8	-	
13	1984	2853,1	+	4
14	1985	2378,4	-	5
15	1986	2680,4	-	
16	1987	2372,9	-	
17	1988	2520,5	-	
18	1989	2252,7	-	
19	1990	2427,5	-	
20	1991	2173,4	-	
21	1992	2074,6	-	5
22	1993	2843,6	+	6
23	1994	2810,1	+	
24	1995	2881,7	+	
25	1996	2825,1	+	
26	1997	2227,2	-	7
27	1998	2976,4	+	8

Nº	Año	PP (mm)	Rachas	
28	1999	3073,2	+	8
29	2000	2624,0	-	9
30	2001	2683,2	-	
31	2002	2021,5	-	
32	2003	3252,9	+	10
33	2004	2584,8	-	11
34	2005	2787,5	+	12
35	2006	2576,5	-	13
36	2007	2878,5	+	14
37	2008	3326,7	+	
38	2009	2322,1	-	15
39	2010	3623,7	+	16
40	2011	2849,6	+	
41	2012	2079,6	-	17
42	2013	1868,3	-	
43	2014	2691,7	-	
44	2015	2395,7	-	
$n_1$	$n_2$	$Me$	$r$	$z$
14	30	2589,9	17	-1,09

### 5.3.1.2. Método de Dobles Acumuladas

Se rectifica la consistencia de los registros con el uso del programa HIDROBAS v3.0. Para la zona alta, la estación a comparar será la estación Los Naranjos (108-017) (Anexo 8), que es la estación más cercana a la estación base que es la estación Finca Lérica (108-001). Sin embargo, es importante resaltar que la estación Los Naranjos solo presenta registros pluviométricos de manera continua y completa desde 1972 hasta el año 2014; por tanto, el análisis del método de doble acumulada solo se analizara con 43 registros anuales y así corroborar la consistencia de los registros.

**Gráfica 1: Consistencia de los registros pluviométricos Estación Finca Lérica - Los Naranjos (Zona Alta)**

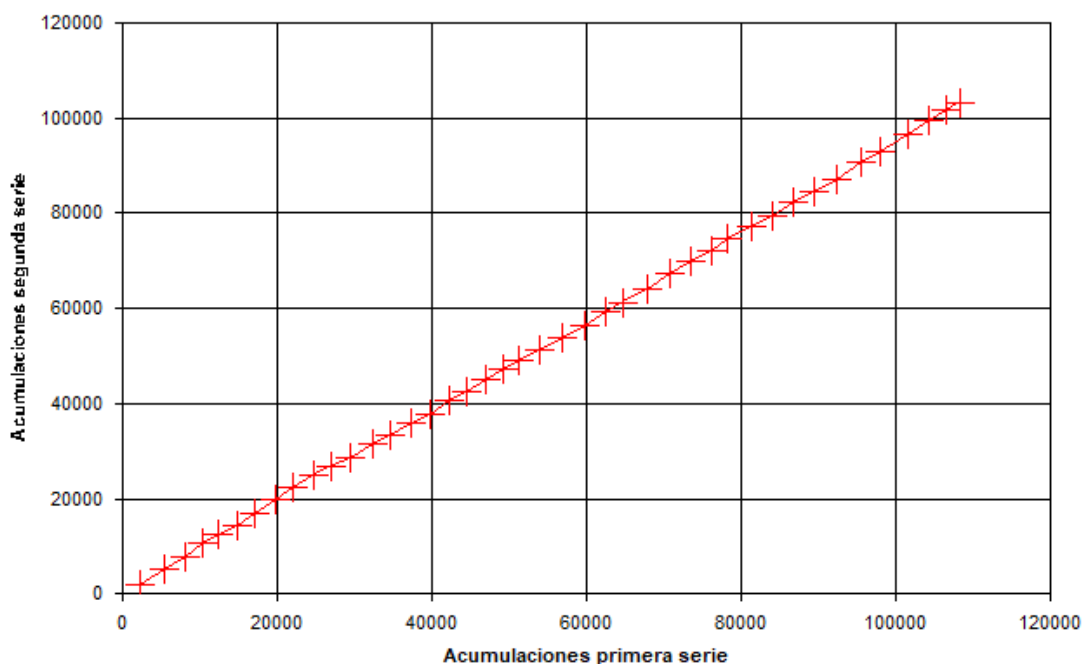


La Gráfica 1 indica que los valores de las dos estaciones son consistentes; puesto que, todos sus valores se encuentran sobre la línea recta, indicando así que los datos de las estaciones en todos y cada uno de los años son aceptables. Así pues, se corrobora la

homogeneidad de los valores de las estaciones Finca Lérica y Los Naranjos, pertenecientes a la zona alta de la cuenca con base a los 43 registros pluviométricos anuales.

No obstante, para la zona baja, la estación a comparar es la estación Veladero Gualaca (108-014) (Anexo 9) la cual corresponde a la estación más cercana a la estación base que es la estación David (108-023). Sin embargo, es importante resaltar que la estación Veladero Gualaca solo presenta registros pluviométricos de manera continua y completa desde 1972 hasta el año 2013; por tanto, el análisis del método de doble acumulada solo se analizara con 42 registros anuales y así corroborar la consistencia de los registros.

**Gráfica 2: Consistencia de los registros pluviométricos Estación David - Veladero Gualaca (Zona Baja)**



La Gráfica 2 corrobora una proporcionalidad en los valores de las dos estaciones, aunque existan años que están medidos por defecto o por exceso, siguen estando dentro

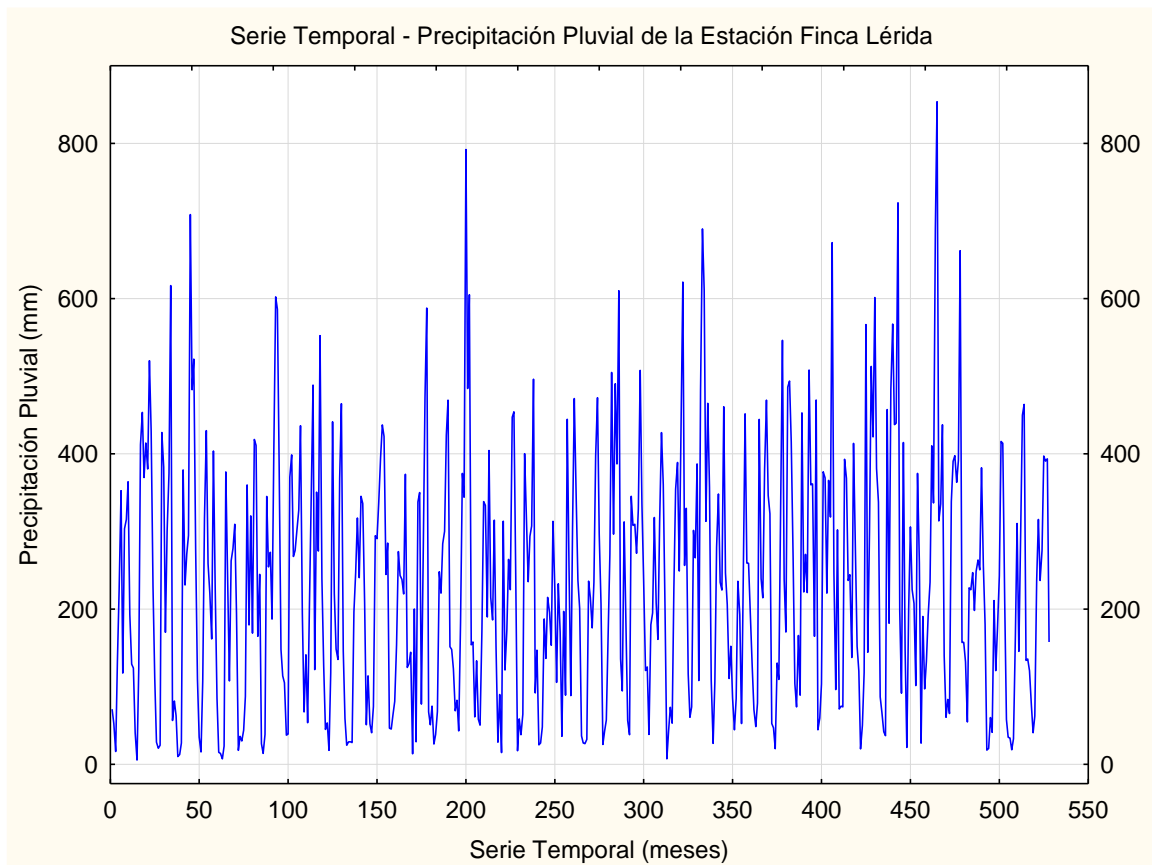
del rango de aceptación sobre la línea recta, indicando así que los valores de las estaciones son aceptables y comprobando la homogeneidad de los valores de las estaciones David y Veladero Gualaca con base a los 42 registros pluviométricos anuales.

### 5.3.2. Modelo ARIMA con la metodología de Box-Jenkins

#### 5.3.2.1. Gráfica de la serie de tiempo:

Se presentan las gráficas de series de tiempo de los valores de la precipitación pluvial, para observar los patrones que poseen.

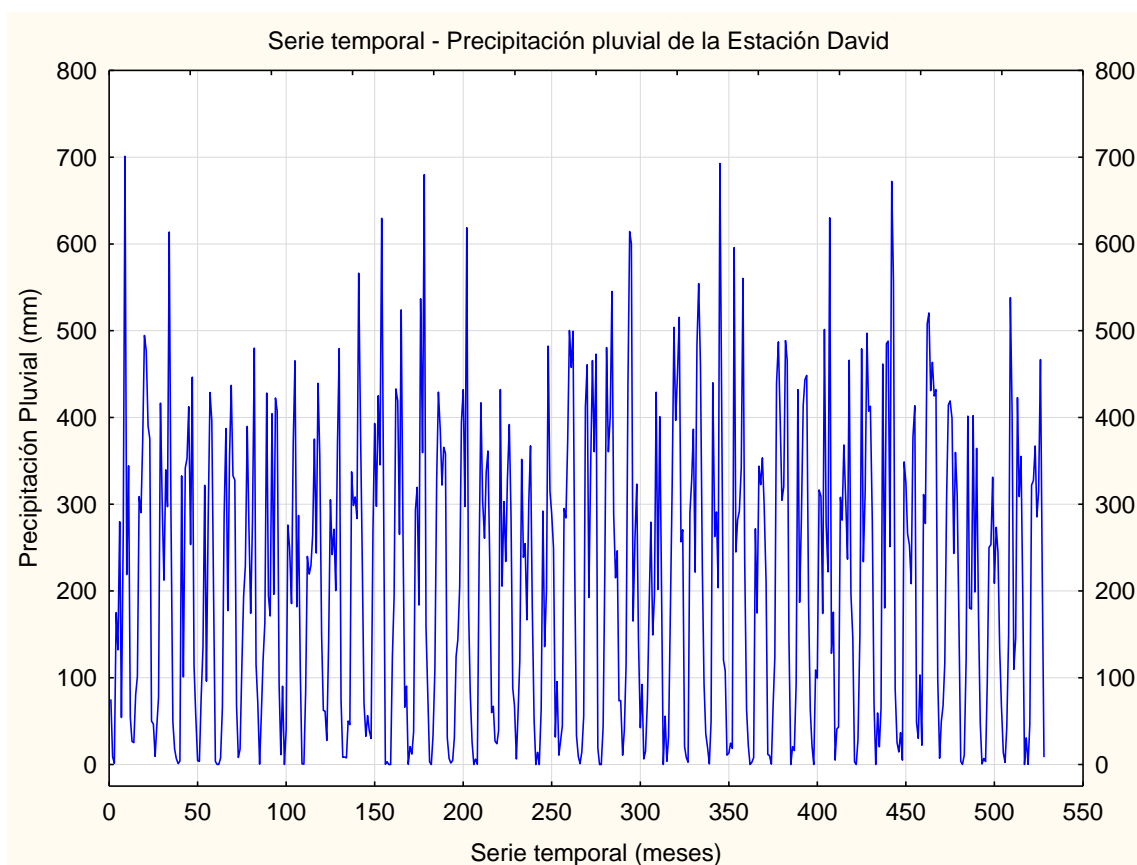
**Gráfica 3: Serie de tiempo estación Finca Lérída**



Las series de tiempo presentadas en las Gráfica 3 y 4 muestran una gran variabilidad estacional y posible variación irregular; con valores muy altos en algunos meses como

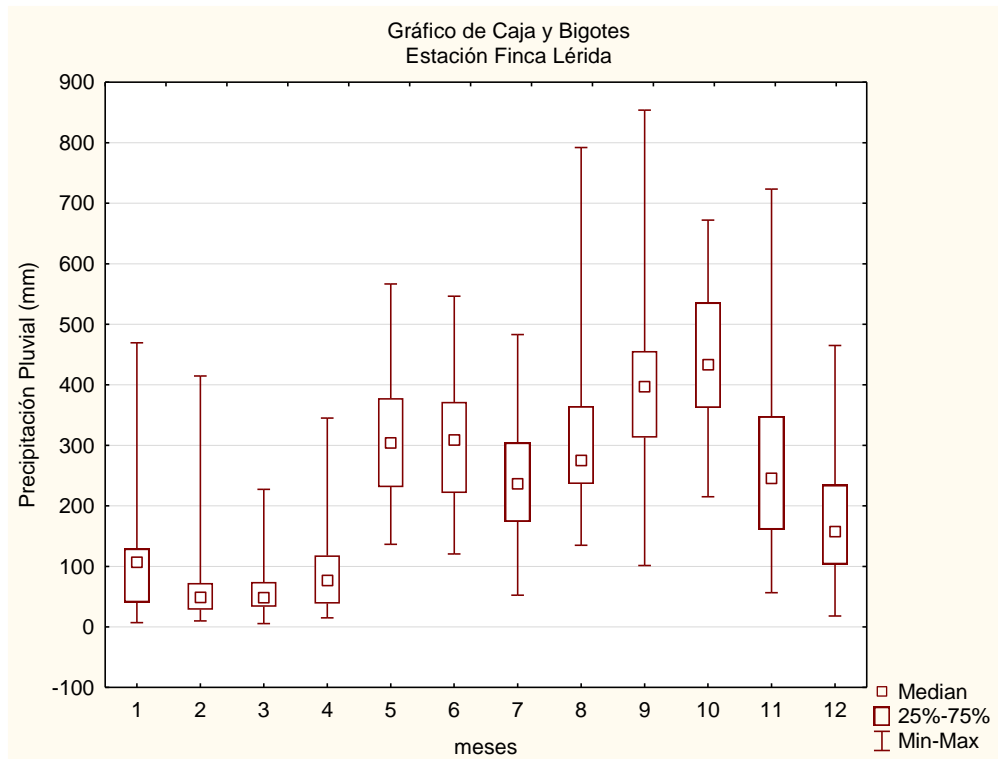
septiembre y octubre, y muy bajos para los meses de enero y febrero de cada año, lo cual se debe a las estaciones lluviosas y secas, estableciendo así un fuerte efecto estacional cada 12 meses para cada uno de los 44 años de la serie.

**Gráfica 4: Serie de tiempo estación David**



Para analizar la estacionariedad en varianza, se observa el gráfico de caja y bigote, para cada estación pluviométrica; que, a su vez permiten corroborar la variabilidad estacional e irregular de la serie de tiempo para cada una de las estaciones en estudio. No obstante, este análisis se contrasta con la prueba de Levene el cual permitirá evaluar la igualdad de las varianzas por meses

**Gráfica 5: Gráfico Caja y Bigotes estación Finca Lérica**



**Tabla 10: Prueba de Levene (homogeneidad en varianza) estación Finca Lérica**

Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
PP-FL001	255946,5	11	23267,87	2361223	516	4576,014	5,084745	0,0000001

**Significancia estadística para la prueba de Levene:**

- Planteamiento de la hipótesis:
  - $H_0$  : La varianza poblacionales de la serie no son iguales por meses
  - $H_1$  : La varianza poblacionales de la serie son iguales por meses
- Estadístico de prueba: Levene  $F_{(11)} = 5,084745$ ; p-valor=0,000 000 1
- Grados de libertad: **11**.
- Región de Aceptación y de Rechazo: punto crítico es,  $\alpha = 0,05$

- Regla de Decisión: Si el P-valor resultante de la prueba de Levene es inferior a 0,05(5%) del nivel de significación  $p_{valor} = 0,00 < 0,05$ , es poco probable que las diferencias obtenidas en las variaciones de la muestra se hayan producido sobre la base de un muestreo aleatorio de una población con varianzas iguales. Por lo tanto, la hipótesis nula de igualdad de varianzas se rechaza y se concluye que hay una diferencia entre las variaciones en la población con una confiabilidad del 95%.

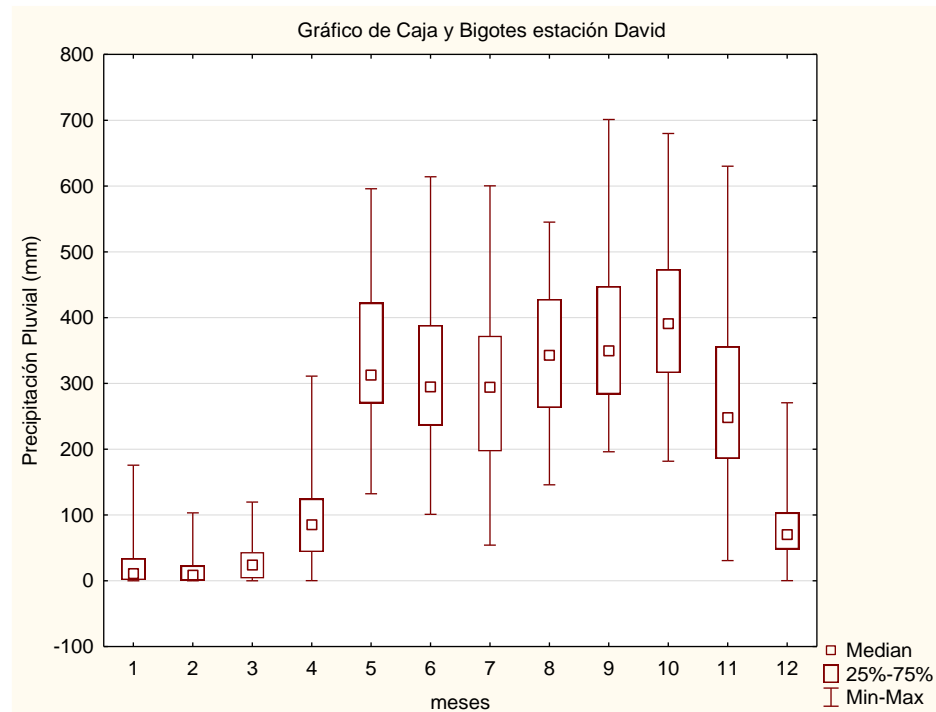
**Tabla 11: Prueba de Levene (homogeneidad en varianza) estación David**

Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
PP-D023	512307,9	11	46573,45	1676078	516	3248,214	14,33817	0,0000001

**Significancia estadística para la prueba de Levene:**

- Planteamiento de la hipótesis:
- $H_0$  : La varianzas poblacionales de la serie no son iguales por meses
  - $H_1$  : La varianzas poblacionales de la serie son iguales por meses
- Estadístico de prueba: Levene  $F_{(11)} = 14,33817$ ; p-valor=0,000 000 1
- Grados de libertad: 11.
- Región de Aceptación y de Rechazo: punto crítico es,  $\alpha = 0,05$
- Regla de Decisión: Si el P-valor resultante de la prueba de Levene es inferior a 0,05(5%) del nivel de significación  $p_{valor} = 0,00 < 0,05$ , es poco probable que las diferencias obtenidas en las variaciones de la muestra se hayan producido sobre la base de un muestreo aleatorio de una población con varianzas iguales. Por lo tanto, la hipótesis nula de igualdad de varianzas se rechaza y se concluye que hay una diferencia entre las variaciones en la población con una confiabilidad del 95%.

**Gráfica 6: Gráfico Caja y Bigotes estación David**



En la Gráfica 5 y 6 las varianzas de las estaciones en estudio no son constantes, debido a que no muestran los puntos a la misma altura ni con el mismo grado de dispersión por período; al igual es corroborado al aplicar la prueba de Levene para la serie de cada estación.

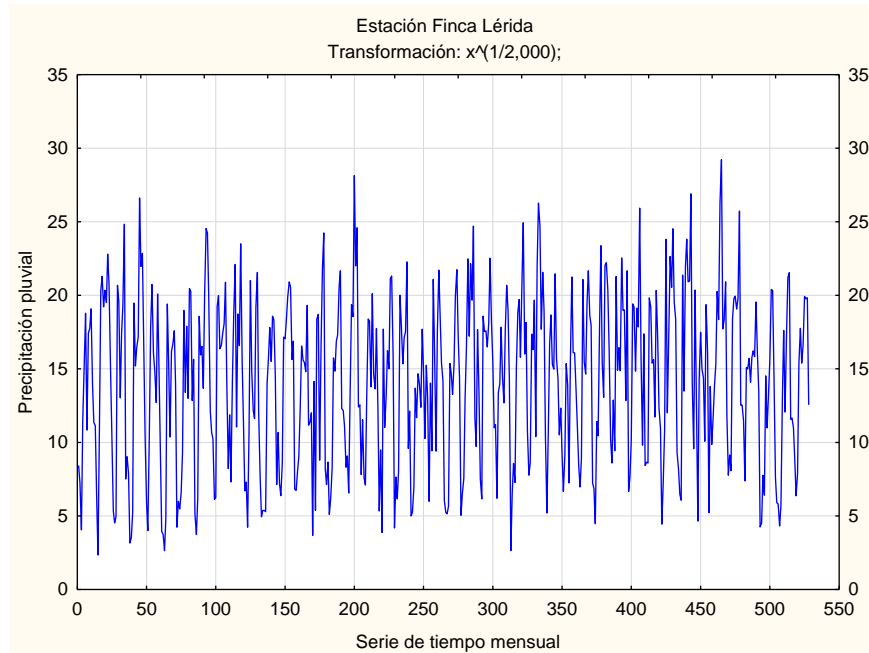
Con respecto a lo anterior, esto es un indicativo para la aplicación de un modelo de descomposición multiplicativo  $ARIMA(p, d, q)(P, D, Q)_s$ , ya que la serie muestra que las amplitudes parecen variar con el tiempo y su estacionalidad marcada; sin embargo, esto se corroboraría en los siguientes apartados.

### 5.3.2.2. Transformaciones de la serie (Estacionariedad en varianza)

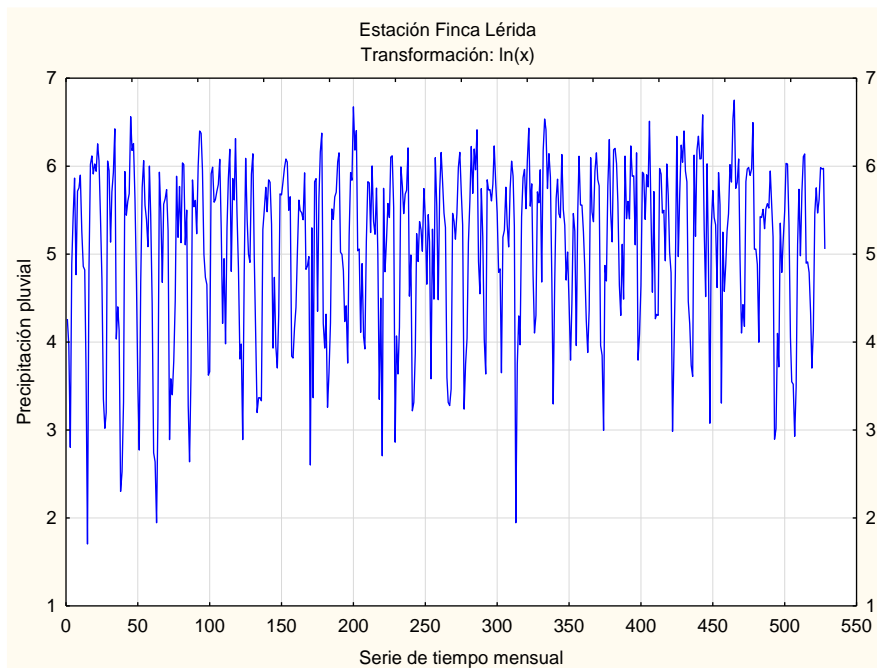
Se presentan las gráficas de las transformaciones Box-Cox, aplicando las transformaciones: logaritmo natural y raíz cuadrada, para cada una de las estaciones en estudio, y así estabilizar la varianza. Cuando la variabilidad de la serie en torno a su

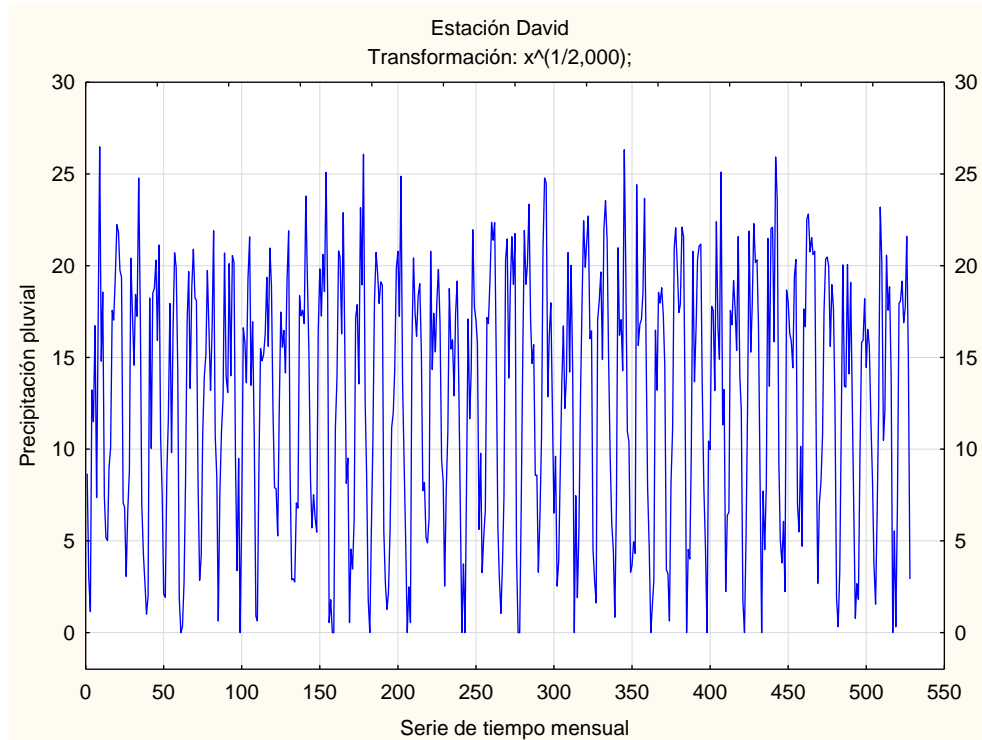
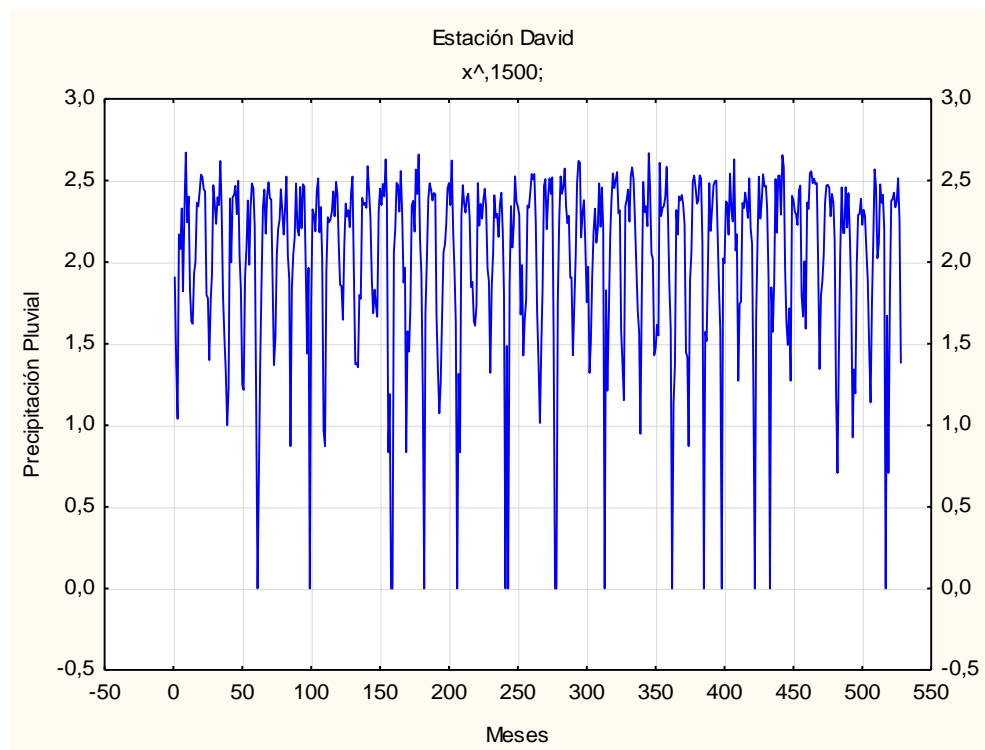
media se mantiene constante a lo largo del tiempo, se comprueba que la serie es constante en varianza; sin embargo, si ambos logran estabilizarla se hace uso del *criterio de parsimonia*.

**Gráfica 7: Transformación raíz cuadrada - estación Finca Lérica**



**Gráfica 8: Transformación logaritmo natural - estación Finca Lérica**



**Gráfica 9: Transformación raíz cuadrada - estación David****Gráfica 10: Transformación  $x^{0,15}$  - estación David**

En la Gráfica 7, 8, 9 y 10 se observa que la amplitud de estos ciclos se ha homogeneizado y que la variabilidad de las series es mucho más estable en el tiempo.

Se observa que para la serie de la estación Finca Lérica, ambas transformaciones logran homogeneizar la serie temporal; por consiguiente, se elige y concluye que la serie es estacionaria en varianza con logaritmo natural, haciendo uso del criterio de parsimonia.

Es importante resaltar que en la serie temporal de la estación David, no se pudo aplicar la transformación logaritmo natural, ya que la misma poseen varios datos nulos; por lo tanto, se realiza un proceso iterativo en el cual se selecciona la transformación que mejor logre homogeneizar la serie en varianza la cual resulto con potencia a la 0,15.

Sin embargo, se observa que en las gráficas transformadas la serie no fluctúa alrededor de una media constante, indicativo que no existe estacionariedad en media.

#### **5.3.2.3. Eliminación de la tendencia (Estacionariedad en media)**

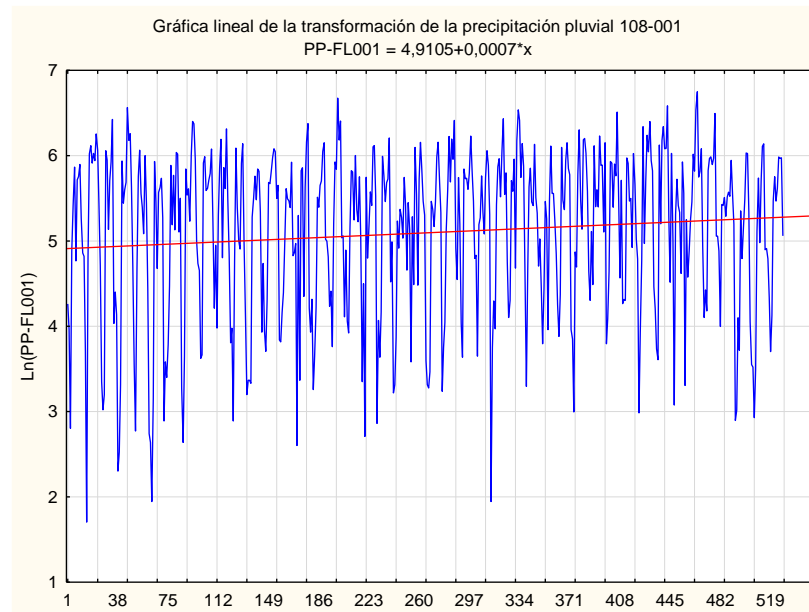
Se procede analizar la estacionariedad en media de las series de la precipitación pluvial.

Una serie no es estacionaria en media cuando:

- La serie de tiempo no fluctúa alrededor de una media constante,
- La función de autocorrelación simple (FAS) y parcial (FAP) decae lentamente hacia cero y, la no existencia de una raíz unitaria en la serie.

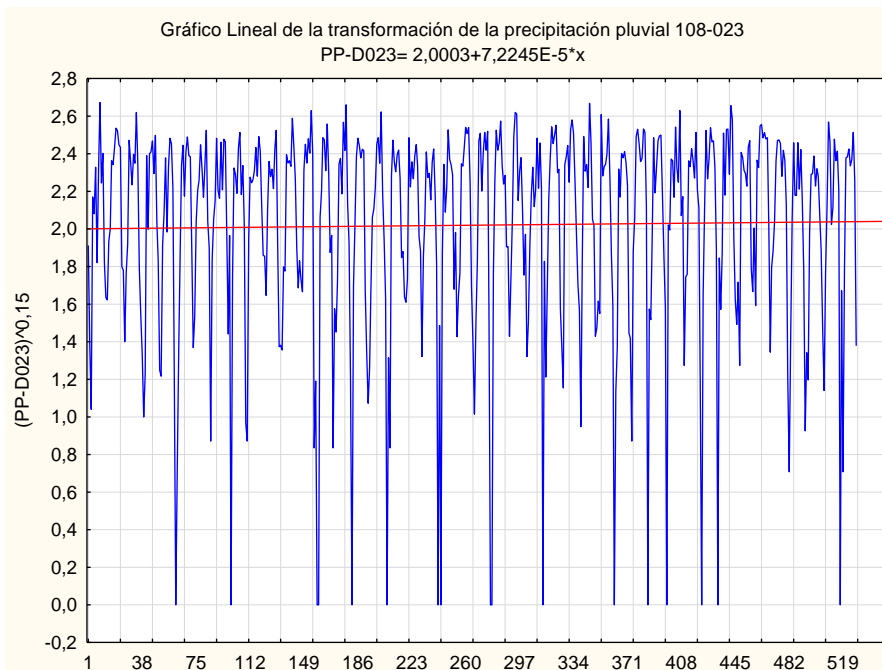
Se presentan las gráficas de las series estacionarias en varianza, además de las gráficas de las FAS y FAP de cada una de las estaciones en estudio, para observar si las mismas poseen o no tendencias en la serie; y de ser así, diferenciarla en la parte estacional y/o regular las veces que sea necesario.

**Gráfica 11: Gráfica lineal de la serie estacionaria en varianza, estación Finca Lérica**

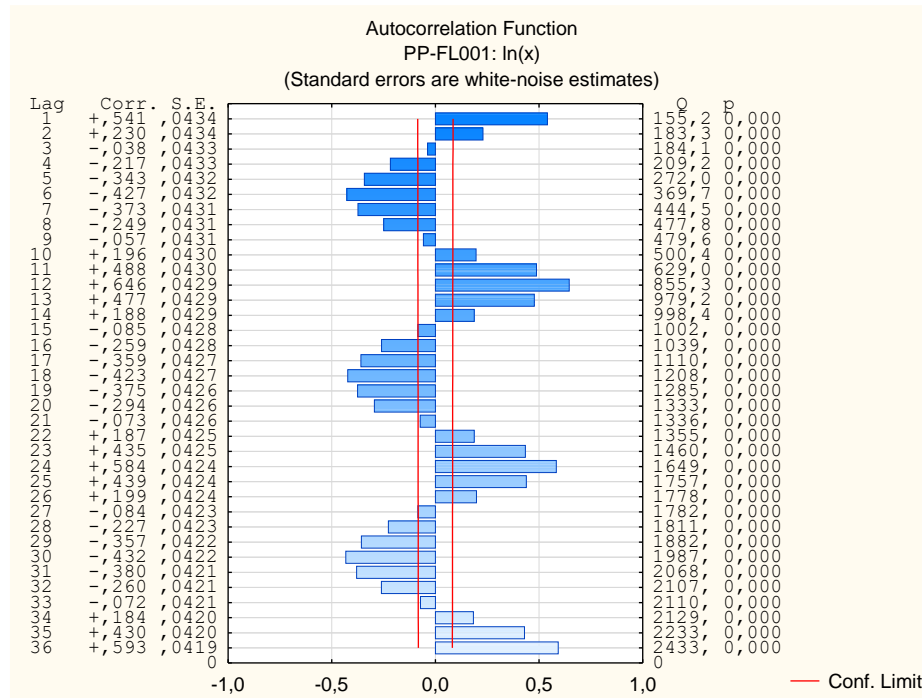


En la Gráfica 11 y 12 de las series estacionarias en varianza se observa que en ambas series de tiempo no fluctúa alrededor de una media constante, con una pequeña inclinación; no obstante, se procede a analizar las funciones de autocorrelación simple al igual que las parciales para verificar si la serie está o no estacionaria en media.

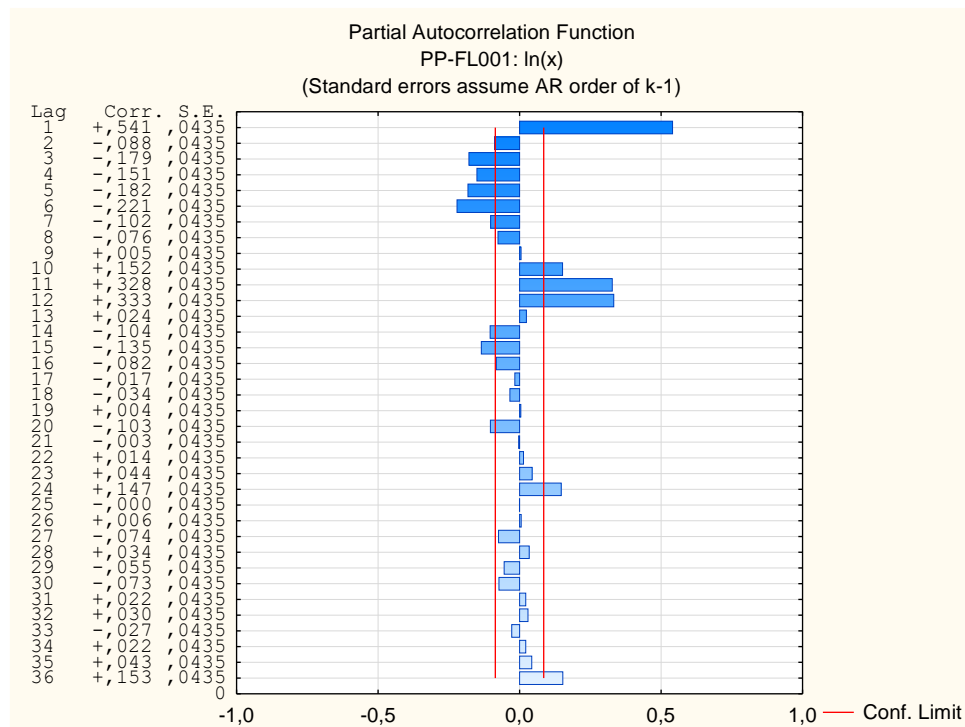
**Gráfica 12: Gráfica lineal de la serie estacionaria en varianza, estación David**



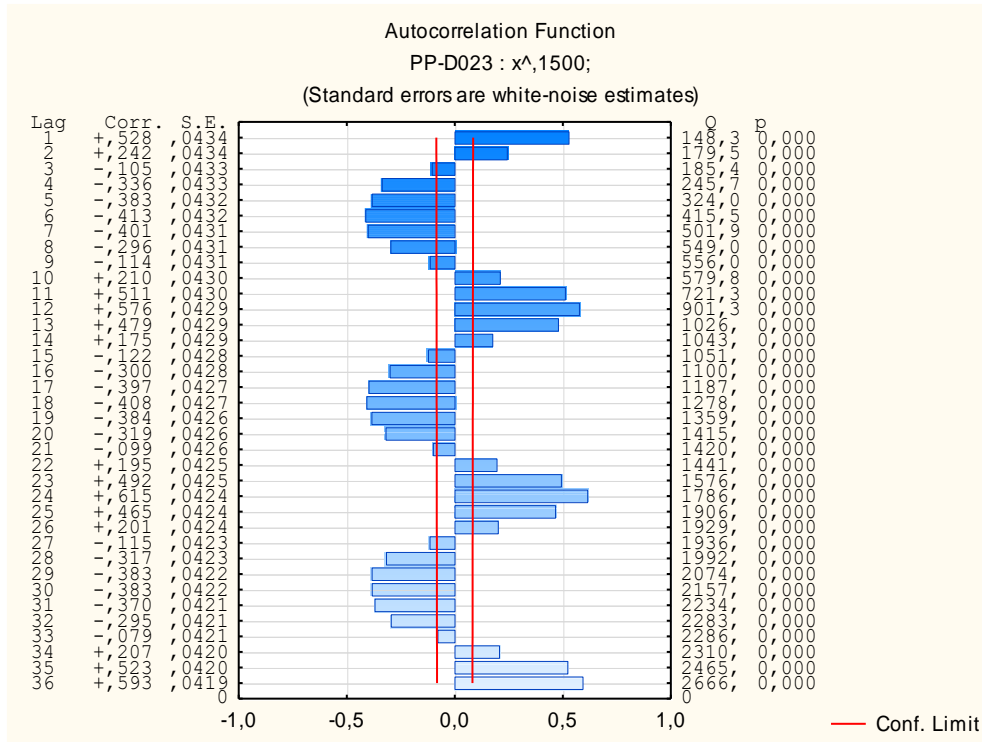
**Gráfica 13: Función de autocorrelación simple de la serie transformada - estación Finca Lérica**



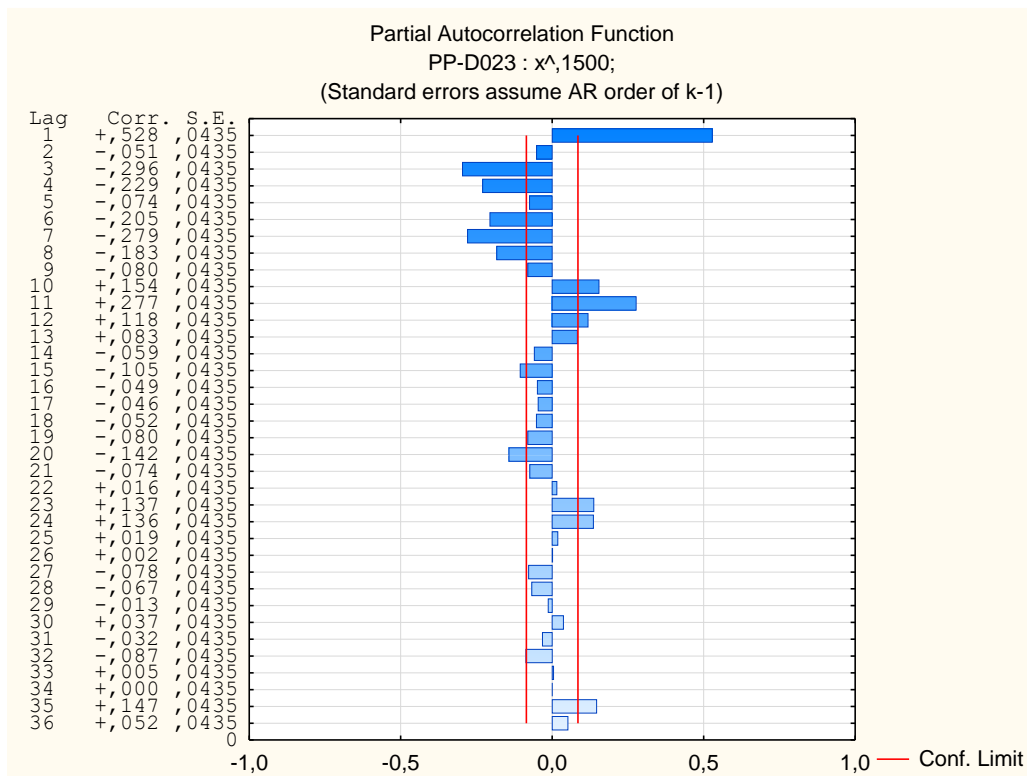
**Gráfica 14: Función de autocorrelación parcial de la serie transformada - estación Finca Lérica**



**Gráfica 15: Función de autocorrelación simple - estación David**



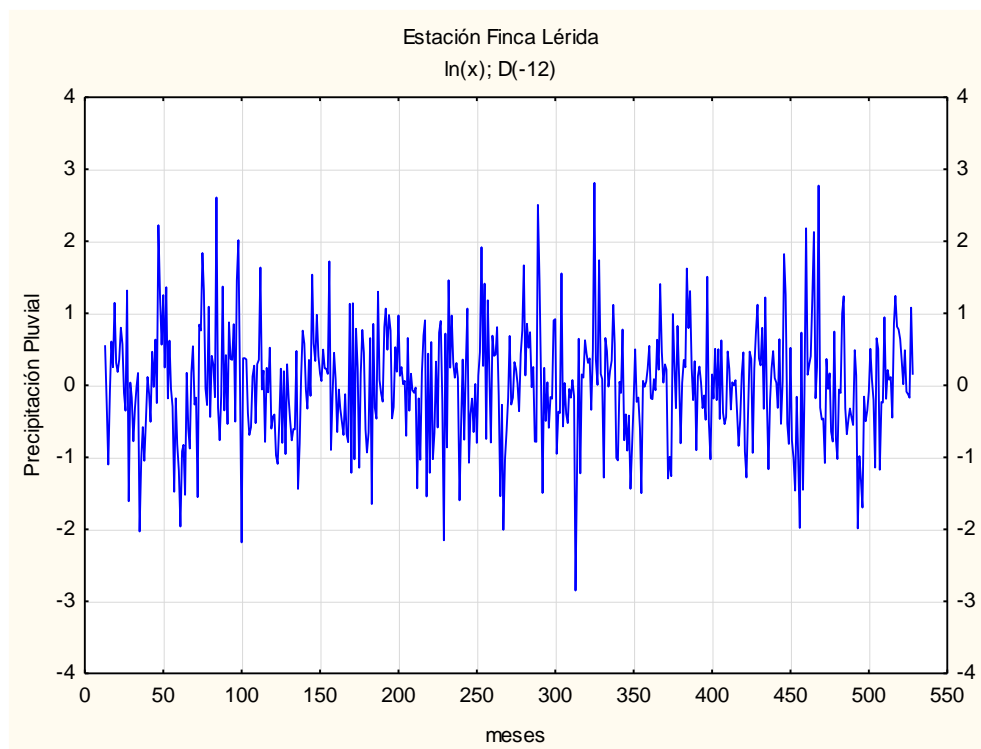
**Gráfica 16: Función de autocorrelación parcial de la serie - estación David**



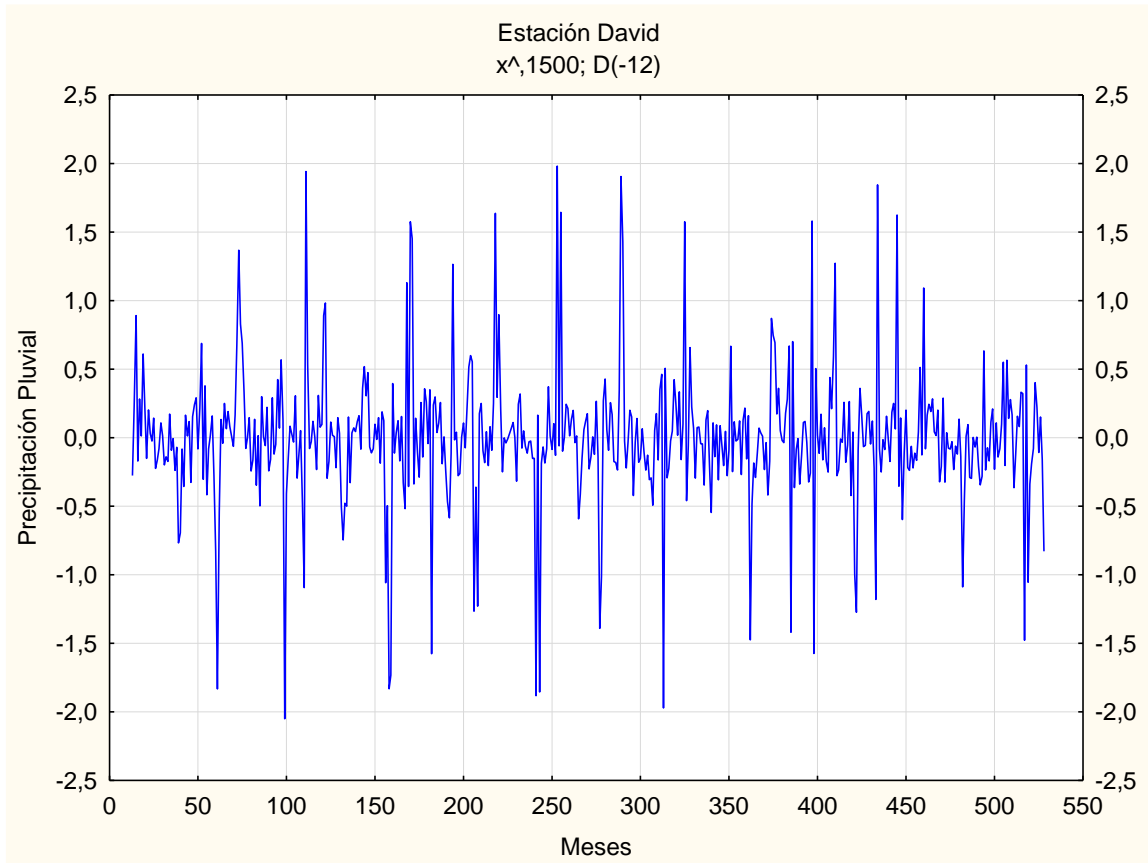
En la Gráfica 13 y 15 de la FAS indican que hay estacionalidad, ya que presenta un comportamiento periódico mensual por año, específicamente en el retardo (lag) 12 que corresponde a diciembre y así se mantiene por los siguientes retardos 24, 36, ..., presentan estructura positiva con decrecimiento lento, esto se presume que es debido a las estaciones secas y lluviosas de los 12 meses del año. Además, en la Gráfica 14 y 16 de la FAP se corrobora la estacionalidad en los períodos 12, 24, 36..., que se mostraban en las gráficas de la serie original de cada estación. Esto es un indicativo de que las series en estudio, son integradas en la parte estacional.

Para la estacionariedad en media de la parte estacional se hace el diferenciado  $D = 1$  con un retardo estacionalidad de  $s = 12$  y se procede a graficar la serie diferenciada en la parte estacional.

**Gráfica 17: Diferenciado estacional de la serie transformada - estación Finca Lérica**



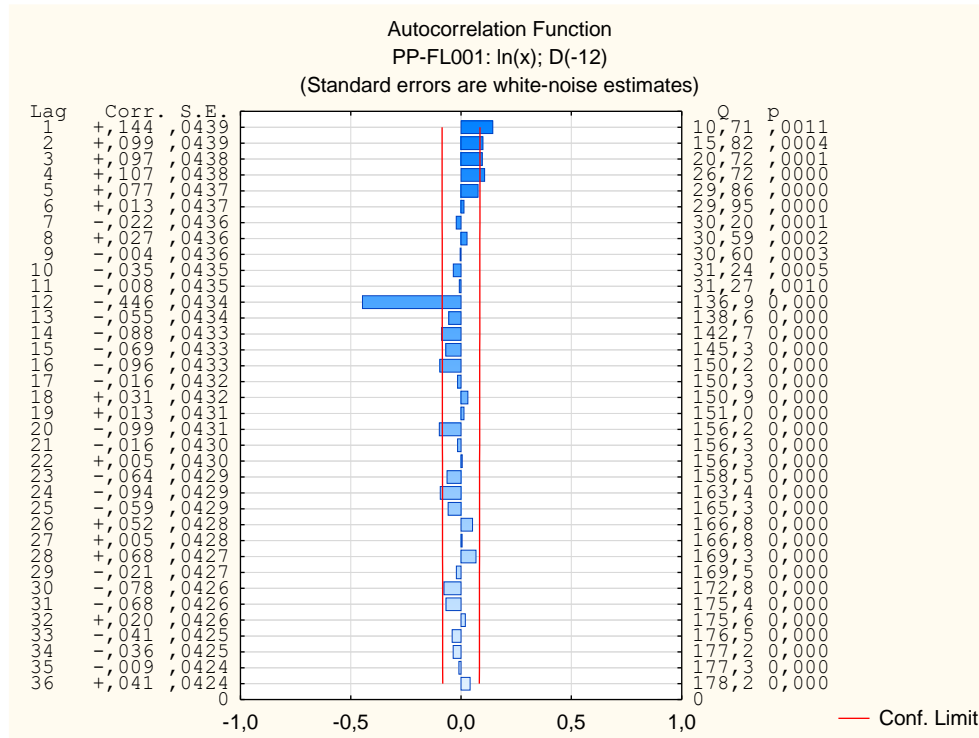
**Gráfica 18: Diferenciado estacional de la serie transformada - estación David**



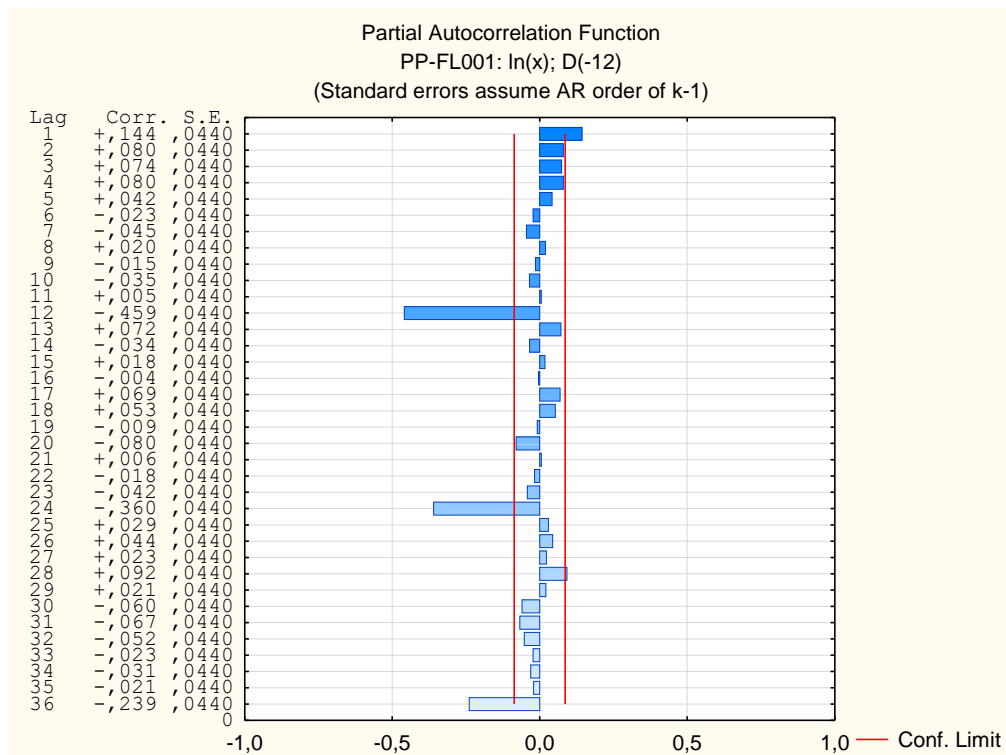
Se observa que en las Gráficas 17 y 18, se estabilizó los picos grandes que representaban la parte estacional; sin embargo, se evidencia una posible tendencia en la parte regular ya que la serie no fluctúa alrededor de la media constante.

Por consiguiente, se rectifica la tendencia en la parte regular con la función de autocorrelación simple y parcial, de la serie transformada y diferenciada en la parte estacional, para cada estación.

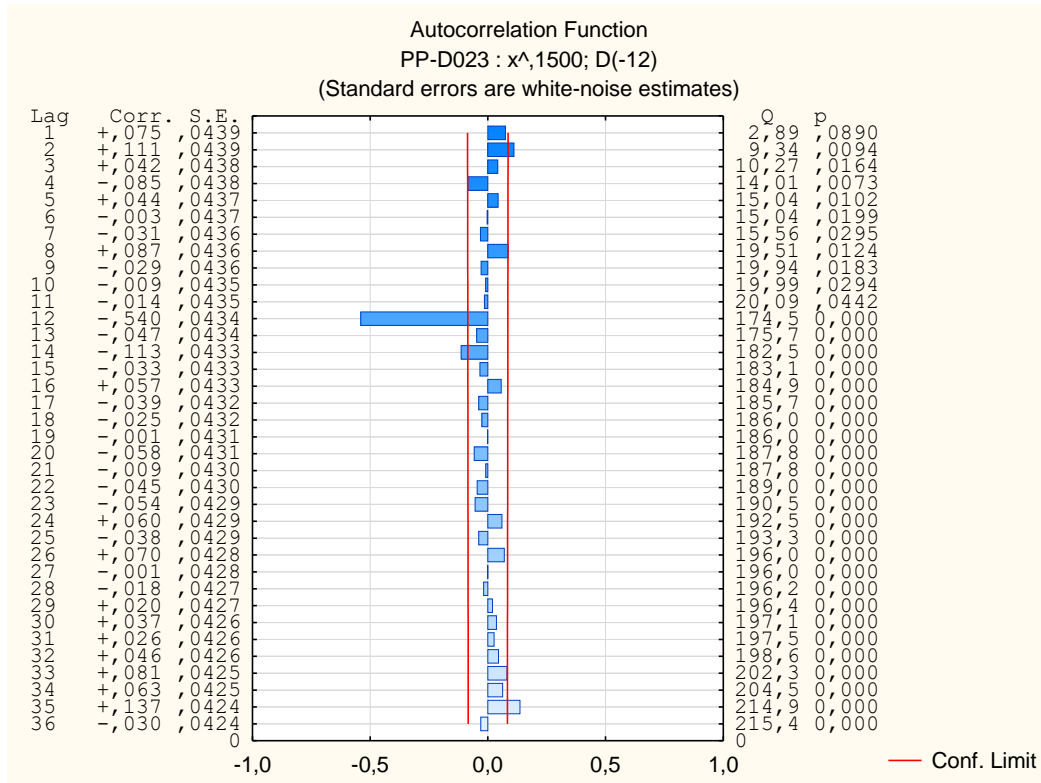
**Gráfica 19: FAS de la serie transformada y diferenciada - estación Finca Lérica**



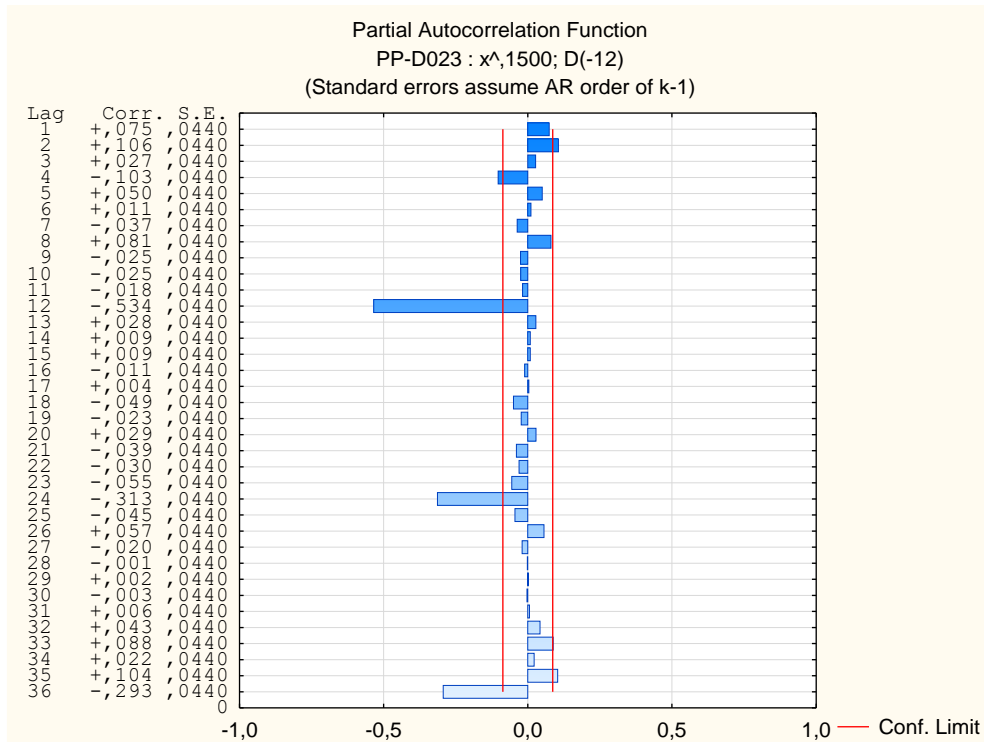
**Gráfica 20: FAP de la serie transformada y diferenciada - estación Finca Lérica**



**Gráfica 21: FAS de la serie transformada y diferenciada - estación David**



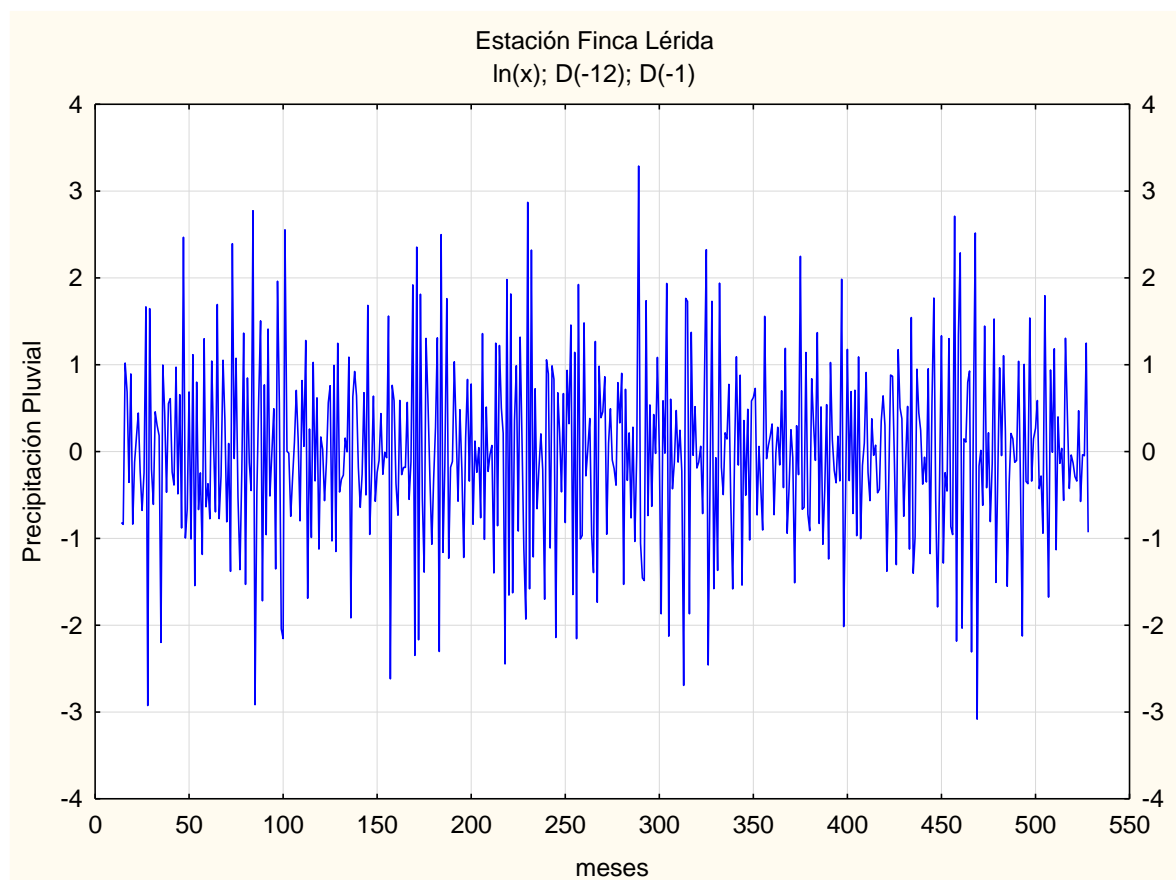
**Gráfica 22: FAP de la serie transformada y diferenciada - estación David**



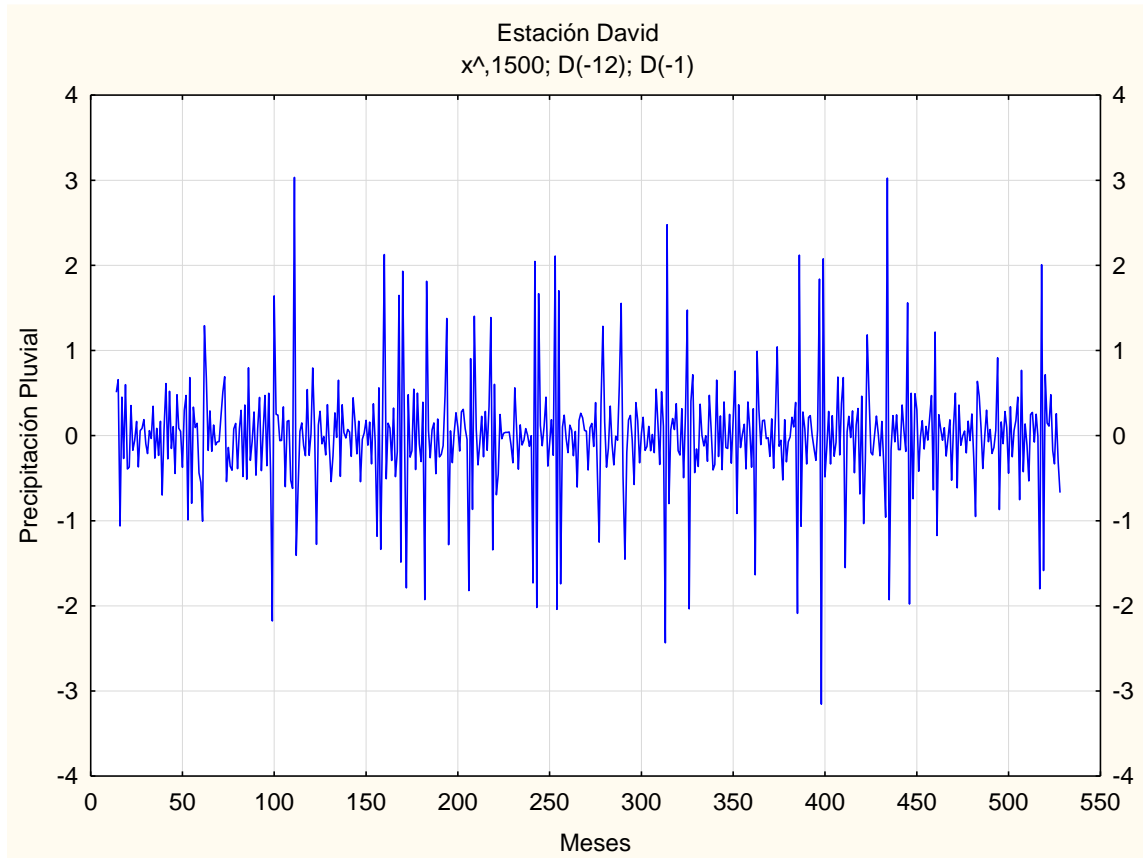
En las Gráficas 19 y 21 de la FAS y Gráficas 20 y 22 de la FAP de la serie transformada y diferenciada, se corrobora que posee tendencia en la parte regular, ya que decae lentamente hasta aproximarse a cero en los once primeros retardos (**lags**) para la estación Finca Lérica; en cambio, para la estación David se observa un comportamiento sinodal, demostrando así que es una serie integrada en la parte regular; y se debe diferenciar con  $d = 1$ , para ambas series.

Una vez diferenciada en la parte regular; se procede a graficar la serie de tiempo transformada y diferenciada, tanto en la parte estacional y regular; para corroborar la estacionariedad en media en cada una de las estaciones en estudio.

**Gráfica 23: Diferenciado de la parte estacional y regular de la serie transformada - estación Finca Lérica**



**Gráfica 24: Diferenciado de la parte estacional y regular de la serie transformada**  
**- estación David**



En la Gráfica 23 y 24 se observa que la serie ahora se encuentra estacionaria en media, tanto en la parte regular como en la estacional; y, de su estacionariedad en varianza.

Para corroborar que la serie cumple con la estacionariedad en media tanto para la parte regular como la estacional, se procede a analizar la hipótesis de raíz unitaria en ambas series, y para la cual se utiliza el contraste de Dickey-Fuller Aumentada (ADF)

**Tabla 12: Contrastes de raíces unitarias, estación Finca Lérida**

ADF model for $z(t)-z(t-1)$ :		
	OLS estimate	t-value Asymptotic critical regions:
$z(t-1)$	-5.6694	-5.7052 < -3.40 (5%) < -3.13 (10%) p-value = 0.00000
$z(t-1)-z(t-2)$	3.8960	3.9795
$z(t-2)-z(t-3)$	3.2120	3.3659
$z(t-3)-z(t-4)$	2.6709	2.8989
$z(t-4)-z(t-5)$	2.2399	2.5374
$z(t-5)-z(t-6)$	1.9208	2.2892
$z(t-6)-z(t-7)$	1.6675	2.1057
$z(t-7)-z(t-8)$	1.4504	1.9523
$z(t-8)-z(t-9)$	1.3090	1.8858
$z(t-9)-z(t-10)$	1.2492	1.9311
$z(t-10)-z(t-11)$	1.2024	1.9938
$z(t-11)-z(t-12)$	1.2725	2.2607
$z(t-12)-z(t-13)$	0.7644	1.4436
$z(t-13)-z(t-14)$	0.3833	0.7819
$z(t-14)-z(t-15)$	-0.0025	-0.0057
$z(t-15)-z(t-16)$	-0.3044	-0.7605
$z(t-16)-z(t-17)$	-0.5558	-1.5838
$z(t-17)-z(t-18)$	-0.6892	-2.2962
$z(t-18)-z(t-19)$	-0.7088	-2.8567
$z(t-19)-z(t-20)$	-0.6604	-3.3910
$z(t-20)-z(t-21)$	-0.6204	-4.3909
$z(t-21)-z(t-22)$	-0.4828	-5.3766
$z(t-22)-z(t-23)$	-0.2817	-6.4303
1	0.0113	0.1672
t	0.00001	-0.0741
Residual s.e.:	69.80450E-002	
R-square:	0.85819	
n:	492	

**Significancia estadística para el contraste de raíces unitaria de la Tabla 12:**

- Planteamiento de la hipótesis:
  - $H_0 : \beta = 0$ ; Existencia de raíz unitaria (serie no estacionaria en media)
  - $H_1 : \beta < 0$ ; No existe raíz unitaria (serie estacionaria en media)
- Estadístico de prueba: Dickey-Fuller  $t_c = -5,7052$  (p-valor=0,000 01)
- Región de Aceptación y de Rechazo: punto crítico es,  $DF_{0,05} = -3,40$

- **Regla de Decisión:** Si el P-valor resultante de la prueba de : Dickey-Fuller es inferior a 0,05(5%) del nivel de significación  $p_{valor} = 0,00 < 0,05$ . Por lo tanto, se rechaza la hipótesis nula de existencia de raíz unitaria a un nivel de significación del 5%; es decir, la serie de tiempo transformada de la estación Finca Lérica, se encuentra estacionaria en media (parte regular y estacional) y varianza.

**Tabla 13: Contrastes de raíces unitarias, estación David**

ADF model for z(t)-z(t-1):		
	OLS estimate	t-value Asymptotic critical regions:
z(t-1)	-7.0116	-6.3259 < -3.40 (5%) < -3.13 (10%) p-value = 0.00000
z(t-1)-z(t-2)	5.1839	4.7446
z(t-2)-z(t-3)	4.5518	4.2746
z(t-3)-z(t-4)	3.9924	3.8803
z(t-4)-z(t-5)	3.4573	3.5062
z(t-5)-z(t-6)	3.0294	3.2301
z(t-6)-z(t-7)	2.6548	2.9956
z(t-7)-z(t-8)	2.3338	2.8055
z(t-8)-z(t-9)	2.1424	2.7598
z(t-9)-z(t-10)	1.9941	2.7625
z(t-10)-z(t-11)	1.9496	2.9090
z(t-11)-z(t-12)	1.9924	3.1930
z(t-12)-z(t-13)	1.3715	2.3385
z(t-13)-z(t-14)	0.8745	1.6152
z(t-14)-z(t-15)	0.4688	0.9557
z(t-15)-z(t-16)	0.1322	0.3032
z(t-16)-z(t-17)	-0.1165	-0.3079
z(t-17)-z(t-18)	-0.2831	-0.8851
z(t-18)-z(t-19)	-0.4222	-1.6170
z(t-19)-z(t-20)	-0.4936	-2.4343
z(t-20)-z(t-21)	-0.4633	-3.1419
z(t-21)-z(t-22)	-0.3924	-4.1345
z(t-22)-z(t-23)	-0.2268	-4.9701
1	0.0028	0.0667
t	0.00001	-0.1038
Residual s.e.:	43.39777E-002	
R-square:	0.87755	
n:	492	

### Significancia estadística para el contraste de raíces unitaria de la Tabla 13:

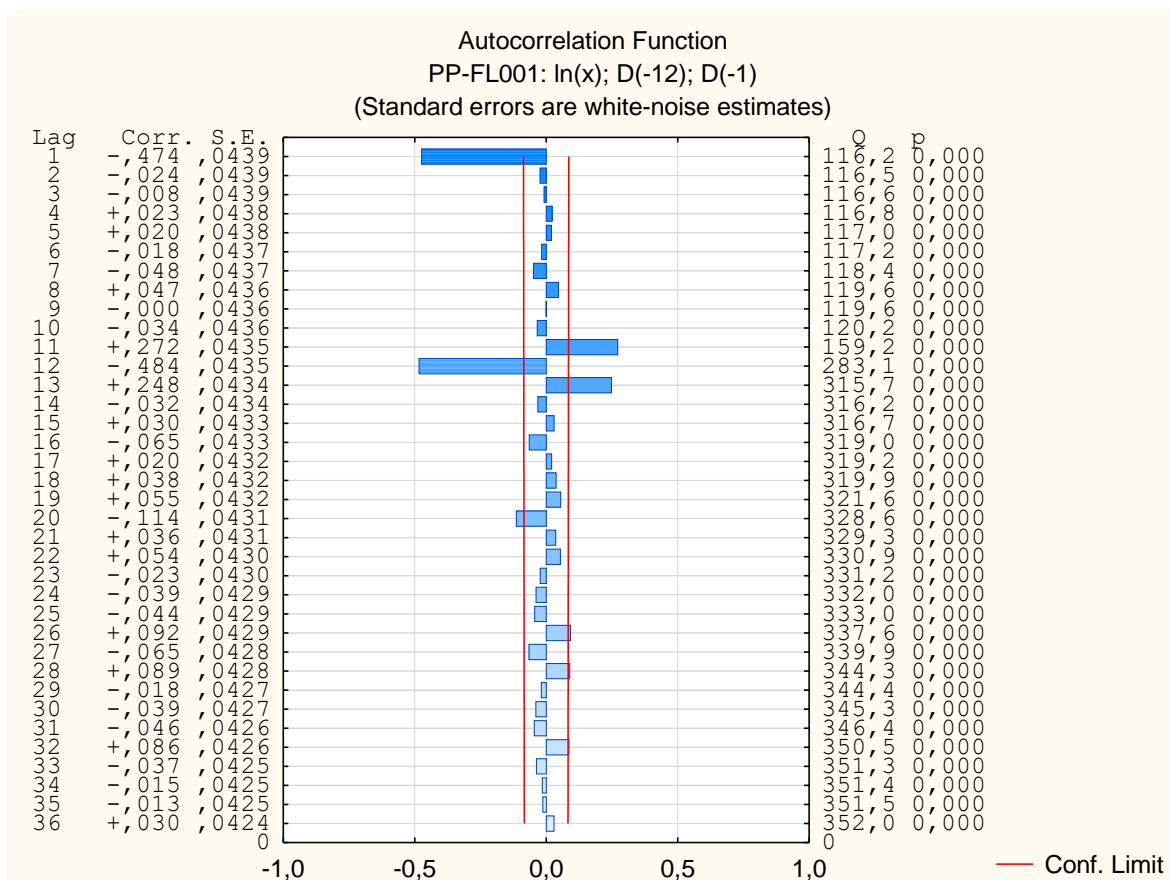
- Planteamiento de la hipótesis:
  - $H_0 : \beta = 0$ ; *Existencia de raíz unitaria (serie no estacionaria en media)*
  - $H_1 : \beta < 0$ ; *No existe raíz unitaria (serie estacionaria en media)*
- Estadístico de prueba: Dickey-Fuller  $t_c = -6,3259$  (**p-valor=0,000 01**)
- Región de Aceptación y de Rechazo: punto crítico es,  $DF_{0,05} = -3,40$
- Regla de Decisión: Si el P-valor resultante de la prueba de : Dickey-Fuller es inferior a 0,05(5%) del nivel de significación  $p_{valor} = 0,00 < 0,05$ . Por lo tanto, se rechaza la hipótesis nula de existencia de raíz unitaria a un nivel de significación del 5%; es decir, la serie de tiempo transformada de la estación David, se encuentra estacionaria en media (parte regular y estacional) y varianza.

#### 5.3.2.4. Identificación del modelo

Luego de estabilizar las series en varianza y media, se procede a la identificación de la posible estructura del modelo, pudiendo ser estas de: Autorregresivos  $AR(p)$ , Media móvil  $MA(q)$ , o Autorregresivo de media móvil  $ARMA(p, q)$ .

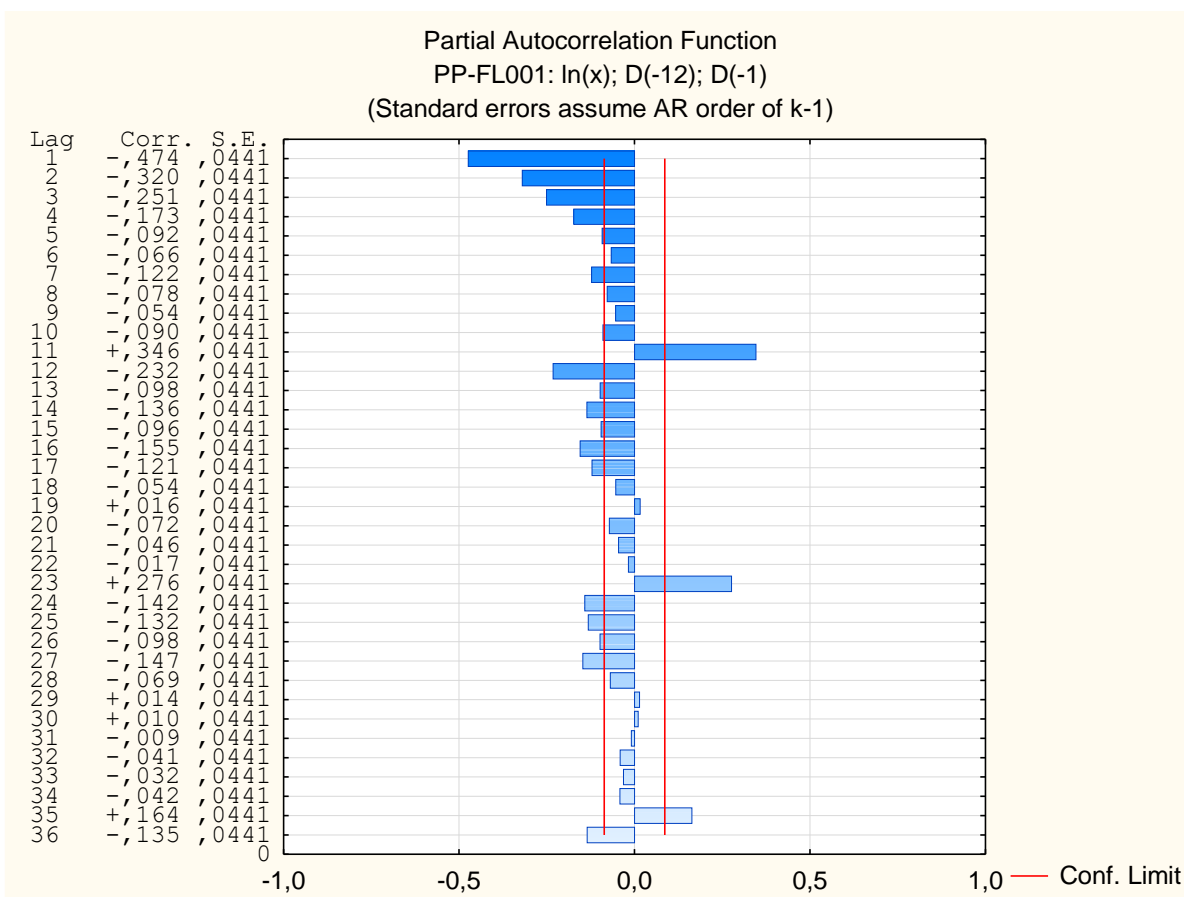
A continuación, se presentan las funciones de autocorrelación simple (FAS) y parcial (FAP) de las series estacionaria en media y varianza; para identificar la posible estructura del modelo para cada una de las estaciones en estudio, tomando como referencia la *Tabla 2: Identificación del Modelo (p. 45)*.

**Gráfica 25: FAS de la serie transformada y diferenciada en la parte estacional y regular sobre los 36 primeros retardos - estación Finca Lérída**



En la Gráfica 25: FAS de la serie transformada y diferenciada en la parte estacional y regular sobre los 36 primeros retardos, no presenta ningún corte o truncado, sino que el primer retardo sobresale de la banda de significatividad y luego decrece rápidamente siguiendo una estructura de seno-coseno; lo mismo sucede para los retardos 11, 12 y 13 que luego decrecen rápidamente para los retardos 24 y 36, producto de la estacionalidad. Basándose únicamente en la FAS podría corresponder a una estructura de tipo *AR* ó *ARMA* de cualquier orden.

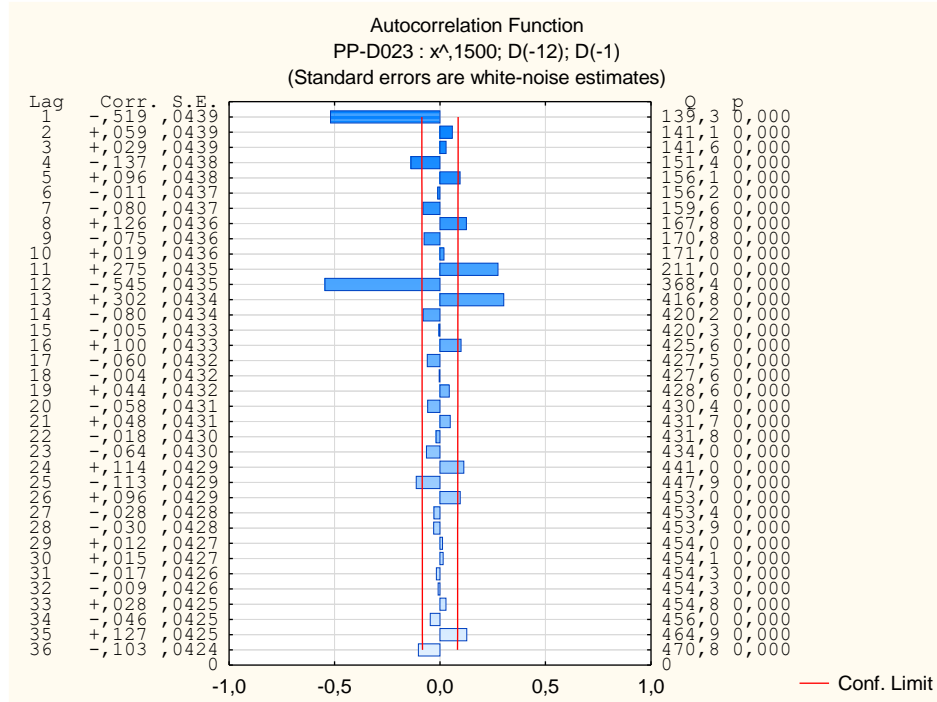
**Gráfica 26: FAP de la serie transformada y diferenciada en la parte estacional y regular sobre los 36 primeros retardos - estación Finca Lérída**



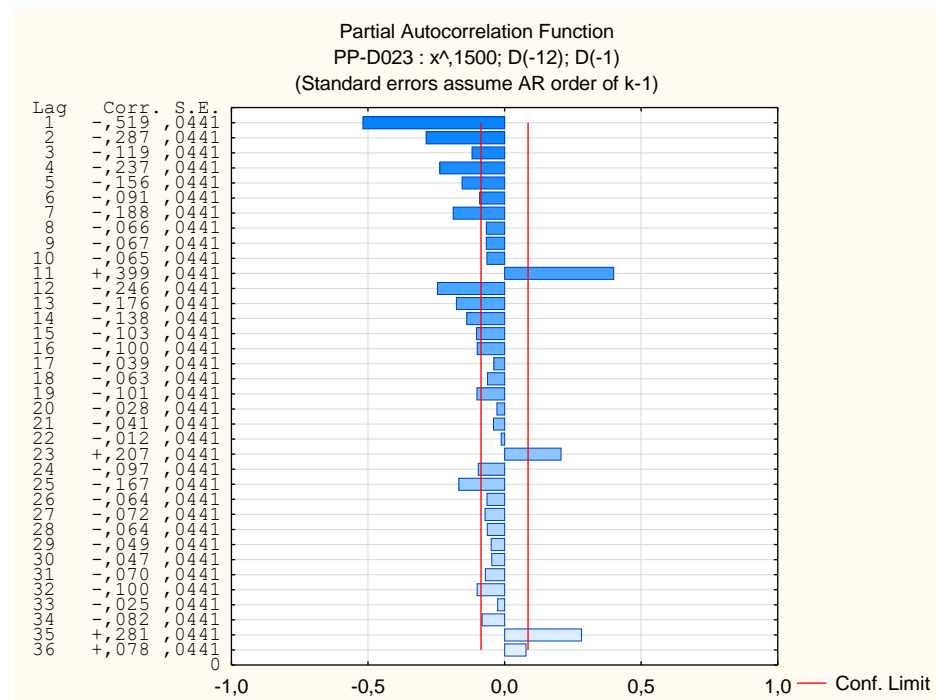
Para respaldarse en la identificación del modelo  $ARMA(p,q)(P,Q)_s$  se acude a la función de autocorrelación parcial (FAP), en la Gráfica 26: FAP de la Serie transformada y diferenciada en la parte estacional y regular sobre los 36 primeros retardos, se observa que hay un decrecimiento lento de tipo exponencial.

Luego de analizar ambos autocorrelogramas y comparándolo con el *Tabla 2: Identificación del Modelo (p. 48)* se asume, que es un modelo de tipo autorregresivo de media móvil,  $ARMA(p,q)(P,Q)_s$ , ya que ninguna de las barras se trunca, y decrecen exponencialmente aproximándose a cero.

**Gráfica 27: FAS de la serie transformada y diferenciada en la parte estacional y regular sobre los 36 primeros retardos - estación David**



**Gráfica 28: FAP de la serie transformada y diferenciada en la parte estacional y regular sobre los 36 primeros retardos - estación David**



En la Gráfica 27 FAS de la serie transformada y diferenciada en la parte estacional y regular sobre los 36 primeros retardos, no presenta ningún corte o truncado, existen diferentes retardos (del 1 al 11) que sobresale de la banda de significatividad siguiendo una estructura de tipo seno-coseno, lo mismo sucede para los retardos 11, 12 y 13, que luego decrecen rápidamente para los retardos 24 y 36; resultado de la estacionalidad.

Basándose únicamente en la FAS no se podría saber el tipo de modelo. Para ello se recurre al análisis de la Gráfica 28 FAP de la serie transformada y diferenciada en la parte estacional y regular sobre los 36 primeros retardos, y se observa que hay un decrecimiento lento de tipo lineal.

Al analizar ambos autocorrelogramas y comparándolo con el *Tabla 2: Identificación del Modelo (p. 48)* se observa que es muy difícil asegurar el tipo de modelo para esta estación; sin embargo, por la complejidad de la misma se elige un modelo tipo autorregresivo de media móvil,  $ARMA(p, q)(P, Q)_s$ .

#### 5.3.2.5. Estimación de los coeficientes del modelo

Luego de tener una percepción de la estructura del modelo para cada una de las estaciones, se pretende estimar los primeros coeficientes o parámetros “ $q$ ” y “ $p$ ” que se estimarán a partir de las funciones de autocorrelogramas simple y parcial, respectivamente (presentadas en el apartado anterior) tomando como referencia los correlogramas simples y parciales teóricos<sup>6</sup> para los modelos *AR* y *MA*.

En este apartado, se presentará el primer modelo con sus coeficientes y se analizará los parámetros de correlación del modelo propuesto; para luego, hacer los correctivos

---

<sup>6</sup> Ver **Anexo 10**: Modelo AR y MA teóricos

pertinentes hasta que se encuentre el modelo que sea más convincente de forma analítica, para cada una de las estaciones en estudio.

Con respecto a la función de autocorrelación simple de la estación Finca Lérica (**Gráfica 25**), se observa que el primer coeficiente es significativamente no nulo y el resto tiende a cero; en cuanto a la función de autocorrelación parcial (**Gráfica 26**), muchos de sus coeficientes son significativamente no nulos y decrecen en forma exponencial; considerando así un modelo de orden  $p = 2$  y  $q = 1$  para la parte regular.

Para obtener los valores de  $P$  y  $Q$ , de la parte estacional; se realiza el mismo análisis que se utilizó para determinar  $p$  y  $q$ , pero estudiando solamente los retardos que son múltiplos de 12; teniendo como resultado dos coeficientes significativos en la FAS y luego decrece rápidamente aproximándose a cero; y para el FAP, se observa que decrece lentamente de forma lineal.

Se considera un modelo estacional de orden  $P = 2$  y  $Q = 2$ ; y se establece, como primer modelo  $ARIMA(2,1,1)(2,1,2)_{12}$  para luego calcular los parámetros de correlación.

**Tabla 14: Estimación de coeficientes y parámetros de correlación del modelo 1**

Modelo 1					
Transformación			Logaritmo Natural		
Retardo estacional			12		
<i>ARIMA</i>					
<i>p</i>	<i>d</i>	<i>q</i>	<i>P</i>	<i>D</i>	<i>Q</i>
2	1	1	2	1	2

Input: PP-FL001 Model:(2,1,1)(2,1,2) Seasonal lag: 12 MS Residual= ,39279								
Paramet.	Constant	p(1)	p(2)	q(1)	Ps(1)	Ps(2)	Qs(1)	Qs(2)
Constant	1,000000						-0,730499	-0,001577
p(1)		1,000000	0,272388	0,600875	0,023203	0,009081	0,016827	-0,025653
p(2)		0,272388	1,000000	0,524949	0,021469	-0,078391	0,013552	-0,011213
q(1)		0,600875	0,524949	1,000000	0,037927	-0,058395	0,025169	-0,038697
Ps(1)		0,023203	0,021469	0,037927	1,000000	-0,143753	0,993598	-0,987891
Ps(2)		0,009081	-0,078391	-0,058395	-0,143753	1,000000	-0,159025	0,221240
Qs(1)	-0,730499	0,016827	0,013552	0,025169	0,993598	-0,159025	1,000000	-0,993582
Qs(2)	-0,001577	-0,025653	-0,011213	-0,038697	-0,987891	0,221240	-0,993582	1,000000

En el Tabla 14, se observa que existen correlaciones altas y bajas entre los parámetros, indicando que posiblemente el modelo está sobre parametrizado o que la especificación propuesta no es la más adecuada.

Se procede a intentar con otro modelo, siendo este  $ARIMA(2,1,2)(2,1,1)_{12}$  y se procede a estudiar los parámetros de correlación.

**Tabla 15:** Estimación de coeficientes y parámetros de correlación del modelo 2

Modelo 2					
Transformación			Logaritmo Natural		
Retardo estacional			12		
ARIMA					
<i>p</i>	<i>d</i>	<i>q</i>	<i>P</i>	<i>D</i>	<i>Q</i>
2	1	2	2	1	1

Input: PP-FL001 Model:(2,1,2)(2,1,1) Seasonal lag: 12 MS Residual= ,39279								
Paramet.	Constant	p(1)	p(2)	q(1)	q(2)	Ps(1)	Ps(2)	Qs(1)
Constant	1,000000						-0,584380	-0,041889
p(1)		1,000000	-0,351121	0,999062	-0,997696	0,072904	-0,063318	-0,054479
p(2)		-0,351121	1,000000	-0,373976	0,399620	-0,036311	-0,002126	-0,009891
q(1)		0,999062	-0,373976	1,000000	-0,998701	0,073233	-0,065932	-0,054576
q(2)		-0,997696	0,399620	-0,998701	1,000000	-0,074846	0,063901	0,047504
Ps(1)		0,072904	-0,036311	0,073233	-0,074846	1,000000	0,384252	0,641094
Ps(2)	-0,584380	-0,063318	-0,002126	-0,065932	0,063901	0,384252	1,000000	0,573896
Qs(1)	-0,041889	-0,054479	-0,009891	-0,054576	0,047504	0,641094	0,573896	1,000000

En la Tabla 15, se evidencia que posee una buena correlación entre los parámetros y se toma como bueno este modelo.

Para la estación David, la función de autocorrelación simple (**Gráfica 27**), se observa que varios coeficientes sobresalen significativamente de la banda de significancia; en cuanto la función de autocorrelación parcial (**Gráfica 28**), muchos de sus coeficientes son significativamente no nulos y decrecen en forma lineal; considerando así un modelo de orden  $p = 2$  y  $q = 1$  para la parte regular. En cuanto para la parte estacional, los valores de orden estacional serían  $P = 2$  y  $Q = 1$ .

Se establece como primer modelo  $ARIMA(2,1,1)(2,1,1)_{12}$  y se corresponde a calcular los parámetros de correlación.

**Tabla 16: Estimación de coeficientes y parámetros de correlación del modelo 1**

Modelo 1					
Transformación			$x^{0,15}$		
Retardo estacional			12		
ARIMA					
$p$	$d$	$q$	$P$	$D$	$Q$
2	1	1	2	1	1

Input: PP-D023							
Model:(2,1,1)(2,1,1) Seasonal lag: 12 MS Residual= ,13738							
Paramet.	Constant	p(1)	p(2)	q(1)	Ps(1)	Ps(2)	Qs(1)
Constant	1,000000						0,002116
p(1)		1,000000	0,157616	0,444420	0,078755	0,096017	-0,007575
p(2)		0,157616	1,000000	0,390254	0,020697	-0,012555	-0,015619
q(1)		0,444420	0,390254	1,000000	0,042284	0,032892	-0,202009
Ps(1)		0,078755	0,020697	0,042284	1,000000	0,381196	0,528807
Ps(2)		0,096017	-0,012555	0,032892	0,381196	1,000000	0,484369
Qs(1)	0,002116	-0,007575	-0,015619	-0,202009	0,528807	0,484369	1,000000

En el Tabla 16, se evidencia que posee una buena correlación entre los parámetros y puede contemplar como bueno el modelo.

Sin embargo, se procura de buscar algún otro modelo que mejore el propuesto, obteniendo *ARIMA* (2 1, 1)(3, 1, 1)<sub>12</sub>. Se analiza los parámetros de correlación, para descartar o considerar el modelo como bueno.

**Tabla 17: Estimación de coeficientes y parámetros de correlación del modelo 2**

Modelo 2					
Transformación			$x^{0,15}$		
Retardo estacional			12		
ARIMA					
<i>p</i>	<i>d</i>	<i>q</i>	<i>P</i>	<i>D</i>	<i>Q</i>
2	1	1	3	1	1

Input: PP-D023								
Model:(2, 1, 1)(3, 1, 1) Seasonal lag: 12 MS Residual= ,13735								
Paramet.	Constant	p(1)	p(2)	q(1)	Ps(1)	Ps(2)	Ps(3)	Qs(1)
Constant	1,000000					0,645492		-0,000702
p(1)		1,000000	0,156304	0,444615	0,051141	0,062083	-0,056795	-0,039501
p(2)		0,156304	1,000000	0,391233	0,015547	-0,013358	-0,001788	-0,016312
q(1)		0,444615	0,391233	1,000000	0,019630	0,016139	0,009926	-0,185322
Ps(1)		0,051141	0,015547	0,019630	1,000000	0,500467	0,364130	0,628811
Ps(2)	0,645492	0,062083	-0,013358	0,016139	0,500467	1,000000	0,427027	0,607817
Ps(3)		-0,056795	-0,001788	0,009926	0,364130	0,427027	1,000000	0,502719
Qs(1)	-0,000702	-0,039501	-0,016312	-0,185322	0,628811	0,607817	0,502719	1,000000

En el Tabla 17, se evidencia que posee una buena correlación entre los parámetros y puede contemplar como bueno el modelo. Sin embargo, ya que ambos modelos soy bueno, se aplica el criterio de parsimonia y se considera el modelo 1 propuesto. No obstante, esta decisión se logra corroborar en el siguiente apartado.

### 5.3.2.6. Contrastes de validez del modelo

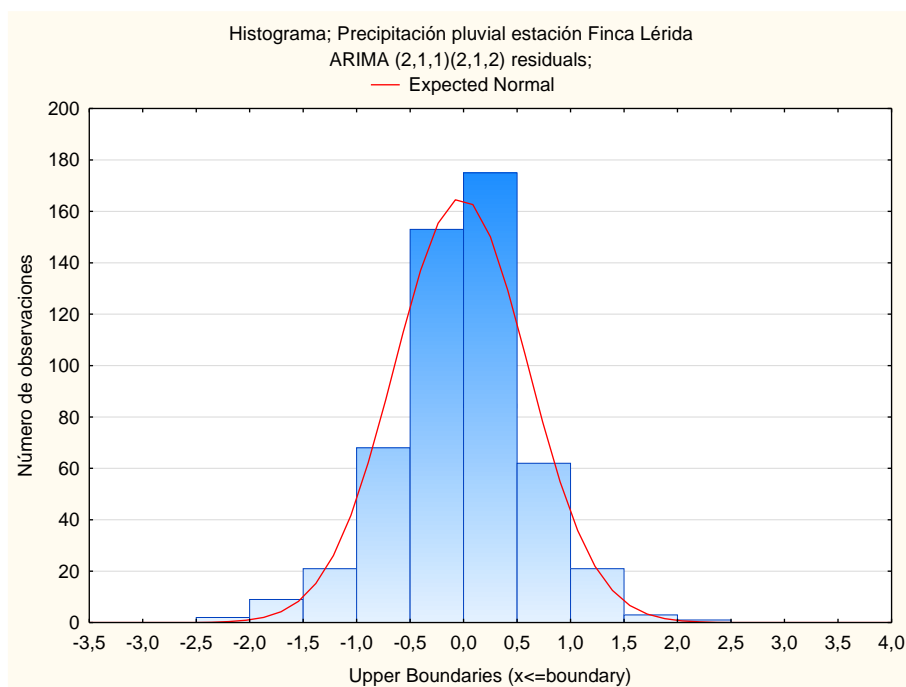
En esta sección se presentan los gráficos y cuadros que validan el contraste de los modelos propuestos, para cada una de las estaciones.

Se inicia con la normalidad de los residuos que se determinará a través de los gráficos de histograma y probabilidad normal, en los que se describe el comportamiento de los residuos al estimar el modelo y contrastarlo con los datos originales (de acuerdo a lo justificado en el acápite c, p.50).

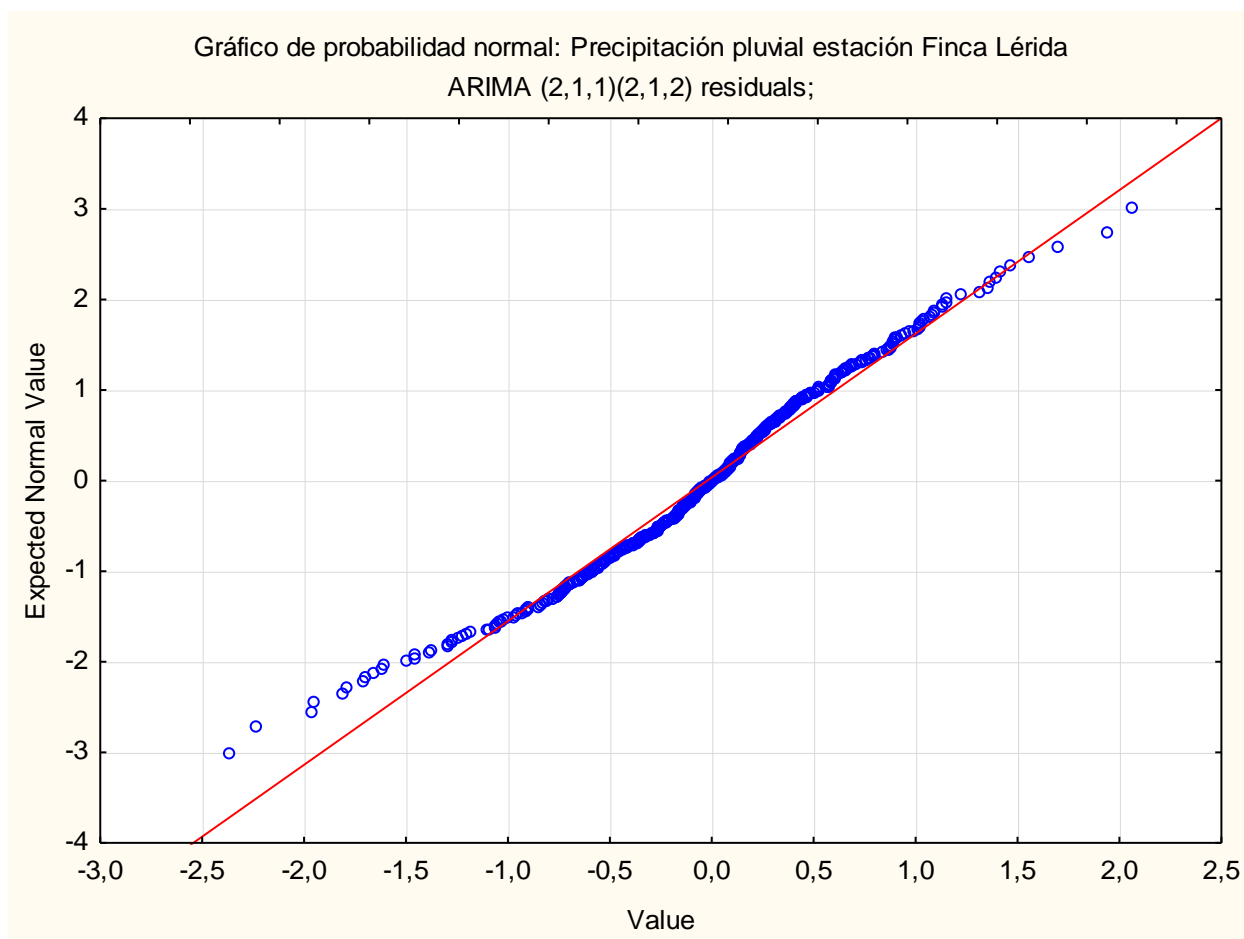
Luego se analizarán los parámetros del modelo observando el valor absoluto del estadístico “*t*” que sea menor que 2,0; y a la vez, que el *p-valor* sea mayor que 0,05. Este proceso nos permite elegir, que parámetro considerar o no, en el modelo (leer acápite d, p. 50).

También, se observará la FAS de los residuos, permitiendo una información análoga a la del contraste de aleatoriedad de *Ljung - Box* de manera gráfica; y se corrobora con la significancia estadística de *Ljung - Box*, con una confiabilidad del 95% (leer acápite e y f, p. 51).

**Gráfica 29: Histograma de residuales del modelo 1:  $ARIMA(2,1,1)(2,1,2)_{12}$  - estación Finca Lérída**



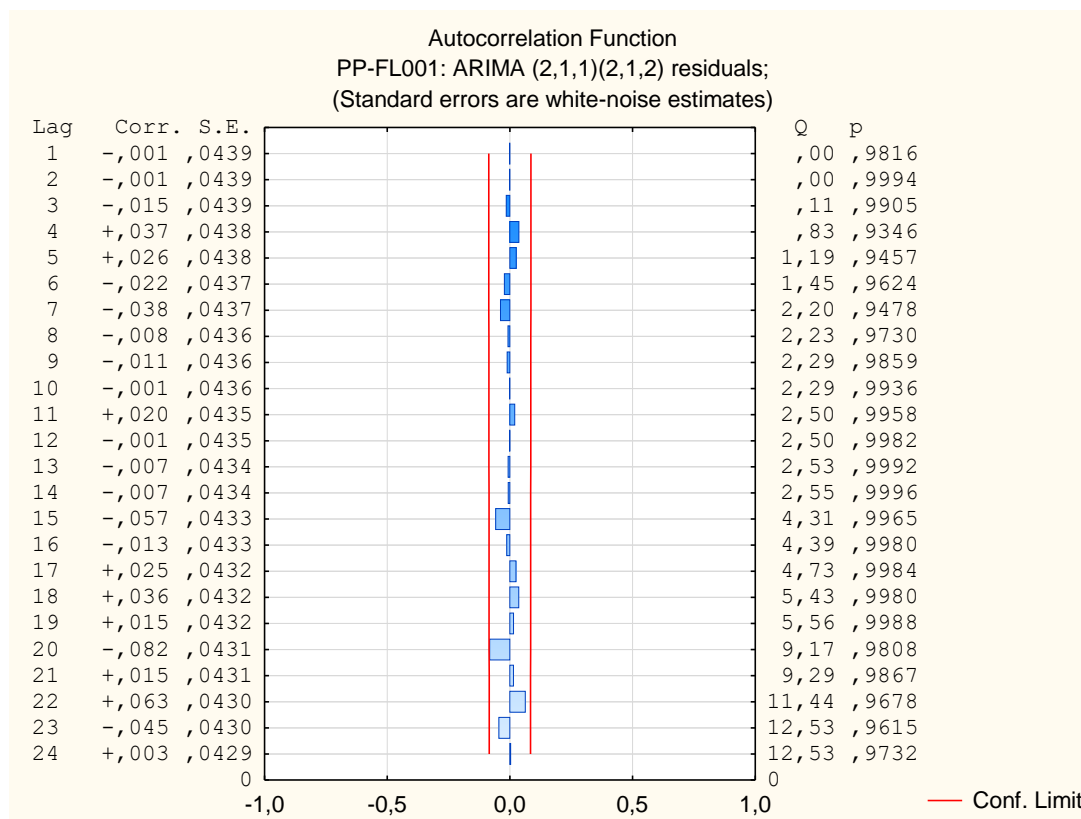
**Gráfica 30: Normalidad de residuales del modelo 1:  $ARIMA(2, 1, 1)(2, 1, 2)_{12}$  - estación Finca Lérica**



**Tabla 18: Parámetros estimados del modelo 1:  $ARIMA(2, 1, 1)(2, 1, 2)_{12}$  - estación Finca Lérica**

Input: PP-FL001 (PP-Estaciones(Tesis-2017).sta)						
Transformations: $\ln(x), D(1), D(12)$						
Model: (2,1,1)(2,1,2) Seasonal lag: 12 MS Residual= ,39279						
Paramet.	Param.	Asympt. Std.Err.	Asympt. t( 507)	p	Lower 95% Conf	Upper 95% Conf
Constant	-0,000672	0,000708	-0,94795	0,343604	-0,002063	0,000720
p(1)	0,051636	0,055790	0,92555	0,355120	-0,057971	0,161243
p(2)	0,014687	0,053102	0,27658	0,782215	-0,089640	0,119015
q(1)	0,859905	0,033221	25,88462	0,000000	0,794638	0,925172
Ps(1)	-0,210004	0,396996	-0,52898	0,597049	-0,989964	0,569956
Ps(2)	-0,082339	0,055750	-1,47694	0,140311	-0,191868	0,027190
Qs(1)	0,642523	0,396129	1,62200	0,105424	-0,135734	1,420780
Qs(2)	0,160064	0,344437	0,46471	0,642338	-0,516636	0,836764

**Gráfica 31: Autocorrelograma de los residuales del modelo 1:  
ARIMA (2, 1, 1)(2, 1, 2)<sub>12</sub> - estación Finca Lérica**



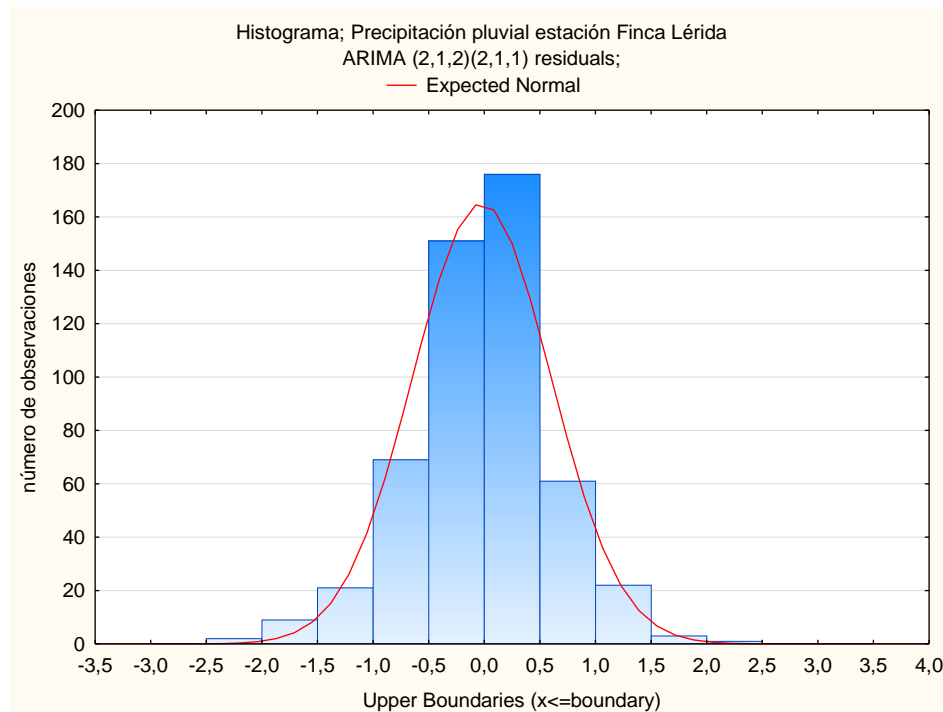
### **Análisis residual del modelo 1: ARIMA(2, 1, 1)(2, 1, 2)<sub>12</sub>**

- La **Gráfica 29** del histograma presenta como resultado, una campana Gaussiana.
- La gráfica de probabilidad normal, **Gráfica 30**, evidencia que los errores tienen una distribución normal aceptable.
- En el **Tabla 18** se observa que la mayoría de los coeficientes cumplen con los estadísticos de “*t*” y de “*p-valor*”; con excepción del parámetro ***q* = 1**, considerándose innecesario en el modelo.
- El Autocorrelograma de residuales (**Gráfica 31**) presenta una estructura de ruido blanco, para cada una de las barras ya que no sobresale ninguna de la banda de confianza; por lo tanto, se concluye que este modelo es adecuado.

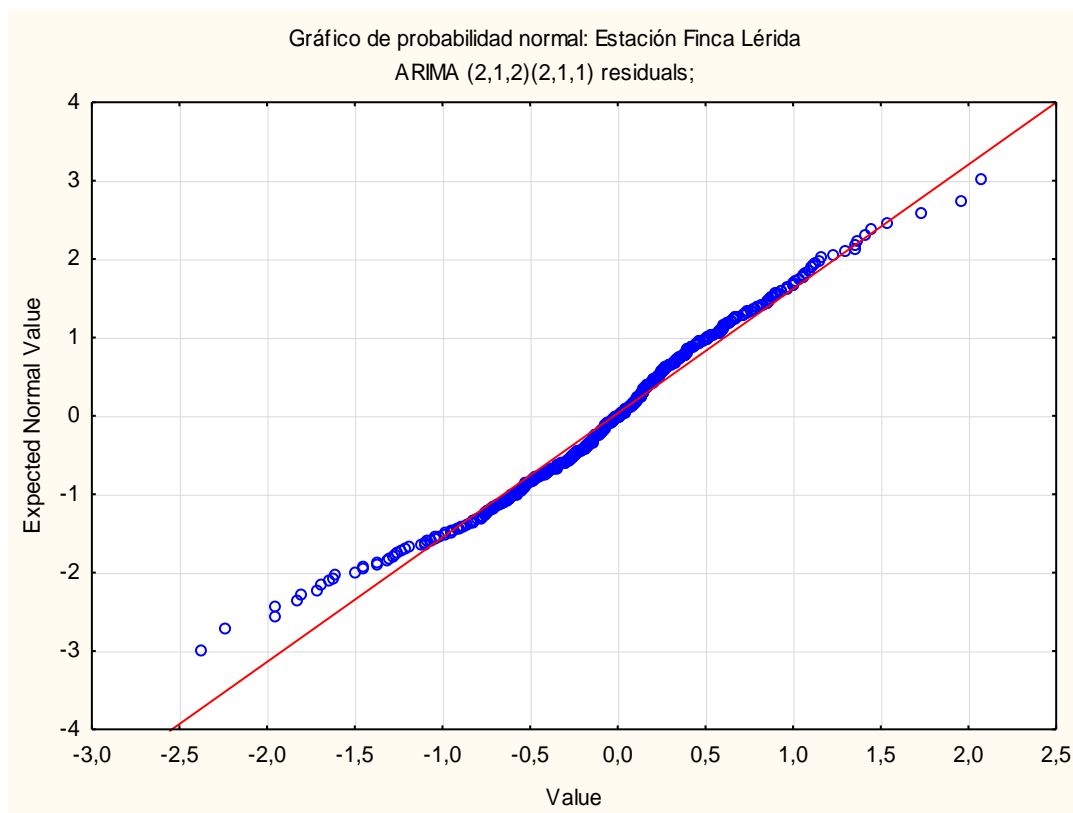
### Significancia estadística de Ljung - Box:

- Planteamiento de la hipótesis:
  - $H_0 : \rho_k = 0$
  - $H_1 : \rho_k \neq 0$
- Estadístico de prueba:  $Q_{(M=24)} = 12,53$ ; siendo  $M = 24$  el número de retardos (lags) utilizados en la *Gráfica 29*.
- Grados de libertad:  $h = 24 - 2 - 7 = 15$ .
- Región de Aceptación y de Rechazo: punto crítico es,  $\chi^2_{0,05;15} = 25,00$
- Regla de Decisión: Si el valor calculado  $Q_{(M=24)}$  es menor que el punto crítico o zona de rechazo  $\chi^2_{0,05;14}$ ;  $Q_{(M=24)} = 12,53 < \chi^2_{0,05;15} = 25,00$ ; entonces, no se rechaza la hipótesis nula ( $H_0$ ); concluyendo que se acepta  $H_0$  y se considerará el modelo  $ARIMA(2, 1, 1)(2, 1, 2)_{12}$  con una confiabilidad del 95%.

**Gráfica 32: Histograma de residuales del modelo 2:  $ARIMA(2, 1, 2)(2, 1, 1)_{12}$  - estación Finca Lérica**



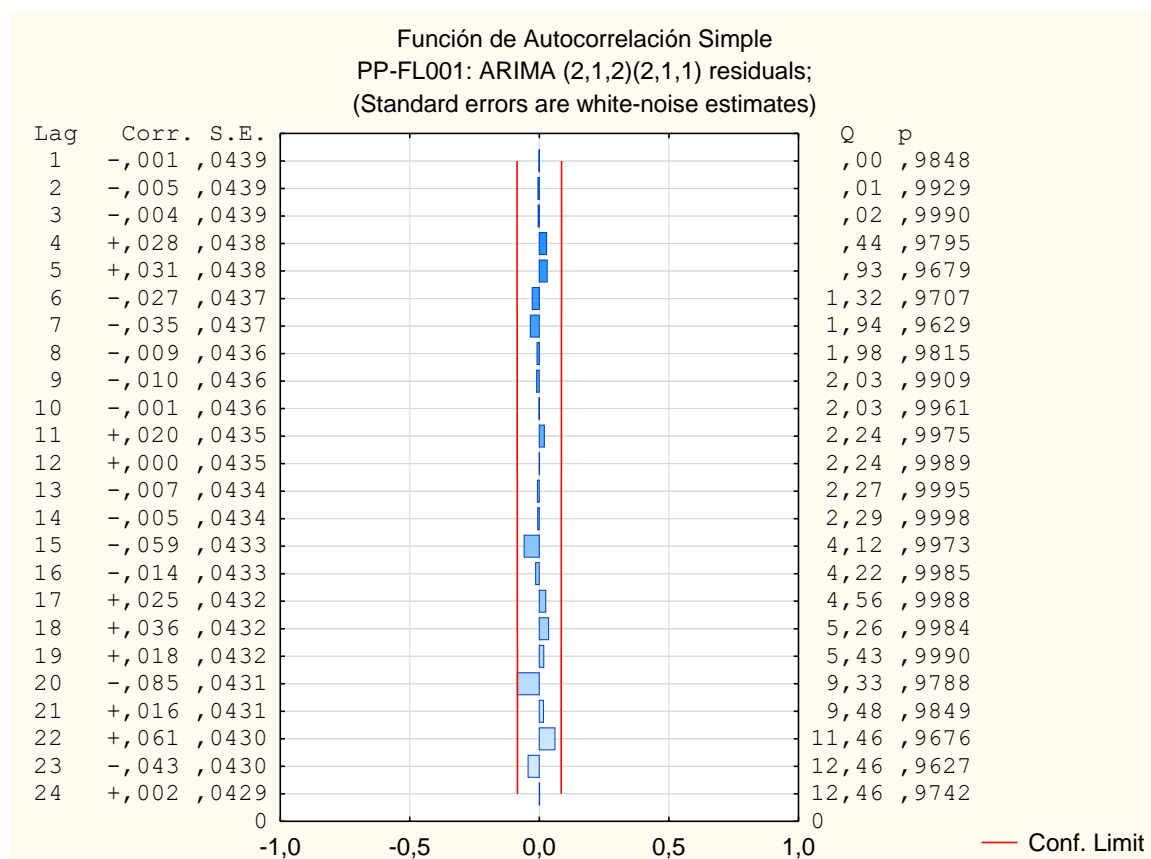
**Gráfica 33: Normalidad de residuales del modelo 2:  $ARIMA(2, 1, 2)(2, 1, 1)_{12}$  - estación Finca Lérída**



**Tabla 19: Estimados del modelo 2:  $ARIMA(2, 1, 2)(2, 1, 1)_{12}$  - estación Finca Lérída**

Input: Finca Lérída (Finca Lérída Oficial.sta)						
Transformations: In(x),D(1),D(12)						
Model:(2, 1, 2)(2, 1, 1) Seasonal lag: 12 MS Residual= ,35628						
Paramet.	Param.	Asympt. Std.Err.	Asympt. t( 434)	p	Lower 95% Conf	Upper 95% Conf
Constant	-0,000915	0,000850	-1,07677	0,282181	-0,00259	0,000755
p(1)	-0,518060	0,745715	-0,69472	0,487605	-1,98372	0,947602
p(2)	0,052820	0,057117	0,92477	0,355602	-0,05944	0,165080
q(1)	0,315064	0,745850	0,42242	0,672925	-1,15086	1,780992
q(2)	0,473087	0,632447	0,74803	0,454849	-0,76995	1,716128
Ps(1)	-0,223921	0,337424	-0,66362	0,507286	-0,88711	0,439267
Ps(2)	-0,108336	0,063930	-1,69461	0,090866	-0,23399	0,017314
Qs(1)	0,609126	0,335791	1,81400	0,070368	-0,05085	1,269105
Qs(2)	0,141009	0,283531	0,49733	0,619208	-0,41626	0,698274

**Gráfica 34: Autocorrelograma de los residuales del modelo 2:  
ARIMA(2, 1, 2)(2, 1, 1)<sub>12</sub> - estación Finca Lérica**



#### **Análisis residual del modelo 2: ARIMA(2, 1, 2)(2, 1, 1)<sub>12</sub>**

- La **Gráfica 32** del histograma presenta como resultado, una campana Gaussiana.
- La gráfica de probabilidad normal, **Gráfica 33**, evidencia que los errores tienen una distribución normal aceptable.
- En el **Tabla 19** se observa que todos los coeficientes cumplen con los estadísticos de “t” y de “p-valor”; considerándose válido el modelo.
- El autocorrelograma de residuales (**Gráfica 34**) presenta una estructura de ruido blanco, para cada una de las barras ya que no sobresale ninguna de la banda de confianza; por lo tanto, se concluye que el modelo es adecuado.

**Se corrobora el Ruido Blanco con la significancia estadística de Ljung - Box:**

➤ Planteamiento de la hipótesis:

- $H_0 : \rho_k = 0$
- $H_1 : \rho_k \neq 0$

➤ Estadístico de prueba:  $Q_{(M=24)} = 12,46$

➤ Grados de libertad:  $h = 24 - 2 - 7 = 15$

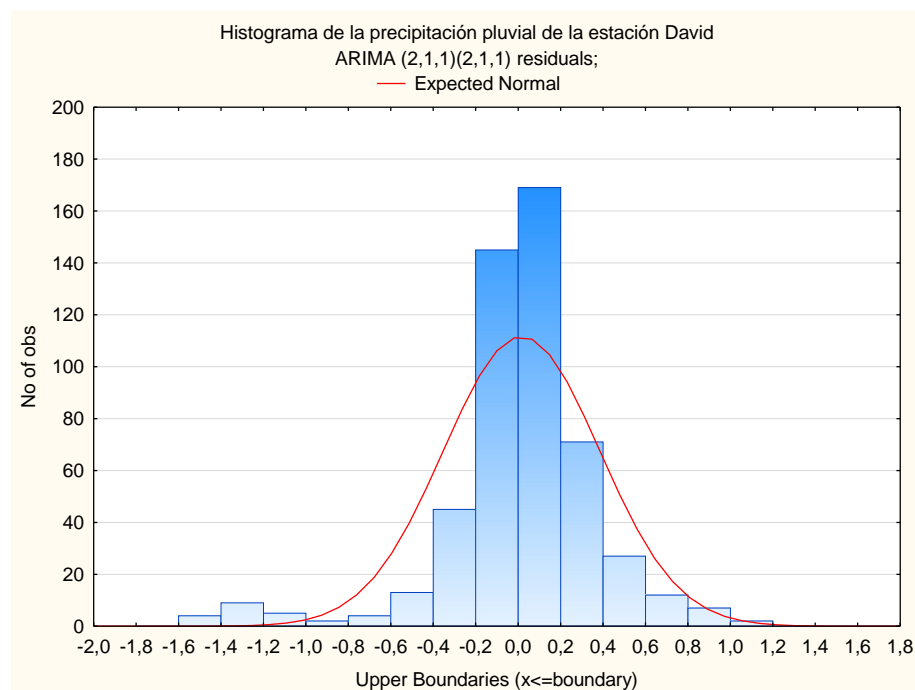
➤ Región de Aceptación y de Rechazo: punto crítico es,  $\chi^2_{0,05;15} = 25,00$

➤ Regla de Decisión: Si  $Q_{(M=24)} = 12,46 < \chi^2_{0,05;15} = 25,00$ ; entonces, no se rechaza la hipótesis nula ( $H_0$ ).

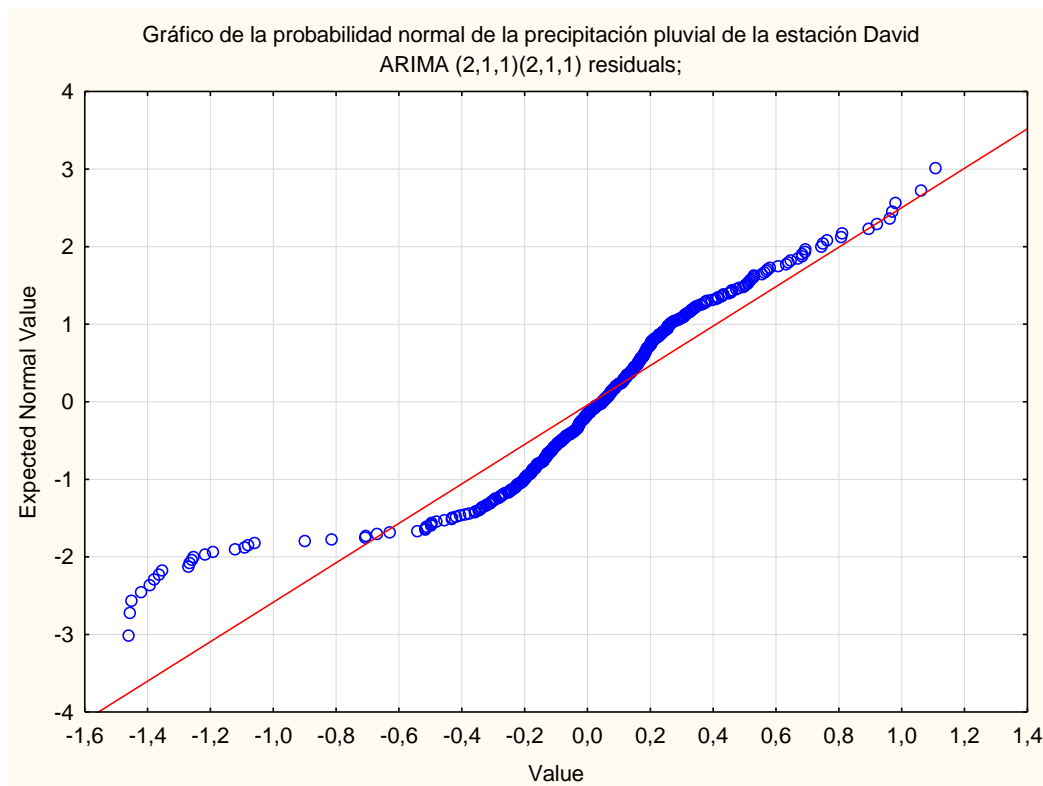
Así pues, se acepta la hipótesis nula ( $H_0$ ) y se toma como adecuado el modelo  $ARIMA(2, 1, 2)(2, 1, 1)_{12}$ , con una confianza del 95%.

**Estación David:**

**Gráfica 35: Histograma de residuales del modelo 1:  $ARIMA(2, 1, 1)(2, 1, 1)_{12}$  - estación David**



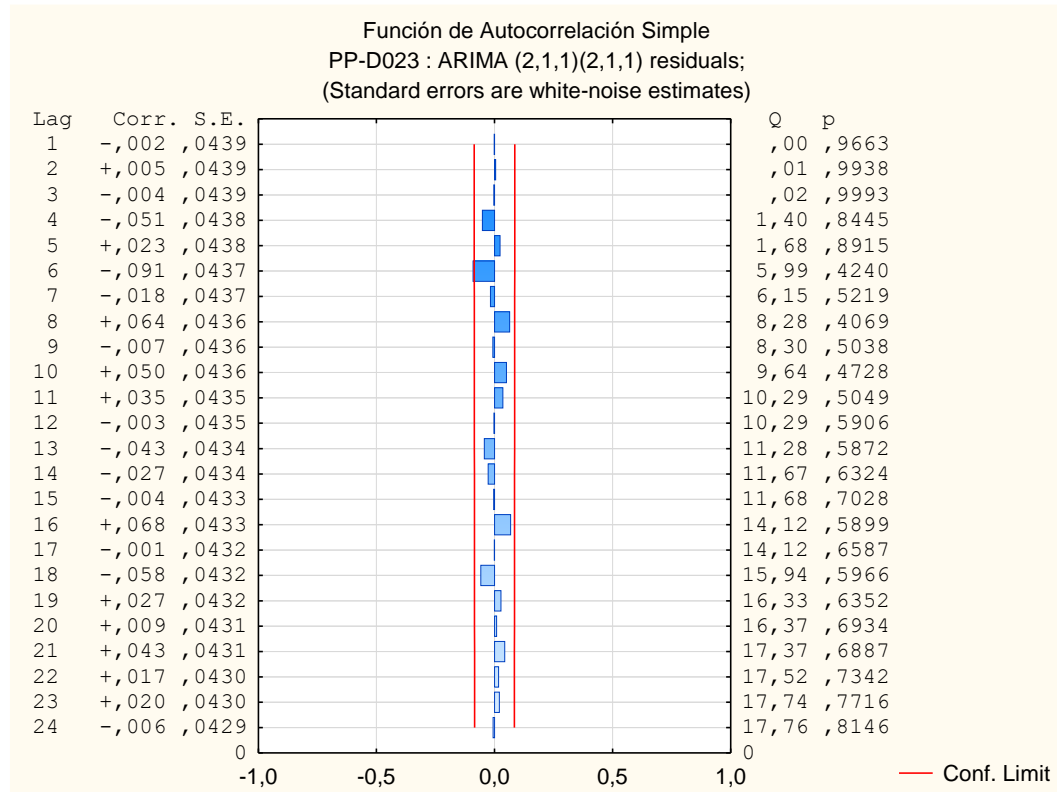
**Gráfica 36: Normalidad de residuales del modelo 1:  $ARIMA(2, 1, 1)(2, 1, 1)_{12}$  - estación David**



**Tabla 20: Parámetros de correlación y Estimados del modelo 1:  
 $ARIMA(2, 1, 1)(2, 1, 1)_{12}$  - estación David**

Transformations: $x^{\wedge}, 1500, D(1), \bar{D}(12)$						
Model: (2,1,1)(2,1,1) Seasonal lag: 12 MS Residual= ,13738						
Paramet.	Param.	Asympt. Std.Err.	Asympt. t( 508)	p	Lower 95% Conf	Upper 95% Conf
Constant	0,000292	0,000284	1,02914	0,303906	-0,000265	0,000850
p(1)	0,019698	0,049666	0,39660	0,691829	-0,077879	0,117275
p(2)	0,106924	0,048315	2,21308	0,027336	0,012003	0,201845
q(1)	0,903468	0,020770	43,49812	0,000000	0,862661	0,944274
Ps(1)	-0,158537	0,054270	-2,92125	0,003641	-0,265160	-0,051915
Ps(2)	-0,027850	0,052810	-0,52736	0,598173	-0,131603	0,075903
Qs(1)	0,834331	0,031418	26,55566	0,000000	0,772606	0,896057

**Gráfica 37: Autocorrelograma de los residuales del modelo 1:**  
**ARIMA(2, 1, 1)(2, 1, 1)<sub>12</sub> - estación David**



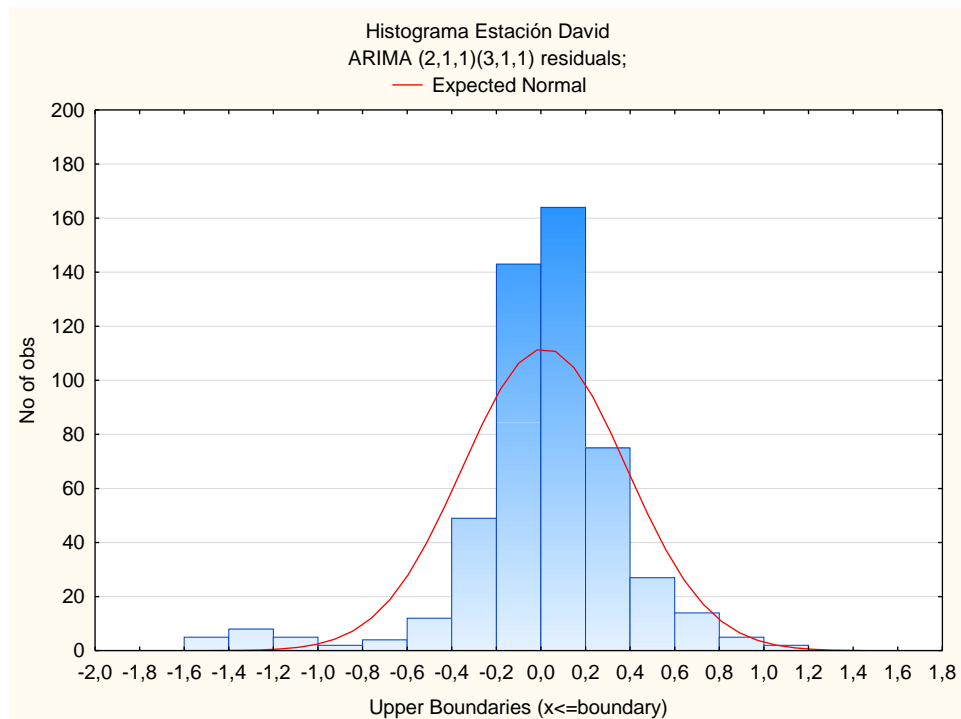
#### Análisis residual del Modelo 1: ARIMA(2, 1, 1)(2, 1, 1)<sub>12</sub>

- La Gráfica 35 del histograma presenta como resultado, una forma de campana Gaussiana sesgada a la izquierda, pero mantiene una distribución con media cero y varianza constante considerándose aceptable.
- La gráfica de probabilidad normal, Gráfica 36, evidencia que los errores tienen una distribución aproximadamente normal, aunque sesgada a la izquierda, se le considera aceptable.
- En el Tabla 20 se observa que la mayoría de los coeficientes no cumplen con los estadísticos de “t” y de “p-valor”; con excepción de  $p = 1$  y  $P = 2$ .
- El autocorrelograma de residuales (Gráfica 37) presenta estructura de ruido blanco, para cada una de las barras ya que no sobresale ninguna de la banda de confianza.

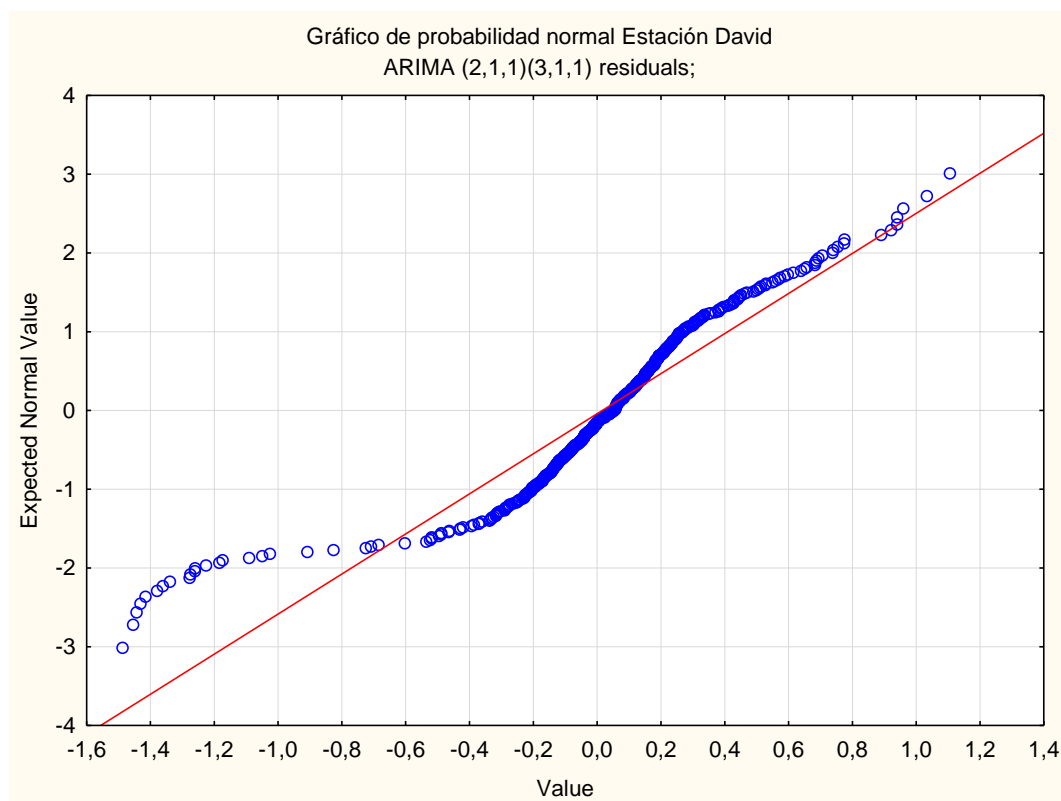
### Significancia estadística de Ljung - Box:

- Planteamiento de la hipótesis:
  - $H_0 : \rho_k = 0$
  - $H_1 : \rho_k \neq 0$
- Estadístico de prueba:  $Q_{(M=24)} = 17,76$ ; siendo  $M = 24$  el número de retardos (lags) utilizados en la Gráfica 35.
- Grados de libertad:  $h = 24 - 2 - 6 = 16$ .
- Región de Aceptación y de Rechazo: punto crítico es,  $\chi^2_{0,05;16} = 26,30$
- Regla de Decisión: Si el valor calculado  $Q_{(M=24)}$  es menor que el punto crítico o zona de rechazo  $\chi^2_{0,05;16}$ ;  $Q_{(M=24)} = 17,76 < \chi^2_{0,05;16} = 26,30$ ; entonces, no se rechaza la hipótesis nula ( $H_0$ ) . Así pues, se acepta la hipótesis nula ( $H_0$ ) y se toma como adecuado el modelo  $ARIMA(2, 1, 1)(2, 1, 1)_{12}$ , con una confianza del 95%.

**Gráfica 38: Histograma de residuales del modelo 2:  $ARIMA(2, 1, 1)(3, 1, 1)_{12}$  - estación David**



**Gráfica 39: Normalidad de residuales del modelo 2:  $ARIMA(2, 1, 1)(3, 1, 1)_{12}$  - estación David**

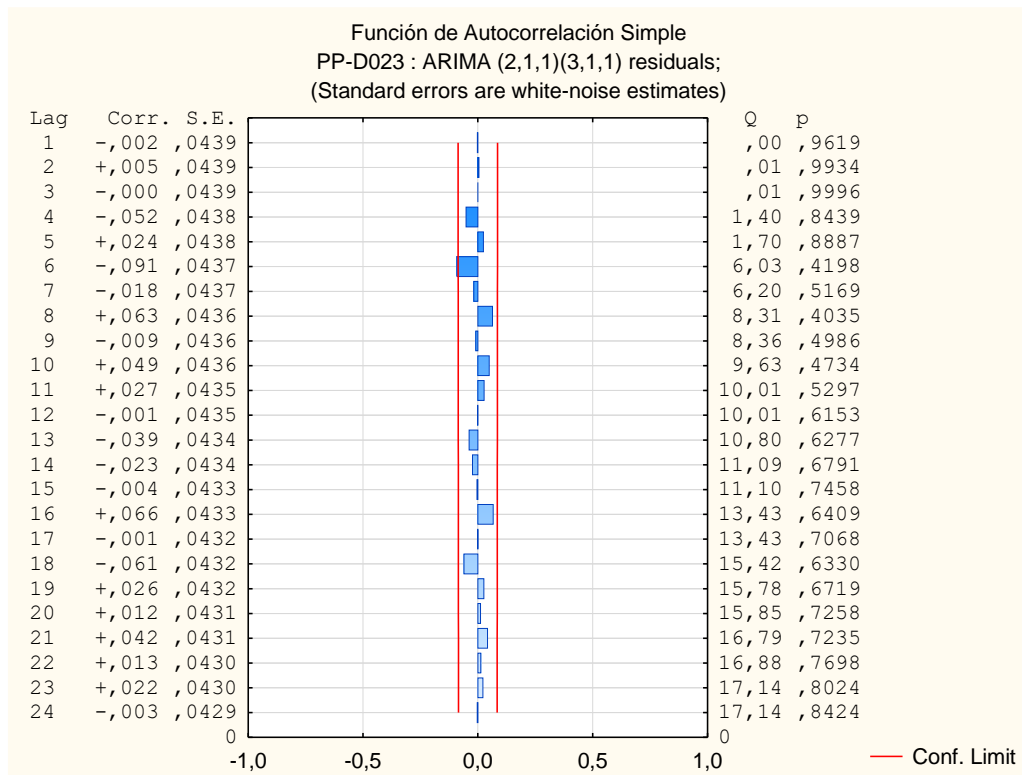


**Tabla 21: Parámetros de correlación y Estimados del modelo 2:  
 $ARIMA(2, 1, 1)(3, 1, 1)_{12}$  - estación David**

Transformations:  $x^A, 1500, D(1), \bar{D}(12)$   
Model: (2,1,1)(3,1,1) Seasonal lag: 12 MS Residual= ,13735

Paramet.	Param.	Asympt. Std.Err.	Asympt. t( 507)	p	Lower 95% Conf	Upper 95% Conf
Constant	0,000282	0,000292	0,96396	0,335525	-0,000292	0,000855
p(1)	0,022553	0,049882	0,45213	0,651366	-0,075447	0,120554
p(2)	0,106991	0,048397	2,21069	0,027503	0,011908	0,202074
q(1)	0,903017	0,020947	43,11010	0,000000	0,861864	0,944170
Ps(1)	-0,180532	0,059233	-3,04782	0,002425	-0,296905	-0,064159
Ps(2)	-0,053733	0,059009	-0,91059	0,362944	-0,169666	0,062200
Ps(3)	-0,056435	0,053512	-1,05463	0,292097	-0,161568	0,048697
Qs(1)	0,813800	0,040439	20,12426	0,000000	0,734352	0,893249

**Gráfica 40: Autocorrelograma de los residuales del modelo 2:  $ARIMA(2, 1, 1)(3, 1, 1)_{12}$  - estación David**



#### Análisis residual del Modelo 2: $ARIMA(2, 1, 1)(3, 1, 1)_{12}$

- La Gráfica 38 del histograma se presenta como resultado, una campana Gaussiana sesgada a la izquierda, pero mantiene una distribución con media cero y varianza constante, considerándose aceptable.
- La gráfica de probabilidad normal, Gráfica 39, evidencia que los errores tienen una distribución aproximadamente normal, aunque sesgada a la derecha, se le considera aceptable.
- En el Tabla 21 se observa que la mayoría de los coeficientes no cumplen con los estadísticos de “ $t$ ” y de “ $p$ -valor”; con excepción de  $p = 1$ ,  $P = 2$  y  $P = 3$
- El autocorrelograma de residuales (Gráfica 40) presenta estructura de ruido blanco, pero cada una de las barras con excepción del retardo 6, que sobresale de la banda de confianza, estipulado por  $\pm 2\sqrt{456} \approx 0,094$ .

**Se corrobora el Ruido Blanco, con la significancia estadística de Ljung-Box:**

- Planteamiento de la hipótesis:
  - $H_0 : \rho_k = 0$
  - $H_1 : \rho_k \neq 0$
- Estadístico de prueba:  $Q_{(M=24)} = 17,14$
- Grados de libertad:  $h = 24 - 2 - 7 = 15$
- Región de Aceptación y de Rechazo: punto crítico es,  $\chi^2_{0,05;15} = 25,00$
- Regla de Decisión: Si  $Q_{(M=24)} = 17,14 < \chi^2_{0,05;15} = 25,00$ ; entonces, no se rechaza la hipótesis nula ( $H_0$ ).

Así pues, se acepta la hipótesis nula ( $H_0$ ) y se toma como adecuado el modelo  $ARIMA(2, 1, 1)(3, 1, 1)_{12}$ , con una confianza del 95%.

**5.3.2.7. Análisis estadístico del error y selección del modelo**

Luego de haberse validado los modelos propuestos, se analizan los errores estadísticos: error cuadrado medio (ECM), raíz cuadrada del error medio (RCEM) y porcentaje de error medio absoluto (PEMA), para cada modelo de las estaciones en estudio. Se determina si el modelo producirá errores de predicción que se juzguen como suficientemente pequeños. Es normal esperar que un modelo produzca errores de pronóstico relativamente bajos sobre una base consistente; sin embargo, al trabajar con la precipitación pluvial se espera que estos sean relativamente altos por la alta variabilidad del mismo.

En cuanto a la selección del modelo para cada estación, se toma en cuenta el de mayor confiabilidad en la validez (estadístico “t”, p-valor, y Ruido Blanco de la FAS) y análisis del error del modelo; no obstante, en último término, la elección del modelo final debe ser un equilibrio entre simplicidad y eficacia, haciendo uso del criterio de parsimonia.

**Tabla 22: Estadísticas de error y análisis residual de los modelos propuestos - estación Finca Lérica**

Modelos	ECM	RCEM	PEMA
<i>ARIMA</i> (1, 1, 1)(1, 1, 1) <sub>12</sub>	0,39236	0,62639	35,04%
<i>ARIMA</i> (2, 1, 2)(2, 1, 1) <sub>12</sub>	<b>0,39279</b>	<b>0,62673</b>	<b>34,87%</b>

En la Tabla 22, se muestra que de los dos modelos, ambos presentan una pequeña variación en los análisis ECM, RCEM y con respecto al análisis de PEMA, el de menor porcentaje fue el modelo *ARIMA*(2, 1, 2)(2, 1, 1)<sub>12</sub>; además, este modelo cumple con todos los contrastes de validez satisfactoriamente y se elige como el mejor modelo para la estación Finca Lérica.

**Tabla 23: Estadísticas de error y análisis residual de los modelos propuestos - estación David**

Modelos	ECM	RCEM	PEMA
<i>ARIMA</i> (2, 1, 1)(2, 1, 1) <sub>12</sub>	<b>0,13738</b>	<b>0,37065</b>	<b>29,09%</b>
<i>ARIMA</i> (2, 1, 1)(3, 1, 1) <sub>12</sub>	0,13735	0,37061	29,03%

En el Tabla 23, se muestra que de los dos modelos, ambos presentan una variación insignificante en los análisis ECM, RCEM y con respecto al análisis de PEMA; sin embargo, se contempla el modelo *ARIMA*(1, 1, 2)(2, 1, 2)<sub>12</sub> por hacer uso de menos parámetros; y se corrobora con el contraste de validez, para seleccionarlo como el mejor para la estación David.

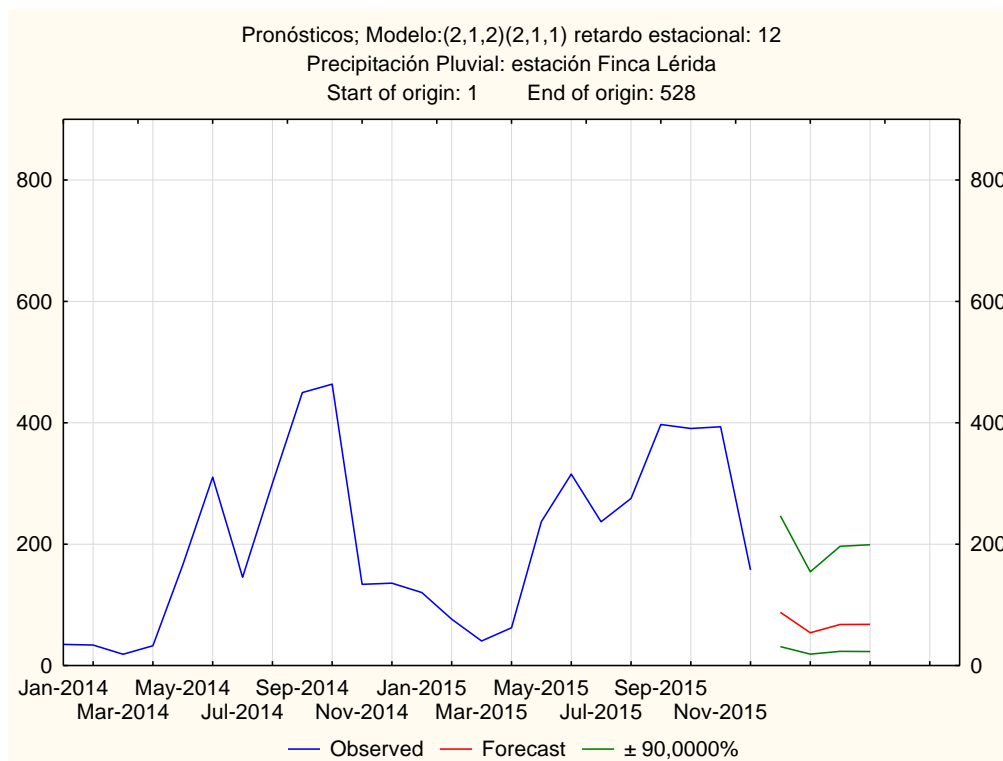
### 5.3.2.8. Pronósticos

Luego de haber cumplido todos los pasos estipulados y haber escogido el mejor modelo que presente un equilibrio entre simplicidad y eficacia, se finaliza proponiendo la predicción para el primer cuatrimestre del año 2016.

**Tabla 24: Pronósticos del modelo  $ARIMA(2, 1, 2)(2, 1, 1)_{12}$  en el primer cuatrimestre del 2016 para la estación Finca Lérica**

N°	Mes	Pronóstico	90% menor	90% mayor
529	ene	87,81941	31,26572	246,6679
530	feb	53,98874	18,86181	154,5336
531	mar	67,64526	23,29515	196,4307
532	abr	67,91021	23,15698	199,1536

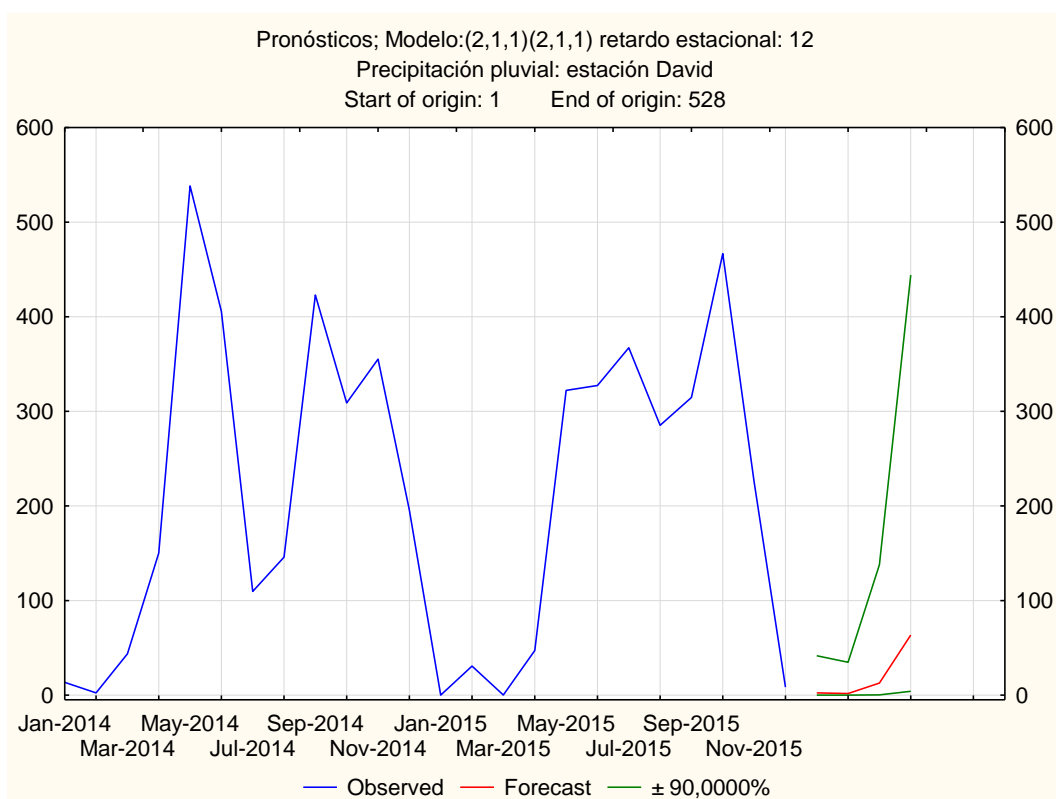
**Gráfica 41: Pronósticos del modelo  $ARIMA(2, 1, 2)(2, 1, 1)_{12}$  en el primer cuatrimestre del 2016 para la estación Finca Lérica**



**Tabla 25: Pronósticos del modelo  $ARIMA(2, 1, 1)(2, 1, 1)_{12}$  en el primer cuatrimestre del 2016 para la estación David**

N°	Mes	Pronóstico	90% menor	90% mayor
529	ene	2,38481	0,014238	41,7159
530	feb	1,75287	0,006793	34,7509
531	mar	12,80988	0,308773	137,8114
532	abr	63,49797	4,028922	444,0455

**Gráfica 42: Pronósticos del modelo  $ARIMA(2, 1, 1)(2, 1, 1)_{12}$  en el primer cuatrimestre del 2016 para la estación David**



En la [Tabla 24](#) y [25](#), se presenta los pronósticos del primer cuatrimestre para el año 2016 con los intervalos de confianza inferior y superior al 90% para cada estación

pluviométrico en estudio. Estos pronósticos se aprecian mejor en la Gráfica 41 y 42, visualizando el comportamiento de lluvia de los dos años previos 2014, 2015 y el pronóstico para el mes de enero, febrero, marzo y abril para el año 2016 con sus intervalos de confianza inferior y superior al 90% de los modelos para cada estación.

### 5.3.3. Comparación de los datos pronosticados versus los reales del año 2016

La **Tabla 26** que se presenta a continuación, contiene los datos reales de las cuatro estaciones en estudio, facilitada por la unidad de meteorología de la Empresa de Transmisión Eléctrica, S.A en el año 2017. La presentación de esta información surge con el objetivo de verificar la validez de los datos pronosticados del primer cuatrimestre del 2016 con el uso de la metodología Box - Jenkins versus los datos reales.

**Tabla 26: Total de precipitación pluvial del primer cuatrimestre para las estaciones meteorológicas en estudio, año: 2016**

mes	Estaciones meteorológicas (mm)	
	Finca Lérica	David
ene	S/D	2,5
feb	S/D	1,5
mar	S/D	20,9
abr	S/D	61,5

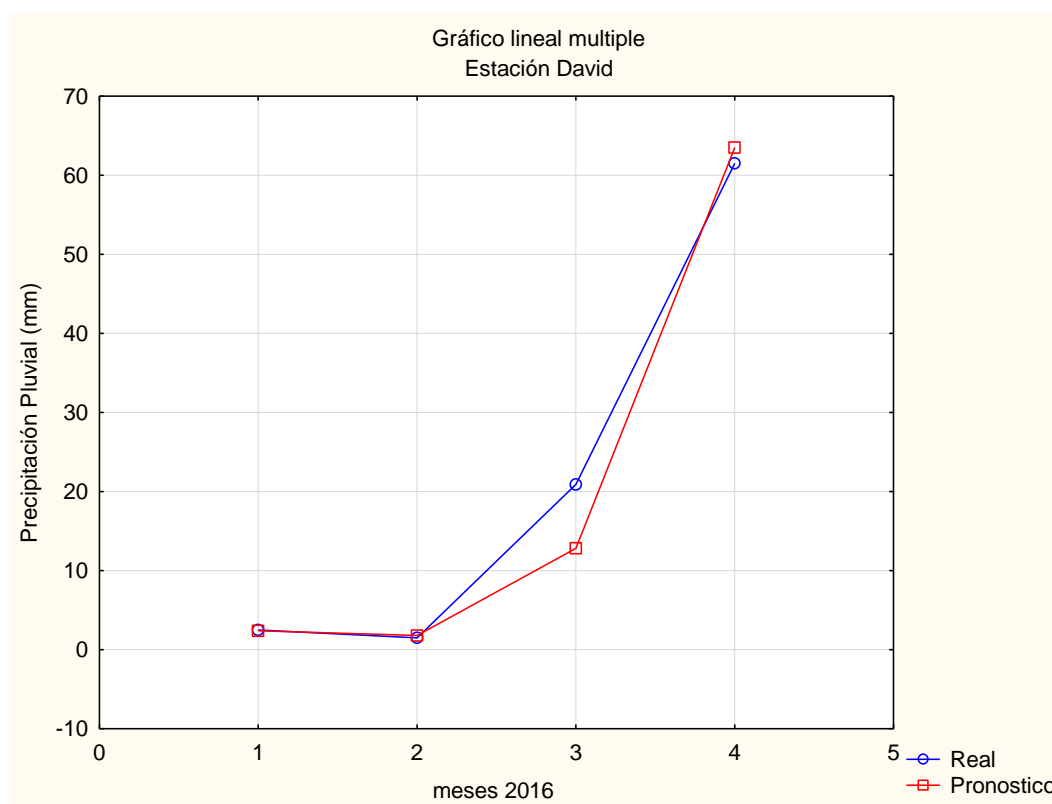
Fuente: Gerencia de Hidrometeorología de la Empresa de Transmisión Eléctrica S.A. (ETESA).

Se puede observar que los registros pluviométricos para la estación Finca Lérica son desconocidos (sin dato). De acuerdo a la Gerencia de Hidrometeorología de la Empresa de Transmisión Eléctrica S.A. (ETESA), el pluviómetro presentó problemas técnicos el cual no lograron registrar la acumulación de precipitación pluvial por varios meses del año 2016. Por tal motivo, solo se procederá al análisis comparativo para la estación David.

**Tabla 27: Comparación de pronósticos del modelo  $ARIMA(2, 1, 1)(2, 1, 1)_{12}$  versus datos reales del primer cuatrimestre, año: 2016 - estación David**

mes	Precipitación pluvial (mm)	
	Real	Pronóstico
ene	2,5	2,4
feb	1,5	1,8
mar	20,9	12,8
abr	61,5	63,5

**Gráfica 43: Comparación de pronósticos del modelo  $ARIMA(2, 1, 1)(2, 1, 1)_{12}$  versus datos reales del primer cuatrimestre, año: 2016 - estación David**



Considerando la *Gráfica 43*, se observa que los valores pronosticados presentan una tendencia considerable con respecto a los datos reales. Además, muestran buena precisión en los datos pronosticados con respecto a los reales; por tanto, se considera como bueno el modelo  $ARIMA(2, 1, 1)(2, 1, 1)_{12}$ .

## 6. CONCLUSIONES Y RECOMENDACIONES

Al reflexionar el desarrollo de este trabajo, se puede considerar que los estudios meteorológicos permiten que las instituciones privadas y gubernamentales puedan tener una mejor comprensión y análisis del comportamiento meteorológico en el territorio nacional; y con ayuda de nuevas líneas de investigación, se podría desarrollar cada vez más protocolos de contingencia contra problemáticas como son inundaciones, desbordamiento de ríos, deslizamientos de laderas, sequías entre otras.

Este documento, refleja las fundamentaciones de la investigación, con ayuda de artículos, libros y otros documentos que se consideran como básicos para poseer un panorama actual de la temática “series de tiempo”, que se encuentra ligado a la metodología Box - Jenkins.

Además, se coincide con otras investigaciones como la de Cruz & García (2008); Verón et al.(n.d.) Y Medinaceli (2007); lo cual consisten en analizar las series temporales de diferentes factores meteorológicos para generar modelos y así contemplarlo como modelo predictivo.

Además, en el trabajo, se presenta la aplicación de la metodología Box-Jenkins que propone una metodología de modelización de las series, es decir, un conjunto de reglas para construir modelos que reproduzcan satisfactoriamente el comportamiento de serie temporales reales. Resultan muy interesantes y significativos que los modelos ARIMA que se generan con la aplicación de la metodología Box - Jenkins logren ajustarse significativamente a la serie real, permitiendo que el modelo presente una perspectiva retrospectiva, actual y predictiva; todo esto, sin contemplar los posibles factores externos presentes en los registros pluviométricos, de los cuales se conocen que son

varios. Es decir, que la metodología Box - Jenkins debe contar con suficientes datos; la misma, se encuentra muy vinculada al pasado y no considera el efecto de variables externas como fenómeno de El Niño y La Niña, frentes fríos e igualmente efectos de tormentas tropicales. Este hecho resulta ser una fuerte limitación del modelo; sin embargo, puede ser útil para proyecciones de corto plazo, las cuales son susceptibles de actualizarse año con año.

### 6.1. En relación a las preguntas-objetivos del trabajo

En el estudio descriptivo de las dos estaciones en estudio, la estación Finca Lérica es la que presenta mayor promedio histórico de precipitación pluvial con 2 780,1 mm, en cambio, la estación David presenta 2 585,6 mm de precipitación pluvial. Además, el registro máximo ha sido de 4 224,3 mm ocurrido en el año 2010 en la estación Finca Lérica, el mayor registrado en los 44 años de estudios; sin embargo, para la estación David su registro máximo fue de 3 623,7 mm en el año 2008 de acuerdo a los registros pluviométricos que corresponde de 1972 hasta el 2015.

También, en la Gráficas 5 y 6 (p. 65, 67), se muestran los registros de lluvia donde los meses de *mayo* y *octubre* registran los valores máximos, los cuales indican que para nuestro país la precipitación tiene un comportamiento *bimodal*.

Se verificó la validez de todos los datos históricos, de las estaciones, presentando una precisión y consistencia dentro del rango de aceptación con una *confiabilidad del 95%* utilizando la prueba de Rachas con su respectiva significancia estadística y el método de Dobles acumuladas corroborándose en las Gráficas 1 y 2 (p. 61-61).

Se encontró que las series de tiempo de la variable (precipitación pluvial) de cada estación no son estacionarias, debido a que todas poseen una tendencia; por lo tanto, no se consideran constantes en media ni en varianza. Además, todas las series fueron aleatorias y las mismas no presentan ciclos.

También, se muestra que las amplitudes de las diferentes series varían con un alto grado de variabilidad estacional y regular con respecto al tiempo, seleccionándose así un modelo de descomposición multiplicativo como lo es el autorregresivo integral de media móvil estacional multiplicativo -  $ARIMA(p, d, q)(P, D, Q)_s$ , como propuesta del modelo a utilizar.

Todos los modelos fueron sometidos a rigurosos y variados análisis estadísticos para contrastar su validez, y así elegir el mejor modelo que se ajuste con una confiabilidad del 95% con el uso de la prueba estadística “Ljung - Box”; sin embargo, es importante señalar que se perdió un poco el criterio de parsimonia ya que fueron muchos los parámetros estimados en los modelos propuestos. El modelo de la estación David ubicada en la zona baja de la cuenca del Río Chiriquí, se ajusta mejor que de la zona alta; una posible causa sería la variabilidad de precipitación que posee la zona alta, dificultando un mejor ajuste en el modelo.

Se identificó que para la estación Finca Lérída el modelo que mejor se ajusta es  $ARIMA(2, 1, 2)(2, 1, 1)_{12}$  y para la estación David el modelo  $ARIMA(2, 1, 1)(2, 1, 1)_{12}$  es el que mejor se ajusta; siendo estos modelos, utilizados para los pronósticos de la precipitación pluvial del primer cuatrimestre del año 2016, para cada una de las estaciones pluviométricas en estudios (Finca Lérída y David) ubicadas en la zona alta y baja de la cuenca del río Chiriquí.

Luego de haber cumplido todos los pasos estipulados y haber escogido el mejor modelo que presente un equilibrio entre simplicidad y eficacia, se genera los pronósticos para el primer cuatrimestre del año 2016 con un 95% de confianza, siendo los siguientes:

Pronóstico de Precipitación Pluvial (mm) Precipitación pluvial (mm)

N°	Mes	Estación Finca Lérica	Estación David	Dato real estación David
529	ene	87,81941	2,38481	2,4
530	feb	53,98874	1,75287	1,8
531	mar	67,64526	12,80988	12,8
532	abr	67,91021	63,49797	63,5

Al comparar los datos pronosticados del modelo de la estación David con respecto a los valores reales del primer cuatrimestre del año 2016, se analizó que el modelo propuesto posee tendencias considerables y muy aceptables con respecto a los valores reales (p. 106). Es relevante mencionar que, los registros pluviométricos reales del año 2016 para la estación Finca Lérica son desconocidos (sin dato). De acuerdo a la Gerencia de Hidrometeorología de la Empresa de Transmisión Eléctrica S.A. (ETESA), el pluviómetro presentó problemas técnicos el cual no lograron registrar la acumulación de precipitación pluvial por varios meses del año 2016.

Las estadísticas de error confirman la confiabilidad y eficacia de los modelos propuestos donde la estación que presentaba menor porcentaje de error medio absoluto (PEMA) fue la de David con 29,09%, y la que presentó mayor porcentaje de error medio absoluto (PEMA) fue la estación Finca Lérica con 34,87%. Recordando, se analiza una variable muy volátil como lo es la precipitación pluvial y con todo esto, el modelo se ajusta bastante a lo ocurrido en el primer cuatrimestre del 2016.

## 6.2. Perspectivas a futuro

Se espera que con la utilización de la metodología Box - Jenkins pueda ser implementada para diferentes estudios que involucren factores hidrometeorológicos en otras regiones; y a su vez, considerarlo como alternativa como modelo predictivo con la finalidad de prevenir eventos adversos contra la agricultura, hidroeléctricas, ganadería entre otras.

Además, para la problemática de la varianza no constante se puede realizar un estudio utilizando modelos más avanzados como lo son los modelos ARCH y GARCH con la aplicación de la metodología Box - Jenkins, ya que estos sí trabajan con varianzas no constantes y permiten omitir la transformación de la serie.

También sería conveniente un modelo como el autorregresivo integrado de media móvil vectorial (ARIMAV) que considere la variable adicional que tome en cuenta los fenómenos de El Niño y La Niña; al igual que fenómenos climatológicos como depresiones tropicales, frente fríos entre otros.

Para la utilización de los modelos ARCH, GARCH y ARIMAV, se podría utilizar otro software avanzado y especializado, como el Eviews.

A raíz de este esfuerzo, en lo personal, se debe compartir toda esta información con la comunidad científica ante una publicación, al igual que para entidades gubernamentales y privadas interesadas en ésta temática con la finalidad que tengan a su saber, otra opción a contemplar en sus tomas de decisiones.

## 7. REFERENCIAS BIBLIOGRÁFICAS

- Breña Moral, J. A. (n.d.). *Análisis de Series Temporales*.
- Casado de Lucas, D. (2006). *Protocolo para la identificación de modelos ARIMA en series temporales*. Madrid. Retrieved from [www.est.uc3m.es/dcasado/](http://www.est.uc3m.es/dcasado/)
- Castro F., E. (2008). *Manual de Procedimientos para las Estaciones Meteorológicas*. Sarapiquí.
- Centro de aguas para zonas áridas y semiáridas de ALC-CAZALAC (2008). *Manual breve de preparación de datos*. CAZALAC.
- Cruz, C. E.; Ramírez, A. A. (2010). *Perspectiva mensual de la lluvia para la cuenca del Río Chiriquí (108)*. Meta de Meteorología 2010: Pronóstico cuantitativo de lluvia mensual con el CPT en Fortuna. Gerencia de Hidrometeorología de la Empresa de Transmisión Eléctrica S.A. (ETESA).
- Cruz, C. E. (2007). *Protocolo para identificar modelos ARIMA y pronósticos con series temporales*. Maestría en estadística aplicada. Universidad de Panamá. Facultad de Ciencias Naturales Exactas y Tecnología. Escuela de Estadística. Panamá.
- Cruz M., I. R., & García P., F. (2008). *MODELACIÓN DE LOS ESCURRIMIENTOS DEL RÍO YAQUI*. *Terra Latinoamericana*, 26, 145-152.
- De Batista, R.; Castillo, I. (1999). *Actualización del diagnóstico de los recursos naturales existentes en la cuenca del Río Caldera*. Tesis para optar por licenciatura. Universidad Autónoma de Chiriquí. Escuela de Geografía e Historia. Panamá.
- Gerencia de Hidrometeorología de la Empresa de Transmisión Eléctrica S.A. (ETESA). (2009). *Open Data - Estaciones meteorológicas*. Retrieved February 18, 2017, from

[http://www.hidromet.com.pa/estaciones\\_meteo.php?f\\_cuenca=RIO+CHIRIQUI&f\\_status=2&f\\_institucion=noopcion&f\\_latitud\\_1=0&f\\_latitud\\_2=0&f\\_longitud\\_1=0&f\\_longitud\\_2=0&f\\_provincia=CHIRIQUI&f\\_distrito=noopcion&f\\_corregimiento=noopcion&f\\_nombre=](http://www.hidromet.com.pa/estaciones_meteo.php?f_cuenca=RIO+CHIRIQUI&f_status=2&f_institucion=noopcion&f_latitud_1=0&f_latitud_2=0&f_longitud_1=0&f_longitud_2=0&f_provincia=CHIRIQUI&f_distrito=noopcion&f_corregimiento=noopcion&f_nombre=)

González, M. P. (Universidad del P. V. (2007). *Análisis de series temporales: modelos*

*ARIMA*. Sarriko-on. Retrieved from

<https://books.google.com.co/books?id=ftqrKnA7dTMC>

Gracia S., A., Vázquez C., D., Reyes O., H., Sáenz S., A., & Limón L., A. (2009).

*Investigación en el ámbito empresarial “Pronósticos , Supervisión e Indicadores Financieros”* (Electrónica, Vol. 1). Retrieved from

[http://www.eumed.net/libros/2009b/543/TIPOS DE PRONOSTICOS EN LA INVESTIGACION EN EL AMBITO EMPRESARIAL.htm](http://www.eumed.net/libros/2009b/543/TIPOS_DE_PRONOSTICOS_EN_LA_INVESTIGACION_EN_EL_ambito_EMPRESARIAL.htm)

Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, P. (2006). *Analisis*

*de los datos cuantitativos. Metodología de la investigación* (cuarta). México:

McGraw-Hill. <https://doi.org/10.6018/turismo.36.231041>

López S., D. A., & Martínez A., C. A. (2013). Modelado de pérdidas en una transmisión

de video por medio de series de tiempo ARIMA y SARIMA. *Revista Tecnura*, 17(37), 53-63.

<https://doi.org/http://dx.doi.org/10.14483/udistrital.jour.tecnura.2013.3.a05>

Marie, G. C. (2007). Utilización de modelos ARIMA para la vigilancia de enfermedades

transmisibles. *Revista Cubana de Salud Publica*, 33(2), 1-11.

<https://doi.org/10.1590/S0864-34662007000200012>

Martinez, E. Z., Silva, E. A. S. da, & Fabbro, A. L. D. (2011). A SARIMA forecasting

model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil. *Revista Da Sociedade Brasileira de Medicina Tropical*, 44(4), 436-40.  
<https://doi.org/10.1590/S0037-86822011000400007>

Medinaceli, E. (2007). MODELO MATEMÁTICO PARA EL PRONÓSTICO DEL CLIMA CON FINES AGRÍCOLAS EN LA CUENCA DEL RÍO PARIA. Retrieved February 18, 2017, from <http://dpicuto.edu.bo/tesis/facultad-nacional-de-ingenieria/carrera-de-ingenieria-civil/225--modelo-matematico-para-el-pronostico-del-clima-con-fines-agricolas-en-la-cuenca-del-rio-paria.html>

Mellado, M. (2004). Pronóstico de las defunciones por medio de los modelos autorregresivos integrados de promedios móviles. *Papeles de Población, Redalyc*, 42, 179-206. <https://doi.org/1405-7425>

Molinero, L. M. (2004). Análisis de series temporales. *LADE Y LE*, 1(Análisis temporal), 111-128. Retrieved from <http://www.seh-lilha.org/tseries.htm>

Mora, C. (2003). Caracterización Hidrológica sub-cuenca Río David. Tesis para optar por licenciatura. Universidad de Panamá. Facultad de ciencias agropecuarias. Panamá.

Murillos, J., Trejos, Á., & Carvajal, P. (2003). Estudio del pronóstico de la demanda de energía eléctrica, utilizando modelos de series de tiempo. *Scientia et Technica*, 23(23), 37-40.

Navío, A. (2013). *La Investigación Educativa*. Barcelona.

Ocaña, R. (2011). ¿Qué es un modelo estadístico?- *Divestadística - Portal de Divulgación Estadística*. Retrieved July 30, 2017, from [http://www.divestadistica.es/es/que\\_es\\_un\\_modelo\\_estadistico.html](http://www.divestadistica.es/es/que_es_un_modelo_estadistico.html)

Olivares, F. (1994). Análisis del recurso hídrico en la cuenca del Río David. Tesis para

optar por licenciatura. Universidad de Panamá. Facultad de ciencias agropecuarias. Panamá.

Ortíz, I. (2006). Caracterización Hidrológica en el área de Boquete cuenca alta del río Caldera. Tesis para optar por licenciatura. Universidad de Panamá. Facultad de ciencias agropecuarias. Chiriquí-Panamá.

Ortíz, J.; Iglesias, D.; Hernández, R.; De Mera, A.; Gómez, M.; López, J. y Ballester, A. (2001). Desarrollo de programas para el tratamiento de series hidrológicas y evaluación de aportaciones hídricas subterráneas (Hidrobás v.3.0). Asociación Española de Hidrología Subterránea. Murcia

Pérez-Tejada, H. E. (2008). *Estadística para las ciencias sociales, del comportamiento y de la salud*. (F. de J. Castro P., Ed.) (3a. edición). México: Cengage Learning Editores, S.A. Retrieved from <http://latinoamerica.cengage.com>

Pulido, I., Roldán, J., López, R., & Gutiérrez, J. (2002). Técnicas de predicción a corto plazo de la demanda de agua. Aplicación al uso agrícola. *Ingeniería Del Agua*, 9(3), 319-331. Retrieved from <http://en.scientificcommons.org/47433887>

UAM. (2004). Modelos ARIMA. In *Curso de Predicción Económica y Empresarial* (pp. 1-6). Retrieved from [https://www.uam.es/docencia/predysim/prediccion\\_unidad4/4\\_3\\_doc3.pdf](https://www.uam.es/docencia/predysim/prediccion_unidad4/4_3_doc3.pdf)

UNESCO. (1982). Guía Metodológica para la elaboración del balance hídrico de América del Sur. Montevideo, Uruguay: ROSTLAC.

UNESCO. (2008). *Balance hídrico superficial de Panamá, Periodo 1971-2002*.

Verón, J., Herrera, C., & Rodríguez, N. (n.d.). *Análisis de serie de tiempo de caudales del río El Tala periodo 1937-1960*. Catamarca, Argentina.

Wikipedia. (2017). Errores de tipo I y de tipo II. Retrieved December 26, 2017, from [https://es.wikipedia.org/wiki/Errores\\_de\\_tipo\\_I\\_y\\_de\\_tipo\\_II](https://es.wikipedia.org/wiki/Errores_de_tipo_I_y_de_tipo_II)

ANEXO 1:

*Estación Finca Lérída (108-001)*



Estación de tipo Pluviométrica (PV), ubicada en el distrito de Boquete a una elevación de 1 700 msnm.

## ANEXO 2:

*Estación Los Naranjos (108-017)*

Estación de tipo B, ubicada en el distrito de Boquete a una elevación de 1 200 msnm.



Pluviómetro de la estación Los Naranjos

## ANEXO 3:

*Estación Veladero Gualaca (108-014)*

Estación de tipo Pluviométrica (PV), ubicada en el distrito de Gualaca a una elevación de 45 msnm.

## ANEXO 4:

*Estación David (108-023)*

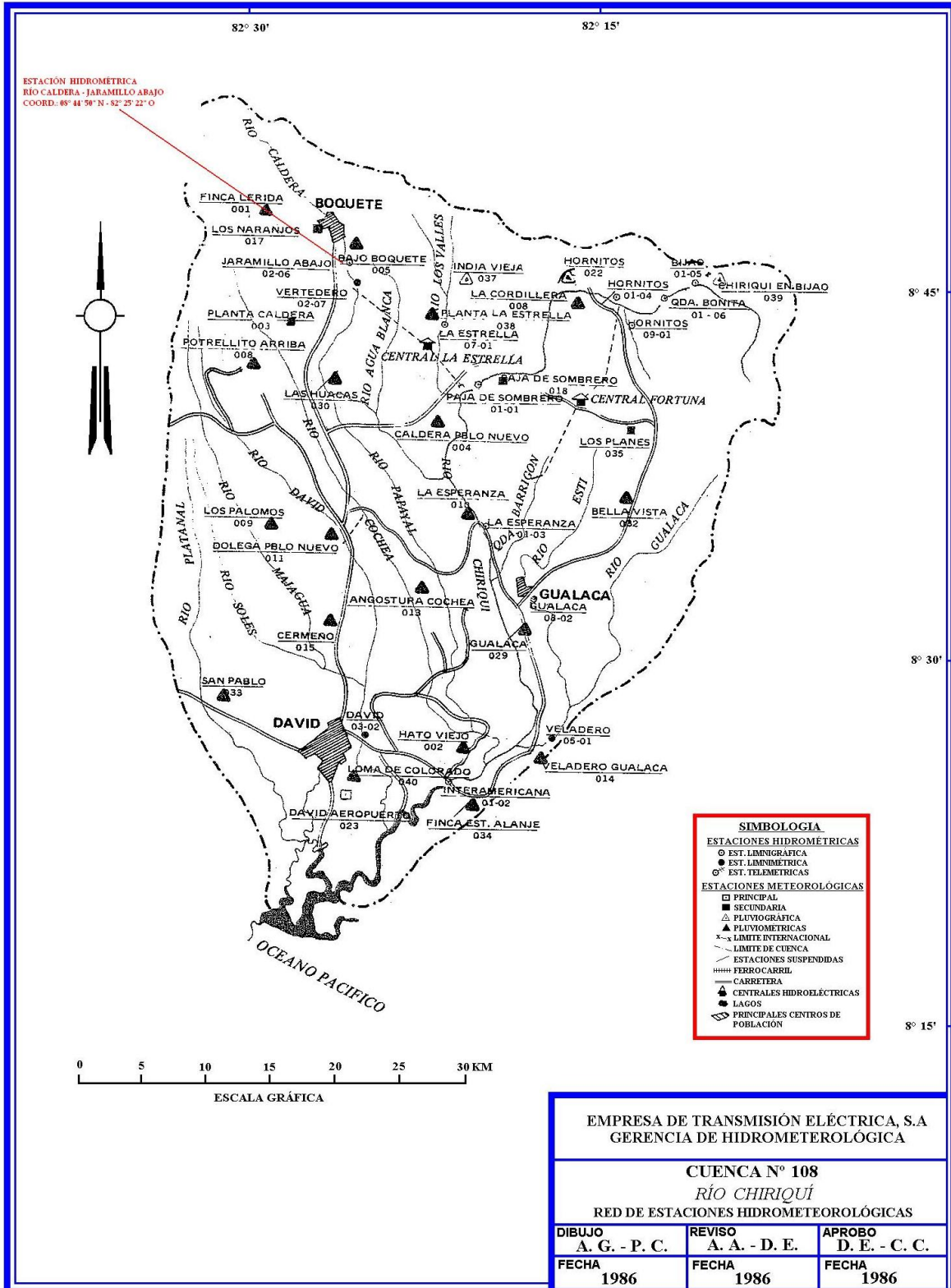
Estación de tipo A, ubicada en el distrito de David a una elevación de 27 msnm.



Pluviómetro de la estación David

ANEXO 5:

Cuenca 108 - río Chiriquí



Fuente: Gerencia de Hidrometeorología de la Empresa de Transmisión Eléctrica S.A. (ETESA), propietaria de todos los derechos.

ANEXO 6: Tabla de Datos: Estación Finca Lérica (108-001) zona alta

AÑO	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Anual
1972	71,0	52,5	16,5	132,5	224,5	352,5	117,5	303,5	315,5	364,0	193,0	129,0	2272,0
1973	124,0	40,5	5,5	122,5	412,0	453,5	369,5	414,0	380,5	520,0	429,5	228,5	3500,0
1974	111,5	28,5	20,5	24,5	427,6	383,1	170,0	301,3	375,1	616,8	56,5	81,5	2596,9
1975	62,5	10,0	12,5	27,5	379,3	231,0	271,0	295,0	708,3	482,6	522,1	278,1	3279,9
1976	110,0	35,0	16,0	107,5	316,8	429,9	258,5	220,0	161,5	403,6	231,1	85,0	2374,9
1977	15,5	14,0	7,0	23,5	376,8	236,0	107,5	262,0	278,0	309,3	194,5	18,0	1842,1
1978	36,0	30,0	44,0	87,5	360,0	179,5	320,1	169,0	418,6	410,9	165,0	244,8	2465,4
1979	26,5	14,0	38,5	345,1	254,3	273,3	187,0	405,3	602,2	586,3	385,7	148,0	3266,2
1980	114,0	105,0	37,5	39,0	370,1	398,7	267,8	275,4	302,5	327,5	436,3	194,6	2868,4
1981	67,5	141,0	53,5	200,0	350,0	488,6	122,0	350,6	274,8	552,3	239,5	126,5	2966,3
1982	45,0	53,4	18,0	118,0	441,3	220,0	148,5	135,0	368,4	464,6	146,3	59,0	2217,5
1983	24,5	29,0	29,0	28,0	196,5	246,8	317,3	240,5	345,3	336,0	208,5	51,0	2052,4
1984	114,0	52,0	40,7	74,5	294,6	290,8	336,8	395,4	437,4	422,6	244,5	285,0	2988,3
1985	46,5	45,5	64,0	81,0	154,0	274,1	243,3	238,0	219,5	373,6	124,8	129,5	1993,8
1986	144,3	13,5	200,1	29,0	338,1	350,1	77,5	279,5	473,6	587,8	67,5	51,0	2612,0
1987	75,0	26,0	38,5	68,0	248,0	220,5	284,3	300,1	423,9	469,3	151,5	148,0	2453,1
1988	122,9	69,0	82,5	43,0	183,5	374,8	344,1	792,1	484,2	605,0	154,0	157,5	3412,6
1989	61,0	133,3	58,0	50,5	171,5	338,4	333,6	189,8	404,3	215,0	186,0	314,6	2456,0
1990	150,5	28,5	90,0	15,0	313,3	121,5	172,5	264,0	225,0	447,3	454,2	252,0	2533,8
1991	17,5	58,5	38,0	64,5	400,3	320,1	235,3	293,8	307,0	496,2	92,0	147,0	2470,2
1992	25,0	27,5	48,0	187,3	136,5	215,0	195,5	153,5	313,0	223,3	105,5	232,5	1862,6
1993	170,0	36,0	197,0	89,0	444,5	256,0	88,5	306,0	471,4	342,3	236,8	201,1	2838,6
1994	36,5	27,5	26,5	32,0	236,0	214,8	175,8	234,5	397,1	472,4	296,5	207,0	2356,6
1995	25,5	42,5	57,0	169,5	270,8	504,9	296,5	490,3	387,0	610,1	136,0	94,5	3084,6
1996	312,4	180,8	56,5	38,0	345,3	307,8	309,3	271,8	328,8	507,6	334,8	236,8	3229,9
1997	120,8	125,5	38,5	179,8	195,0	318,0	208,0	161,0	312,5	427,4	360,3	202,9	2649,7
1998	7,0	42,5	73,5	53,0	227,0	354,5	388,9	249,0	427,5	621,4	256,3	329,8	3030,4
1999	116,5	60,5	74,0	301,4	266,0	387,1	108,0	481,1	689,9	610,6	313,0	465,1	3873,2
2000	356,6	107,5	27,0	106,3	279,5	348,3	234,5	224,5	460,6	246,5	205,8	110,5	2707,6
2001	152,0	85,5	44,5	84,5	235,8	192,5	52,5	238,8	451,5	259,3	259,0	191,1	2247,0
2002	127,0	70,5	48,5	79,0	444,4	239,5	214,5	381,1	469,3	346,5	322,8	52,5	2795,6
2003	47,0	20,0	130,5	109,3	323,6	546,4	238,5	170,5	486,1	493,8	415,4	266,0	3247,1
2004	104,0	74,0	166,0	89,0	452,7	222,0	270,5	221,3	508,1	360,1	361,3	165,0	2994,0
2005	469,4	44,5	59,5	103,5	376,9	369,0	220,5	365,8	318,5	672,2	247,0	96,3	3343,1
2006	302,1	71,2	74,9	74,0	393,0	368,5	236,9	244,3	137,6	413,5	288,9	152,0	2756,9
2007	120,2	19,8	50,3	118,3	566,6	144,3	299,9	512,6	421,8	601,4	381,3	337,7	3574,2
2008	86,8	67,0	41,9	36,9	457,0	181,8	483,1	567,2	437,6	439,6	723,4	197,7	3720,0
2009	91,6	414,6	147,5	21,7	202,6	305,8	225,0	207,3	101,5	374,9	259,8	27,3	2379,6
2010	190,5	97,2	132,5	192,2	234,7	410,4	336,7	689,6	853,8	313,6	335,6	437,5	4224,3
2011	140,0	60,5	83,6	65,3	338,5	390,5	397,6	363,4	391,6	661,9	157,0	157,2	3207,1
2012	132	54,5	227,4	224,9	246,9	198,3	248,7	263,2	250,5	382	256,2	180,4	2665,0
2013	18,1	20,4	60,3	41,2	210,9	120,6	175,3	242,8	416	413,6	209,5	57,6	1986,3
2014	34,8	33,8	18,7	32,7	166,1	310,3	145,6	300,5	449,8	463,8	133,9	135,8	2225,8
2015	120,4	76,4	40,6	62,2	237,1	315,4	236,9	275,4	397,1	390,6	393,5	157,5	2703,1
2016	S/D	S/D	S/D	S/D	S/D	S/D	S/D	S/D	132,4	S/D	S/D	S/D	132,4
Máx.	469,4	414,6	227,4	345,1	566,6	546,4	483,1	792,1	853,8	672,2	723,4	465,1	4224,3
Mín.	7,0	10,0	5,5	15,0	136,5	120,6	52,5	135,0	101,5	215,0	56,5	18,0	1842,1
Promedio	107,0	48,8	48,3	76,8	304,0	309,1	236,9	275,4	397,1	433,5	245,8	157,5	2780,1

ANEXO 7: Tabla de Datos: Estación David (108-023) zona baja

AÑO	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Anual
1972	75,1	8,9	1,3	175,0	132,2	279,9	54,2	270,9	701,1	219,0	344,3	54,6	2316,5
1973	26,6	25,2	80,6	101,7	308,7	290,3	373,8	494,6	477,1	390,0	375,6	50,1	2994,3
1974	46,6	9,4	44,5	77,6	416,5	301,1	212,5	339,7	297,2	613,8	291,2	49,0	2699,1
1975	17,9	6,7	1,0	3,9	332,7	369,9	342,0	352,5	412,4	253,6	446,4	113,9	2652,9
1976	56,7	4,4	3,7	76,1	134,7	321,8	96,1	280,8	428,7	396,2	202,4	3,5	2005,1
1977	0,0	0,1	7,4	66,0	286,7	387,2	177,3	353,3	437,0	333,0	327,8	68,2	2444,0
1978	8,1	17,6	116,1	194,5	229,1	389,7	273,4	174,3	278,8	480,0	115,5	71,8	2348,9
1979	0,4	57,8	119,8	163,5	428,1	194,7	171,2	404,4	196,1	422,5	407,4	91,7	2657,6
1980	11,4	90,2	0,0	40,1	276,0	249,8	185,4	371,2	465,3	181,8	287,0	109,2	2267,4
1981	0,8	0,4	83,4	239,9	219,4	229,5	264,8	375,1	243,7	439,3	356,9	149,5	2602,7
1982	62,4	61,5	27,8	142,0	305,2	242,1	271,2	200,4	369,2	479,5	74,2	8,3	2243,8
1983	8,6	7,6	50,0	46,0	337,3	298,6	308,5	283,6	566,3	383,1	235,1	70,3	2595,0
1984	32,6	56,4	38,9	30,0	268,6	393,1	297,7	425,0	345,4	629,6	335,5	0,3	2853,1
1985	3,2	0,0	0,0	124,3	193,4	432,9	419,2	265,3	524,0	259,6	66,1	90,4	2378,4
1986	0,3	20,7	12,0	37,8	293,9	319,5	183,8	536,8	359,4	679,9	158,1	78,2	2680,4
1987	3,1	0,0	33,1	109,3	327,8	429,3	386,1	322,1	365,6	358,7	30,8	7,0	2372,9
1988	1,6	4,8	31,2	124,7	143,8	204,5	394,3	432,0	297,4	618,8	184,9	82,5	2520,5
1989	25,8	0,0	6,2	0,3	245,1	417,0	303,8	260,8	337,0	361,5	235,4	59,8	2252,7
1990	67,1	26,8	23,9	39,0	432,1	205,6	303,0	234,1	329,3	391,9	286,4	88,3	2427,5
1991	67,7	6,4	61,4	120,4	351,7	238,8	254,7	166,7	302,6	367,3	183,9	51,8	2173,4
1992	0,0	14,0	0,0	63,8	292,1	135,9	201,1	482,2	315,1	289,3	249,4	31,7	2074,6
1993	95,5	10,7	27,5	44,6	295,0	284,1	371,1	500,2	457,6	499,4	224,5	33,4	2843,6
1994	9,0	1,1	14,5	55,5	412,4	460,8	192,6	346,1	465,6	360,6	473,1	18,8	2810,1
1995	0,0	0,0	43,8	226,1	480,6	360,5	401,7	545,2	288,9	215,2	246,3	73,4	2881,7
1996	73,7	10,8	40,1	113,9	435,8	614,2	600,3	165,3	262,2	323,2	143,2	42,4	2825,1
1997	92,3	6,4	15,2	74,1	182,0	279,4	149,4	195,2	429,0	201,7	400,9	201,6	2227,2
1998	0,0	55,7	3,6	32,2	169,6	320,7	503,9	396,7	451,2	515,5	256,7	270,6	2976,4
1999	20,7	8,1	2,6	290,6	327,2	386,5	221,8	483,6	554,4	459,5	225,8	92,4	3073,2
2000	35,1	19,9	0,7	49,7	440,0	262,8	291,0	203,8	692,8	399,8	120,4	108,0	2624,0
2001	10,8	13,3	24,5	18,6	595,8	244,9	282,2	291,8	342,2	560,3	235,2	63,6	2683,2
2002	22,0	0,0	2,3	8,0	271,7	174,6	344,0	322,6	353,4	299,3	211,9	11,7	2021,5
2003	10,3	0,4	67,4	123,6	439,3	487,1	397,4	304,6	320,2	488,7	466,0	147,9	3252,9
2004	0,0	20,5	16,1	93,3	432,1	187,0	281,0	413,8	443,7	448,4	186,6	62,3	2584,8
2005	21,2	0,0	109,0	99,6	316,4	308,8	174,1	501,5	276,7	222,0	630,1	128,1	2787,5
2006	175,6	5,0	40,8	43,1	308,0	281,7	368,4	308,6	236,7	465,9	196,6	146,1	2576,5
2007	3,0	0,0	28,7	149,3	479,1	233,7	316,3	497,1	407,2	413,0	281,3	69,8	2878,5
2008	0,0	59,4	20,4	64,8	461,6	180,4	485,2	488,1	251,1	672,2	556,1	87,4	3326,7
2009	25,4	14,4	36,7	5,0	348,8	325,2	265,9	252,7	208,4	377,4	413,6	48,6	2322,1
2010	30,1	103,2	22,2	311,2	277,9	508,1	520,5	430,9	463,7	424,7	432,1	99,1	3623,7
2011	7,2	49,5	67,7	116,7	308,0	414,4	419,1	399,1	243,3	359,8	311,1	153,7	2849,6
2012	3,5	0,1	10,9	121,0	401,6	180,3	179,1	402,2	198,8	364,4	170,2	47,5	2079,6
2013	0,6	7,1	3,3	94,8	250,2	253,7	331,2	208,8	273,4	244,7	132,3	68,2	1868,3
2014	13,5	2,4	43,6	150,2	538,2	405,6	109,7	145,8	423,0	308,9	355,2	195,6	2691,7
2015	0,0	30,7	0,1	47,2	322,1	327,3	367,2	285,3	314,7	466,7	225,9	8,5	2395,7
2016	2,5	1,5	20,9	61,5	106,0	178,5	S/D	S/D	443,1	S/D	S/D	S/D	814,0
Máx.	175,6	103,2	119,8	311,2	595,8	614,2	600,3	545,2	701,1	679,9	630,1	270,6	3623,7
Mín.	0,0	0,0	0,0	0,3	132,2	135,9	54,2	145,8	196,1	181,8	30,8	0,3	1868,3
Promedio	11,1	8,5	24,2	85,5	312,6	299,9	294,4	342,9	349,4	391,0	247,9	70,1	2585,6

## ANEXO 8:

## Tabla de Datos: Estación Los Naranjos (108-017) zona alta

AÑO	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Anual
1972	179,3	22,5	26,5	523,2	642,3	257,3	269,9	300,0	470,4	415,8	290,5	119,2	3516,9
1973	0,0	50,8	50,5	104,8	360,8	741,9	632,1	575,5	1032,4	760,0	540,9	101,8	4951,5
1974	31,0	4,0	111,3	14,0	379,9	458,7	219,2	710,3	438,2	632,2	327,6	48,0	3374,4
1975	2,0	21,0	7,0	70,5	455,2	529,9	357,7	414,3	375,5	440,3	445,7	157,0	3276,1
1976	4,8	0,0	4,5	221,0	255,3	285,6	160,3	319,7	397,4	342,4	344,5	16,0	2351,5
1977	0,0	0,0	5,0	55,0	367,4	369,6	107,1	407,0	503,4	418,2	380,0	153,3	2766,0
1978	24,0	8,0	134,5	185,5	368,0	343,1	461,4	343,3	452,1	447,7	361,8	80,0	3209,4
1979	1,0	22,5	92,0	258,4	522,6	241,3	244,8	481,9	346,1	468,1	555,5	155,5	3389,7
1980	16,5	74,0	0,5	50,0	293,6	389,1	304,3	339,3	446,3	653,8	347,8	200,1	3115,3
1981	3,0	19,5	54,0	340,9	407,3	576,7	284,1	642,9	472,1	708,7	464,5	97,0	4070,7
1982	231,1	26,0	106,0	229,8	282,0	299,3	254,0	351,7	441,7	789,1	110,0	5,5	3126,2
1983	0,0	19,5	77,5	130,8	439,7	460,1	230,8	355,3	560,5	456,3	410,8	151,0	3292,3
1984	52,5	99,5	134,0	98,5	464,8	558,6	567,2	398,4	573,3	726,8	672,4	10,0	4356,0
1985	5,0	0,0	10,0	243,3	229,8	446,4	305,6	425,6	579,2	497,4	274,8	180,1	3197,2
1986	2,5	99,3	63,5	36,5	332,9	411,2	238,0	471,9	368,9	793,7	96,5	43,0	2957,9
1987	26,0	0,0	98,0	87,5	291,5	680,5	499,4	343,0	419,9	384,6	209,8	169,5	3209,7
1988	0,0	7,5	40,5	233,1	334,9	511,7	524,4	795,6	442,6	822,2	407,8	44,0	4164,3
1989	8,0	0,0	62,0	35,5	218,0	474,9	393,1	352,1	593,2	402,3	312,5	179,5	3031,1
1990	55,0	10,0	49,0	182,6	619,1	435,5	400,1	381,4	348,8	614,5	362,2	87,0	3545,2
1991	120,3	0,0	105,0	175,8	229,0	455,4	316,9	308,3	501,4	455,1	276,8	107,0	3051,0
1992	0,0	0,0	0,0	49,0	385,1	300,1	260,8	379,3	236,3	491,9	340,1	69,0	2511,6
1993	95,3	2,0	108,0	151,5	446,4	514,7	286,3	613,3	723,2	512,3	355,9	177,1	3986,0
1994	11,5	28,0	39,5	56,0	441,2	295,0	366,8	462,3	452,0	552,7	301,8	65,5	3072,3
1995	0,0	0,0	59,5	504,7	517,0	390,3	552,2	638,9	444,6	420,3	205,5	98,0	3831,0
1996	86,5	8,5	100,5	103,0	352,1	473,7	383,3	335,8	391,8	456,1	289,5	65,0	3045,8
1997	122,3	14,5	42,0	108,5	256,5	219,5	204,3	196,0	379,5	194,5	541,8	47,5	2326,9
1998	0,0	51,5	0,0	48,5	337,5	341,5	357,0	674,0	302,0	436,0	211,0	375,5	3134,5
1999	47,0	142,3	10,5	126,0	269,5	644,8	380,5	452,0	656,6	450,5	306,8	153,0	3639,5
2000	27,5	13,5	48,0	99,5	440,0	532,0	358,5	324,5	632,0	390,0	262,0	143,0	3270,5
2001	4,5	0,0	26,5	47,0	343,1	292,0	187,5	462,6	448,7	574,0	393,0	71,6	2850,5
2002	21,6	4,0	31,0	21,5	135,1	211,6	484,5	509,5	595,0	389,2	320,0	75,0	2798,0
2003	0,0	49,0	113,5	250,7	573,0	649,0	517,1	531,1	380,0	696,4	497,6	164,0	4421,4
2004	0,0	16,0	16,8	122,7	481,8	292,9	355,8	347,0	473,2	539,0	171,7	61,0	2877,9
2005	4,0	27,5	150,5	120,5	364,7	363,3	484,3	477,0	455,2	281,5	544,4	51,6	3324,5
2006	130,7	9,0	55,4	67,7	605,4	409,6	573,6	211,8	403,3	575,0	295,9	232,8	3570,2
2007	24,2	0,0	217,4	505,9	515,0	256,2	531,6	608,4	589,7	424,1	237,6	132,8	4042,9
2008	2,3	255,3	46,2	67,8	474,4	214,2	613,3	622,0	558,2	818,9	724,1	77,1	4473,8
2009	53,7	21,6	70,2	32,4	655,2	399,6	355,4	524,2	269,8	442,8	393,4	126,6	3344,9
2010	13,5	131,0	54,0	372,1	454,1	493,5	452,6	514,0	619,9	563,2	610,6	177,5	4456,0
2011	56,5	49,5	127,4	218,2	388,4	577,3	567,0	446,7	573,7	451,7	574,4	168,3	4199,1
2012	0,0	0,0	6,4	202,4	516,3	254,6	144,3	534,5	187,6	475,7	178,1	92,1	2592,0
2013	9,5	30,0	36,0	122,5	520,0	419,0	534,4	578,7	369,7	604,5	225,6	67,7	3517,6
2014	9,8	19,3	3,9	197,2	508,8	450,8	166,2	278,4	473,5	372,0	332,7	231,6	3044,2
2015	12,2	26,3	1,0	140,3	365,3	443,3	234,7	39,7	S/D	S/D	S/D	S/D	1262,8
Máx.	231,1	255,3	217,4	523,2	655,2	741,9	632,1	795,6	1032,4	822,2	724,1	375,5	4951,5
Mín.	0,0	0,0	0,0	14,0	135,1	211,6	107,1	196,0	187,6	194,5	96,5	5,5	2326,9
Mediana	34,5	31,5	60,3	159,8	406,4	416,8	369,5	452,1	473,9	519,6	360,6	116,9	3401,9
Promedio	9,8	16,0	50,5	122,5	388,4	411,2	357,7	446,7	452,0	468,1	340,1	101,8	3276,1

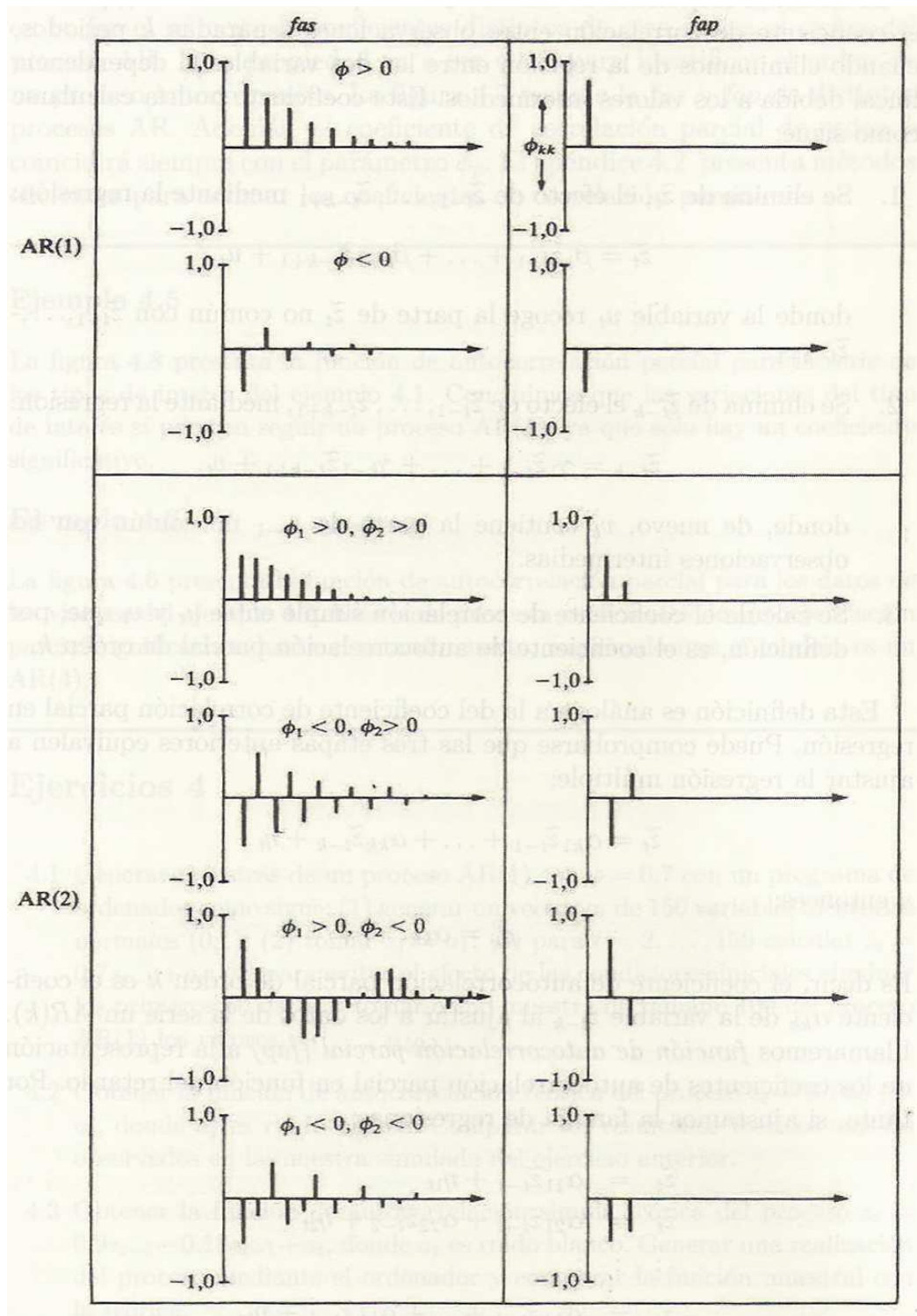
## ANEXO 9:

## Tabla de Datos: Estación Veladero Gualaca (108-014) zona baja

AÑO	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Anual
1972	49,9	47,1	12,7	83,3	176,2	290,2	101,5	259,3	352,6	297,3	142,9	59,8	1872,8
1973	81,1	20,5	22,7	112,8	392,5	489,6	334,5	366,9	326,6	619,7	459,0	132,6	3358,5
1974	44,2	6,1	10,6	22,5	417,9	427,5	177,7	238,9	532,7	431,2	44,3	57,8	2411,4
1975	28,6	5,2	2,5	13,6	292,8	237,5	378,7	233,3	680,1	402,7	640,5	196,4	3111,9
1976	54,1	11,4	5,2	59,4	241,6	398,0	195,8	197,7	190,1	341,8	156,7	70,3	1922,1
1977	3,2	2,2	0,0	3,2	324,5	247,0	73,5	343,7	241,1	191,2	118,5	4,5	1552,6
1978	5,2	5,3	69,3	54,5	389,7	232,0	329,3	180,1	479,4	398,4	160,2	205,7	2509,1
1979	11,8	3,8	82,9	328,7	308,0	383,9	198,9	337,6	436,5	492,1	256,7	88,6	2929,5
1980	47,1	42,1	36,5	42,6	404,3	347,4	258,8	314,3	352,1	284,3	370,4	110,2	2610,1
1981	25,7	57,6	41,9	127,6	359,2	479,3	136,6	222,7	350,8	481,1	186,1	81,9	2550,5
1982	17,6	9,2	11,2	72,1	380,3	285,0	103,8	109,8	442,8	471,6	80,0	9,3	1992,7
1983	2,0	18,6	26,0	40,0	101,5	296,7	312,7	241,4	362,3	344,1	172,4	28,9	1946,6
1984	36,2	21,4	21,1	67,1	186,6	366,6	396,6	470,5	533,9	433,8	163,0	134,8	2831,6
1985	31,8	6,3	19,9	129,6	127,5	309,7	211,4	342,4	237,6	279,5	93,2	66,6	1855,5
1986	55,9	37,9	89,5	7,3	223,0	348,2	64,0	215,3	560,9	709,7	72,4	8,3	2392,4
1987	18,6	5,7	63,2	24,2	201,2	199,8	207,1	352,5	384,7	401,1	59,1	79,0	1996,2
1988	140,9	13,6	79,0	29,3	176,3	383,3	296,6	823,7	587,0	54,4	122,1	54,7	2760,9
1989	15,2	55,1	8,1	24,1	159,4	331,1	239,2	185,5	508,8	214,3	149,5	181,9	2072,2
1990	44,4	14,7	23,8	0,0	331,6	117,6	162,8	239,8	234,7	410,7	559,0	155,5	2294,6
1991	2,5	17,7	60,2	83,3	348,4	339,9	198,0	218,8	394,3	409,0	81,9	64,8	2218,8
1992	2,5	3,6	24,0	190,9	140,1	219,8	171,2	177,7	300,3	258,3	139,1	146,4	1773,9
1993	84,5	7,8	46,5	82,7	399,3	269,8	78,4	309,1	470,5	370,1	195,4	99,9	2414,0
1994	10,3	5,7	13,8	92,4	334,2	223,9	120,4	250,5	406,1	489,2	299,4	112,4	2358,3
1995	17,8	16,7	29,4	119,2	174,9	640,2	259,0	439,0	373,8	550,6	98,6	89,7	2808,9
1996	131,8	75,2	26,7	47,5	328,3	317,6	329,3	310,0	275,5	411,7	249,1	133,3	2636,0
1997	82,5	46,6	58,4	185,6	147,7	223,7	184,8	111,9	263,4	494,3	305,3	46,3	2150,5
1998	10,6	41,9	14,1	16,4	171,4	356,9	436,3	338,8	477,1	489,2	180,4	257,0	2790,1
1999	63,5	48,3	24,6	222,4	229,8	304,5	83,4	501,3	673,9	745,8	267,4	220,6	3385,5
2000	192,4	40,2	5,6	73,0	374,1	333,2	274,8	224,1	572,6	201,4	176,0	71,9	2539,3
2001	75,6	19,9	57,3	33,4	294,5	210,3	158,6	205,3	479,8	294,8	234,8	108,1	2172,4
2002	36,2	13,5	6,1	27,4	316,4	252,6	170,1	433,5	545,3	413,4	201,8	23,0	2439,3
2003	7,3	3,2	68,3	193,9	381,5	478,9	232,0	135,4	381,8	460,2	193,8	115,2	2651,5
2004	62,4	22,9	42,1	30,2	356,2	132,7	270,9	167,7	353,0	425,7	247,4	90,0	2201,2
2005	250,7	11,1	118,3	36,1	274,7	369,2	240,9	373,3	268,5	598,4	205,8	48,7	2795,7
2006	187,7	21,3	20,7	65,2	134,2	382,1	286,3	207,0	156,6	414,2	245,9	97,1	2218,3
2007	31,4	2,1	21,8	117,3	503,3	187,6	210,6	340,2	382,5	490,0	215,0	213,0	2714,8
2008	34,5	29,0	21,2	19,6	478,1	230,1	485,7	556,2	406,7	450,3	503,3	128,4	3343,1
2009	25,7	201,3	108,7	11,5	216,6	408,4	161,4	189,0	125,9	419,2	305,3	58,4	2231,4
2010	78,6	68,1	47,9	226,2	260,8	469,8	415	760,8	630,9	359,6	248,8	230,5	3797,0
2011	75	27,8	93,4	65,7	285,6	430	416,4	342,1	318,2	674,5	108,0	75,2	2911,9
2012	60,8	21,5	146	223,4	281,4	213,2	161	286,1	148,9	297,7	138,5	56,1	2034,6
2013	3,1	2,6	16,8	6,8	193,8	139,3	160,9	177,2	454,9	410,9	178,5	43,4	1788,2
2014	7,5	2,6	22,7	18,3	192,3	188,5	105,9	217,6	302,2	S/D	S/D	S/D	1057,6
Máx.	250,7	201,3	146,0	328,7	503,3	640,2	485,7	823,7	680,1	745,8	640,5	257,0	3797,0
Mín.	2,0	2,1	0,0	0,0	101,5	117,6	64,0	109,8	125,9	54,4	44,3	4,5	1552,6
Mediana	53,4	26,9	40,4	81,2	281,4	316,8	230,6	303,1	396,6	414,0	214,9	101,3	2460,6
Promedio	36,2	18,2	25,3	62,3	289,2	313,7	208,9	254,9	383,6	412,6	183,3	89,2	2412,7

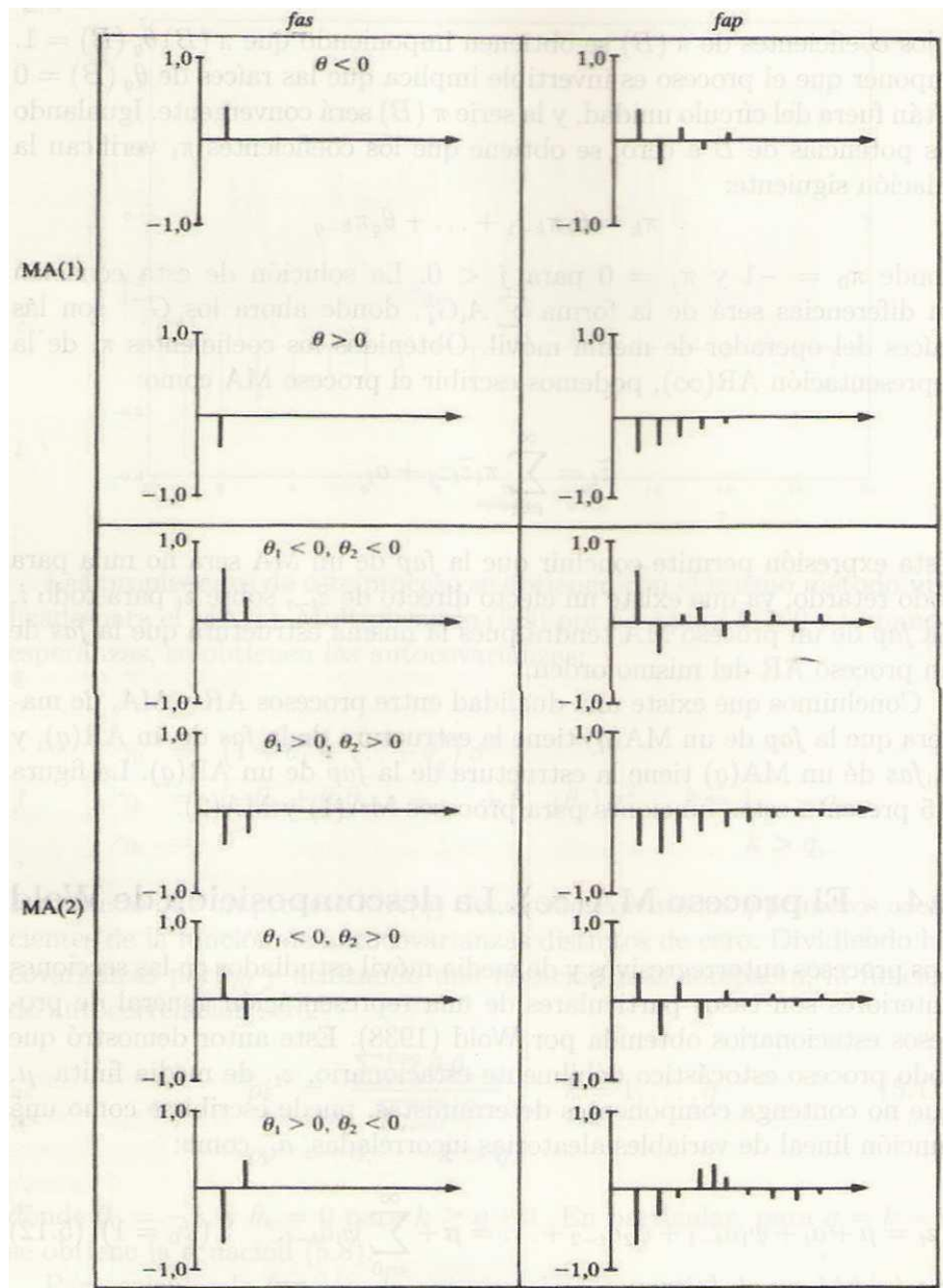
ANEXO 10:

Modelos AR y MA teóricos



Fuente: Análisis de Series Temporales (Breña Moral, n.d., pp. 87-88)

## Modelos AR y MA teóricos continuación



TIPO	FAS	FAP
AR(p)	Muchos coeficientes no nulos	Primeros p no nulos, resto 0
MA(q)	Primeros p no nulos, resto 0	Muchos coeficientes no nulos
ARMA(p,q)	Muchos coeficientes no nulos	Muchos coeficientes no nulos

Fuente: Breña, J. Análisis de Series Temporales. pp. 87-88

## ANEXO 11:

Valores críticos de Distribución Chi-Cuadrado  $\chi^2$ 

Grados de libertad	Probabilidad de un valor superior - Alfa ( $\alpha$ )				
	0,1	0,05	0,025	0,01	0,005
1	2,71	3,84	5,02	6,63	7,88
2	4,61	5,99	7,38	9,21	10,60
3	6,25	7,81	9,35	11,34	12,84
4	7,78	9,49	11,14	13,28	14,86
5	9,24	11,07	12,83	15,09	16,75
6	10,64	12,59	14,45	16,81	18,55
7	12,02	14,07	16,01	18,48	20,28
8	13,36	15,51	17,53	20,09	21,95
9	14,68	16,92	19,02	21,67	23,59
10	15,99	18,31	20,48	23,21	25,19
11	17,28	19,68	21,92	24,73	26,76
12	18,55	21,03	23,34	26,22	28,30
13	19,81	22,36	24,74	27,69	29,82
14	21,06	23,68	26,12	29,14	31,32
15	22,31	25,00	27,49	30,58	32,80
16	23,54	26,30	28,85	32,00	34,27
17	24,77	27,59	30,19	33,41	35,72
18	25,99	28,87	31,53	34,81	37,16
19	27,20	30,14	32,85	36,19	38,58
20	28,41	31,41	34,17	37,57	40,00
21	29,62	32,67	35,48	38,93	41,40
22	30,81	33,92	36,78	40,29	42,80
23	32,01	35,17	38,08	41,64	44,18
24	33,20	36,42	39,36	42,98	45,56
25	34,38	37,65	40,65	44,31	46,93
26	35,56	38,89	41,92	45,64	48,29
27	36,74	40,11	43,19	46,96	49,65
28	37,92	41,34	44,46	48,28	50,99
29	39,09	42,56	45,72	49,59	52,34
30	40,26	43,77	46,98	50,89	53,67
40	51,81	55,76	59,34	63,69	66,77
50	63,17	67,50	71,42	76,15	79,49
60	74,40	79,08	83,30	88,38	91,95
70	85,53	90,53	95,02	100,43	104,21
80	96,58	101,88	106,63	112,33	116,32
90	107,57	113,15	118,14	124,12	128,30
100	118,50	124,34	129,56	135,81	140,17