

---

# **ANÁLISIS DE TENDENCIA Y HOMOGENEIDAD DE SERIES CLIMATOLÓGICAS**



*Río DAGUA*

## **Lina M. Castro, M.Sc.**

Escuela de Ingeniería de Recursos Naturales y del Ambiente.

Universidad del Valle, Cali, Colombia.

*linacahe@yahoo.com*

## **Yesid Carvajal Escobar, Ph.D.**

Profesor Titular

Escuela de Ingeniería de Recursos Naturales y del Ambiente.

Universidad del Valle, Cali, Colombia.

*yecarvaj@gmail.com*

## **RESUMEN**

Para la planeación y diseño de muchos proyectos relacionados con el agua es necesario el uso de información hidroclimatológica. Aunque con los años la recolección de ésta ha ido mejorando, aún muchos de los registros en los que se basa la meteorología aplicada presentan serias deficiencias, tanto en calidad como en cantidad, observándose series con cambios, falta de información, tendencias y datos atípicos; esto tergiversa los resultados de cualquier simulación o modelación. Debido a la importancia económica y social que presenta la correcta predicción y el uso de modelos a partir de esta información, se hace necesario el análisis exploratorio de los datos con el fin de determinar cambios y/o tendencias en la serie hidroclimatológica. En

---

*\*Recibido: Mayo 10 2010 \*Junio 8 2010*

---

este artículo se presentan las herramientas gráficas y cuantitativas disponibles para el análisis exploratorio de datos, con el objetivo fundamental de dar a conocer una serie de métodos en forma conjunta y organizada, que pueden ser programables o que se encuentran en cualquier paquete estadístico. Al final se muestra la aplicación de estas pruebas en series de precipitación con algunas conclusiones y recomendaciones.

## PALABRAS CLAVE

Series cronológicas, tendencia, pruebas paramétricas, pruebas no paramétricas, varianza.

## ABSTRACT

*Planning and design, for many water related projects, require the acquisition and use of hydroclimatology data. Although during the past years, collection has been improved, many of the records on which the applied meteorology are based on, still have quality and quantity issues, from changes in time series, missing data, trends to atypical data, changing dramatically the results of any modelling of simulation. Due to the economic and social importance of having accurate predictions and the use of models from this information, it becomes necessary the exploratory data analysis, in order to determine changes/trends in the time series of hydroclimatology data.*

*Along this document graphic and quantitative tools available to perform the exploratory data analysis are presented. with the purpose to spread the information about many techniques already available in several statistical software packages. Finally it is presented a case study for the application of these techniques on time series of precipitation, including conclusions and recommendations.*

## KEY WORDS

*Time series, trend, parametric test, non parametric test, variance.*

## 1. INTRODUCCIÓN

Para la planeación y diseño de muchos proyectos relacionados con el agua es necesario el uso de información hidroclimatológica proveniente de eventos hidrometeorológicos gobernados por las leyes del azar (Mesa *et al.*, 1997). Luego, las predicciones dependen de la calidad y la cantidad de la información y de la correcta aplicación de los métodos estadísticos disponibles, pues se pretende realizar inferencias de la población a partir de una muestra. Aunque con los años la recolección de información hidroclimatológica (IH) ha ido mejorando, aún muchos de los registros en los que se basa la meteorología aplicada son deficientes tanto en calidad como en cantidad, observándose series con falta de información, con cambios, tendencias y datos atípicos, circunstancias que tergiversan los resultados obtenidos de la modelación o simulación. Para el uso correcto de la IH en la ingeniería de los recursos hídricos, ésta debe cumplir con el supuesto de estacionalidad, consistencia y homogeneidad, sin dejar aparte, en algunos casos, la independencia entre las observaciones. La causa de la no estacionalidad se debe principalmente a procesos naturales y antropogénicos. Dentro de los procesos naturales se pueden mencionar (para series de caudal) incendios forestales, derrumbes y explosiones volcánicas, entre otros; y dentro de los procesos antropogénicos se tienen la tala indiscriminada de bosques, el sobrepastoreo, la destrucción de la cobertura vegetal, los cambios en el uso del suelo, la prácticas agrícolas inadecuadas, la obras de ingeniería que alteran la morfología y dinámica fluvial, etc. También cabe mencionar el cambio climático como fenómeno reciente que está alterando el clima y con éste las temperaturas y la distribución temporal y espacial de la precipitación, ocasionando la recurrencia de sequías e inundaciones, etc., fenómenos que alteran significativamente los registros hidroclimatológicos y que pueden causar cambios y tendencias en los mismos.

Por lo anterior resulta necesario el análisis exploratorio de los datos con el fin de determinar cambios o tendencias en la Serie hidroclimatológica (SH), y aunque es un análisis que toma tiempo, es una parte esencial de cualquier análisis estadístico. El uso de la información sin previa evaluación de su estructura, consistencia, homogeneidad, etc., constituye un enfoque de caja negra que incrementa el grado de incertidumbre sobre la validez de los resultados obtenidos (Lobo, 2004).

Antes de realizar cualquier tipo de análisis hidrológico o climatológico, la información debe ser ampliamente explorada usando métodos gráficos y estadísticos cuantitativos; si ello no se realiza se pueden hacer suposiciones que no son ciertas (independencia, normalidad, homocedasticidad, entre otras) y modelar las SH de forma incorrecta. Los métodos gráficos, como herramienta exploratoria de los datos, se usan con dos propósitos: revelar las características de una posible distribución o las relaciones que existen entre las variables, que de otra manera no podrían ser descubiertas (Maidment, 1993). Los métodos cuantitativos, por su parte, sirven para determinar la estacionalidad de la serie y la estabilidad en la varianza y la media; éstas pruebas, a su vez, se subdividen en paramétricas y no paramétricas; el uso de una u otra dependerá de si los datos de la serie hidroclimatológica (SH) se distribuyen siguiendo una distribución normal o no.

Las pruebas paramétricas son usadas indiscriminadamente sin realizar ni siquiera una prueba de normalidad antes de su aplicación; en muchas ocasiones esa suposición no es válida y en otras la sospecha de que no sea adecuada no resulta fácil de comprobar, por tratarse de muestras pequeñas. En estos casos se dispone de dos posibles mecanismos: los datos se pueden transformar de tal manera que se asemejen a una distribución normal, o bien se puede acudir a pruebas estadísticas que no se basan en ninguna suposición en cuanto a la distribución de probabilidad que siguen los datos, y por ello se denominan pruebas no paramétricas (o de distribución libre).

Los procedimientos descritos en este documento (que por cierto aparecen dispersos en algunas otras publicaciones) pretenden ser una guía para dar a conocer algunos de los métodos más utilizados en el análisis exploratorio de datos, en forma conjunta y organizada, y que son fácilmente programables e incluso se encuentran en cualquier paquete estadístico.

## 2. METODOLOGÍA

El análisis exploratorio de una serie hidroclimatológica consiste en detectar por medios gráfico y cuantitativos la existencia o no de alguna tendencia y/o cambio, y la homogeneidad de la serie. El esquema metodológico de un análisis exploratorio (figura 1) empieza por un análisis gráfico (gráfica de serie de tiempo, gráfica de doble

masa, diagrama de cajas, histogramas, gráfica de normalidad), continúa con la prueba de normalidad (Shapiro Wilk o Smirnov Kolmogorov) para confirmar o no la posible distribución normal de los datos, y termina con un análisis confirmatorio, por medio de pruebas estadísticas paramétricas y no paramétricas. Como se observa en el esquema, si existe tendencia y/o falta de homogeneidad en la información se debe proceder a usar una parte de la misma o remover de ésta la tendencia o la falta de homogeneidad.

### 2.1 Análisis exploratorio gráfico

El análisis exploratorio de los datos (EDA) por medio gráfico se realiza con el fin de comprobar tendencias y cambios en la serie de tiempo por medio visual. Es considerado como el primer análisis a realizar antes de cualquier análisis confirmatorio (cuantitativo) y, más aún, antes de utilizar la información hidroclimatológica para modelos y simulaciones. Dentro del análisis exploratorio gráfico se recomienda utilizar la gráfica de serie de tiempo, el diagrama de cajas, la gráfica de doble masa y la gráfica de normalidad, descritas a continuación:

**Gráfica de series de tiempo:** representa los datos ordenados cronológicamente en las ordenadas y el tiempo en las abscisas. Son gráficos en los cuales se pueden observar claramente las tendencias, los cambios, la irreversibilidad, y la intermitencia, entre otros.

**Diagrama de cajas:** se considera un resumen de la información, ya que brinda una idea de la tendencia central, la variabilidad, la simetría y la presencia de puntos atípicos. En el diagrama de cajas se muestra el percentil 50 (la mediana), 25 y 75 (límite inferior y superior de la caja respectivamente), el menor y el mayor valor observados sin ser considerados atípicos, además de puntos atípicos (1.5 veces la longitud de la caja) y extremos (3.0 veces la longitud de la caja). Para determinar si existe un cambio en la medida de la tendencia central en la serie hidroclimatológica se divide la misma en dos o más partes, de tal forma que se pueda observar a partir del diagrama de cajas de cada una de ellas si existen diferencias entre las características estadísticas de cada una de las partes de la serie (Maidment, 1993; Sánchez, 1999; Smith & Campuzano, 2000).

**Gráfica de doble masa:** ampliamente usada para evaluar la consistencia de las observaciones a lo largo del

tiempo. Para ello, se compara la serie de estudio con una serie patrón que no presente ningún problema de homogeneidad, tendencia o cambio. Para realizar el contraste gráfico es necesario graficar en el eje de las abscisas los valores acumulados de la estación patrón y en el eje de las ordenadas los valores acumulados de la estación en estudio (Lobo, 2004). Si la serie en estudio no presenta cambios en la media, la gráfica debe manifestar una relación estable de proporcionalidad entre la estación patrón y la estación en estudio. Si por el contrario se observan quiebres en la pendiente de la gráfica, saltos o picos, se puede concluir que la serie presenta un cambio en la media y/o presenta puntos atípicos (Smith & Campuzano, 2000; Lobo, 2004).

**Gráfica de normalidad:** consiste en graficar la información en un papel de probabilidad normal. Si la gráfica muestra una línea recta, indicará que la información se distribuye normalmente, de otra manera la información no se distribuye siguiendo esta distribución y será necesario realizar una transformación a la variable. La transformación más utilizada es la propuesta por Box & Cox (1964), según la cual si  $\{x_t\}$  es una serie cronológica asimétrica se determinarán los parámetros  $a$  y  $\lambda$  tales que la serie  $\{y_t\}$  después de la transformación (Ec. (I)) minimice su asimetría:

$$y_t = \frac{(x_t - a)^\lambda}{\lambda} \quad (I)$$

Cuando  $\lambda$  tiende a cero esta transformación se convierte en:

$$y_t = \ln(x_t - a) \quad (II)$$

Si la serie original se supone distribuida log-normal entonces esta transformación la convierte en normal.

## 2.2 Análisis confirmatorio

Para describir el comportamiento estadístico de los datos hidroclimatológicos se han utilizado diversas funciones de distribución, pero de hecho la mayoría de los modelos suponen una distribución normal de la variable, lo cual implica realizar una prueba estadística confirmatoria para rechazar o no la hipótesis nula de que la distribución de los datos sigue ese tipo de distribución teórica. Las pruebas más utilizadas son las de Smirnov

Kolmogorov, Lilliefors, Cramer Von Misesy Shapiro Wilk, entre otras, descritas a profundidad en Kottegoda y Rosso (1997); Behar (1997) y Walpole *et al.* (1999), entre otros.

Posterior al análisis gráfico y a la aplicación de las pruebas de normalidad conviene realizar un análisis más riguroso a partir de técnicas estadísticas, con el fin de determinar si la serie no presenta tendencia y es homogénea y los cuales son supuestos necesarios para realizar posteriormente modelaciones y simulaciones con las series hidroclimatológicas. Para realizar el análisis confirmatorio existen numerosas pruebas estadísticas paramétricas y no paramétricas; el uso de una u otra depende de la cantidad y la calidad de la información disponible en los datos y del cumplimiento o no del supuesto de normalidad en la distribución de los datos.

La mayoría de las pruebas paramétricas suponen que los datos se asemejan a una distribución normal, son sensibles a la cantidad de datos, a las asimétricas y a la presencia de datos atípicos. Las pruebas no paramétricas, en cambio, no requieren de ningún supuesto de normalidad o de otra distribución conocida, lo que indica que son útiles bajo un amplio rango de distribuciones de la población; además, en la mayoría de los casos, los resultados estadísticos se derivan únicamente a partir de procedimientos de ordenación y recuento, por lo que su base lógica es de fácil comprensión. Cuando se trabaja con muestras pequeñas ( $n \leq 10$ ), en las que se desconoce si es válido suponer la normalidad de los datos, conviene utilizar pruebas no paramétricas, al menos para corroborar los resultados obtenidos a partir de la utilización de la teoría basada en la distribución normal. A continuación se presentan algunas de las pruebas paramétricas y no paramétricas más usadas en el análisis cuantitativo para comprobar la homogeneidad (o estacionariedad) de la serie (falta de tendencia) y detectar cambios en la varianza y en la media.

### 2.2.1 Comprobación del carácter estacionario de la serie

Para comprobar que no existe tendencia en la serie, es decir que no hay correlación entre el orden en que se tomaron las observaciones y el incremento (o decremento) en magnitud de los datos de la serie (Dahmen y Hall, 1990), se usa el método de rango de

correlación de Spearman, cuya expresión es una medida de la asociación lineal entre los rangos y números de orden de la serie original y la serie ordenada en forma creciente. También existe el coeficiente de correlación de Pearson como una medida de asociación lineal; dos variables pueden estar perfectamente relacionadas, pero si la relación no es lineal, el coeficiente de correlación de Pearson no será un estadístico adecuado para medir su asociación y, mucho menos, si el supuesto de normalidad es violado. El coeficiente de correlación de Spearman ( $R_{sp}$ ) se define como :

$$R_{sp} = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (R_i - i)^2 \quad (III)$$

donde  $i$  es el número de orden de la variable en orden cronológico y  $R_i$  es el número de orden de la observación de la serie cronológica original que ocupa la variable ordenada en forma ascendente o creciente. Para rechazar o no la hipótesis nula,  $H_0: R_{sp}=0$  (no existe tendencia), hipótesis alterna  $H_a: R_{sp} < 0$  (hay tendencia), se usa el siguiente estadístico :

$$t_t = R_{sp} \left[ \frac{n-2}{1-R_{sp}^2} \right]^{1/2} \quad (IV)$$

Donde  $t_t$  tiene una distribución t Student con  $n-2$  grados de libertad. Para un nivel de significancia ( $\alpha$ ) la hipótesis nula se rechaza si  $t_t$  se encuentra en la siguiente zona de rechazo:

$$\{-\infty, t((n-2), \alpha/2)\} \cup \{t((n-2), (1-\alpha/2)) + \infty\}$$

### 2.2.2. Estabilidad en la varianza

Detectar cambios en la varianza y en la media son pasos fundamentales para determinar si la serie presenta homogeneidad o no; el hecho de que no lo sea significa que los parámetros estadísticos varían, ya sea debido a causas naturales o antropogénicas. Se recomienda hacer en primer lugar el test para la estabilidad de la varianza, debido básicamente a dos razones : la inestabilidad de la varianza implica que la serie cronológica no es estacionaria, así que no se puede usar para análisis posteriores; y algunos tests de estabilidad en la media requieren que haya estabilidad en la varianza (Dahmen y Hall, 1990). A continuación se describe el Test F y el

Test Siegel Turkey, pruebas estadísticas usadas en el desarrollo del documento para establecer la estabilidad de la varianza, pero existen otras que pueden ser consultadas en Maidment (1993), Kotegoda y Rosso (1997); Sheskin (1997), y entre otros.

**Test F:** es una prueba paramétrica que relaciona las varianzas de dos conjuntos de información que resultan de dividir la serie hidroclimatológica en dos partes iguales. Se conoce como distribución F o Fisher a la distribución de la relación entre varianzas de muestras que vienen de una distribución normal; sin embargo, Dahmen y Hall (1990) afirman que si las muestras no vienen de una distribución normal, el Test F dará una buena estimación de la estabilidad de la varianza. El test estadístico se denota como (Snedecor y Cochran, 1983):

$$F_t = \frac{\sigma_1^2}{\sigma_2^2} = \frac{s_1^2}{s_2^2} \quad (V)$$

donde  $s^2$  representa la varianza de cada subconjunto y se calcula a partir de la siguiente expresión :

$$s = \left[ \frac{\sum_{i=1}^n x_i^2 - \frac{\left( \sum_{i=1}^n x_i \right)^2}{n}}{n-1} \right]^{1/2} \quad (VI)$$

donde  $x_i$  denota la observación y  $n$  el número total de datos en la muestra (subconjunto). La hipótesis nula para el test es,  $H_0: s_1^2 = s_2^2$  (igualdad de varianzas), y la hipótesis alterna,  $H_a: s_1^2 < s_2^2$ . La prueba se rechaza si el estadístico estimado se encuentra en la siguiente zona de rechazo, para un nivel de significancia  $\alpha$  :

$$\{0, F((n_1-1), (n_2-1), \alpha/2)\} \cup \{F((n_1-1), (n_2-1), (1-\alpha/2)), +\infty\}$$

Donde  $n_1$  y  $n_2$  son la cantidad de datos para cada uno de los subconjuntos, y  $(n_1-1)$ ,  $(n_2-1)$  son los grados de libertad de la distribución.

**Test Siegel – Tukey:** es una prueba no paramétrica utilizada para determinar si uno de los dos grupos

(subconjunto de la serie original) tiende a tener más valores extremos que el otro grupo, o dicho de otra manera determina si uno de los dos grupos presenta mayor dispersión que el otro con respecto a la medida de tendencia central (Sheskin, 1997).

### 2.2.3. Estabilidad de la media

Se puede determinar si la serie hidroclimatológica es estable en la media a partir de la comparación de subconjuntos de la información. En la mayoría de las ocasiones se recomienda dividir la serie original en dos partes, de tal forma que se puedan aplicar tests estadísticos de comparación de medias para determinar si vienen de la misma población (Maidment, 1993). Para establecer la estabilidad en la varianza existen variadas pruebas paramétricas y no paramétricas, tales como: test Test - t, U test Mann-Whitney, test Signed Rank test, test Kruskal-Wallis, test Mann-Whitney-Wilcoxon, entre otros, que pueden ser consultados en Sheskin (1997); Kottegoda y Rosso (1997). A continuación se describen dos de los test usados en el desarrollo del documento.

**Test t:** Es una prueba paramétrica que involucra el cálculo y la comparación de las medias de dos subconjuntos de la serie cronológica (los mismos subconjuntos que se usaron para determinar la estabilidad de la varianza con el Test - F). La hipótesis nula es  $H_0: \bar{X}_1 = \bar{X}_2$ ; y la hipótesis alterna,  $H_a: \bar{X}_1 \neq \bar{X}_2$ . El estadístico se calcula con la siguiente expresión:

$$t_t = \frac{\bar{X}_1 - \bar{X}_2}{\left[ \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \cdot \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \right]^{1/2}}$$

donde  $n_1$  y  $n_2$  son los números de datos en los dos arreglos subconjuntos,  $\bar{X}$  representa la media de cada subconjunto y  $s^2$  su varianza. La prueba exige que las varianzas no sean significativamente diferentes. La prueba se rechaza si  $t_t$  cae dentro de la siguiente región de rechazo para un nivel de significancia  $\alpha$ :

$$\{-\infty, t_{((n_1-1)+(n_2-1), \alpha/2)}\} \cup \{t_{((n_1-1)+(n_2-1), (1-\alpha/2)}, +\infty\}$$

donde  $(n_1-1)+(n_2-1)$  son los grados de libertad de la distribución.

**U Mann - Whitney:** es una de las pruebas no paramétricas más poderosas y constituye la alternativa más útil ante la prueba paramétrica (t) cuando el investigador desea evitar las suposiciones que ésta exige o si la medición en la investigación es más vaga que la escala de intervalo. La hipótesis nula de contraste es que las dos muestras, de tamaño  $n_1$  y  $n_2$ , respectivamente, proceden de poblaciones continuas idénticas; la hipótesis alterna puede ser unilateral o bilateral y únicamente supone que la tendencia central de una población difiere de la otra, pero no una diferencia de forma o de dispersión (Maidment, 1993; Sheskin, 1997).

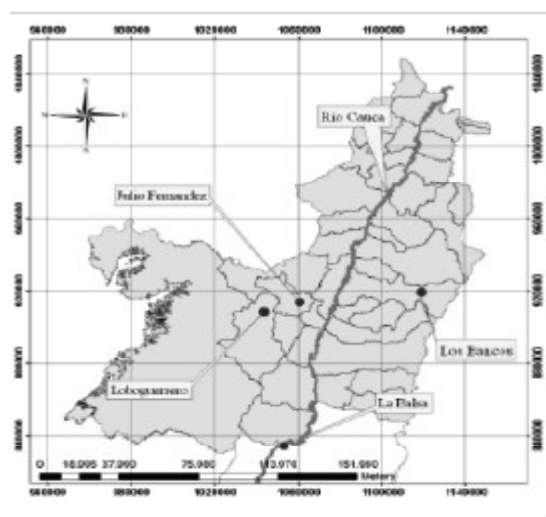
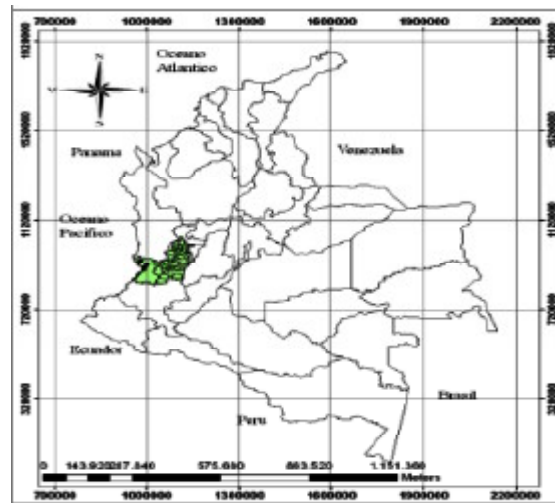


Figura 2. Zona de estudio: Dpto. Valle del Cauca (Colombia)

Tabla 1. Estadística descriptiva de las zonas de precipitación normal en las estaciones La Balsa, Julio Fernández, Loboguerrero y Los Bancos.

Estación	La Balsa	Julio Fernández	LoboGuerrero	Los Bancos
n	49	52	34	36
Años de registro	1955-2006	1968-2005	1973-2006	1971-2006
Media	2002.8	1089.4	852.2	850.8
Desvest	515.4	183.8	268.5	262.7
Varianza	265622.7	33797.9	72092	69019
Coef. Var	25.73	16.88	31.51	30.88
Min	750	773.3	528	332
Q1	1781	929.8	718.4	674.8
Mediana	2018	1098.7	764.5	819
Q3	2339	1229.3	910.8	982.3
Max	3001	1440.3	1917	1539
IQR	558	299.5	192.3	307.5
Coef. Simetría	-0.44	0.2	2.47	0.52

### 3. RESULTADOS Y DISCUSIÓN

#### 3.1. Zona de estudio e información utilizada

Para la aplicación del análisis exploratorio gráfico y confirmatorio se usaron series de precipitación total multianual de cuatro estaciones pluviográficas ubicadas en el departamento del Valle del Cauca (Colombia) (Figura 2). Se escogieron al azar las estaciones Julio Fernández, Loboguerrero, La Balsa y Los Bancos, actualmente manejadas por la Federación Nacional de Cafeteros y la Corporación Autónoma Regional del Valle del Cauca (CVC). Las estaciones presentaron períodos de registro que variaron entre 34 y 52 años, y presentaron, en términos generales, menos del 10% de datos faltantes, que fueron completados haciendo uso del método racional deductivo (Chávarri, 2005).

Donde n es el número total de datos, Desvest representa la desviación típica el Coef. Var el coeficiente de variación, min el menor valor, max el mayor valor de la serie de datos, Q1 el primer cuartíl, Q3 el tercer cuartíl, IQR el rango intercuartilico.

#### 3.2 Aplicación del análisis exploratorio gráfico y cuantitativo

Previa a la aplicación de las pruebas gráficas y confirmatorias para el análisis exploratorio de datos se procedió a completar los datos faltantes de las series de precipitación con el uso del método racional deductivo (Chavarrí, 2005) que permite estimar los registros apo-

yándose en la información que brindan los años completos de cada serie.

De los estadísticos descriptivos mostrados en la Tabla 1 se puede decir que las medidas de tendencia central, media y mediana para las estaciones La Balsa, Julio Fernández y Los Bancos no muestran una diferencia mayor al 10%, comparando la mediana con respecto a la media, excepto la estación Loboguerrero que presenta una mediana que varía en un 11% con respecto al valor de la media. Los coeficientes de variación de las series no sobrepasan el 32%, para el caso de Loboguerrero los datos tienden en promedio a estar dispersos en un 31% del valor de la media. Las series presentan asimetrías positivas y negativas, y es particularmente alta la de la estación Loboguerrero influenciada por valores extremos hacia la derecha.

En las Figuras 3 a 6 se muestra la aplicación de las gráficas de series de tiempo, diagramas de cajas, gráfica de doble masa y la gráfica de probabilidad. Las gráficas de series de tiempo no muestran ninguna tendencia definida, pero La Balsa, Loboguerrero y Los Bancos muestran una alta dispersión en los datos en algunas partes de las series. Se observa en la estación La Balsa un salto en el año 1968, que puede producir un cambio en la medida de tendencia central y también en la varianza (suposición que será posteriormente comprobada con el análisis confirmatorio). El diagrama de cajas, por su parte, muestra que en algunas de las series existe una diferencia en la dispersión de los datos; es así como las estaciones La Balsa, Loboguerrero y Los Bancos muestran rangos intercuartilicos en cada

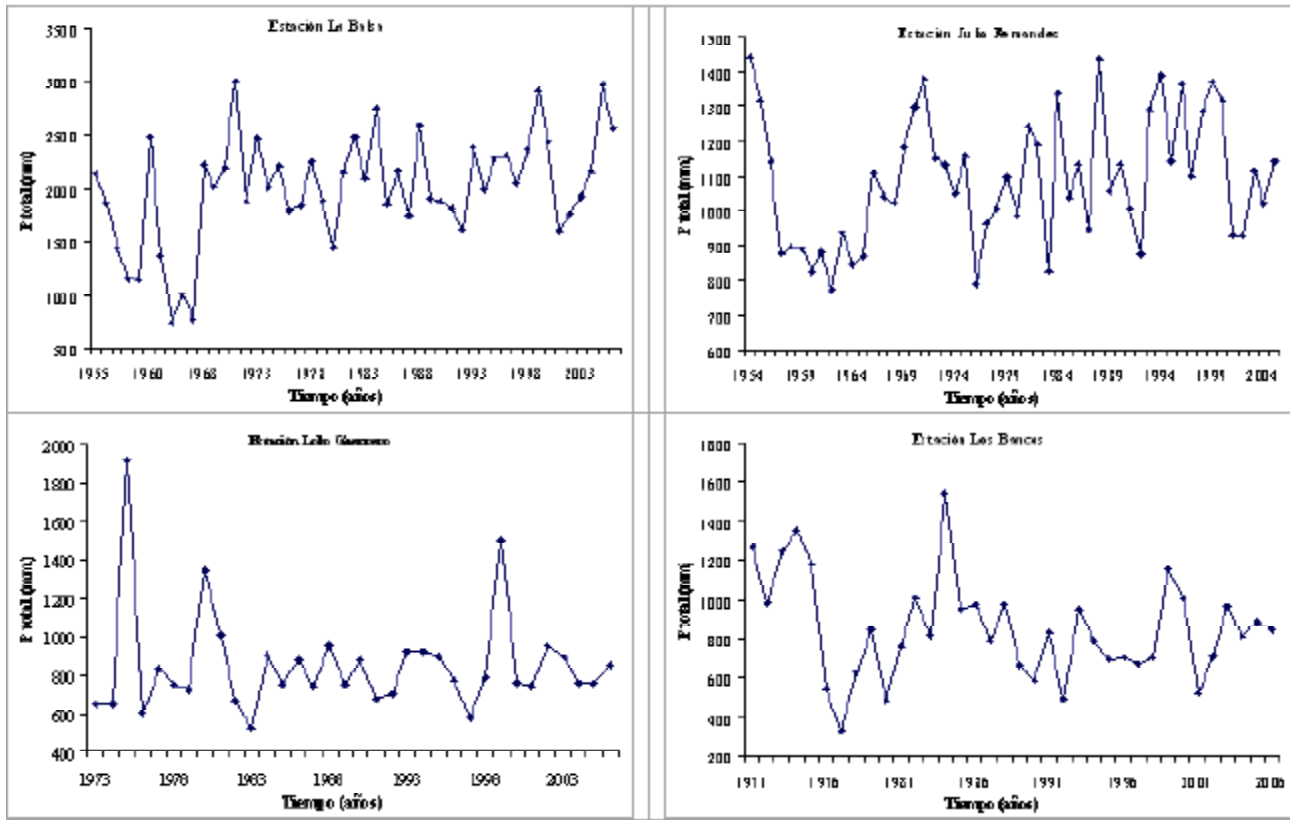


Figura 3. Precipitación anual en las estaciones La Bolsa, Loboguerrero, Julio Fernández, Los Bancos

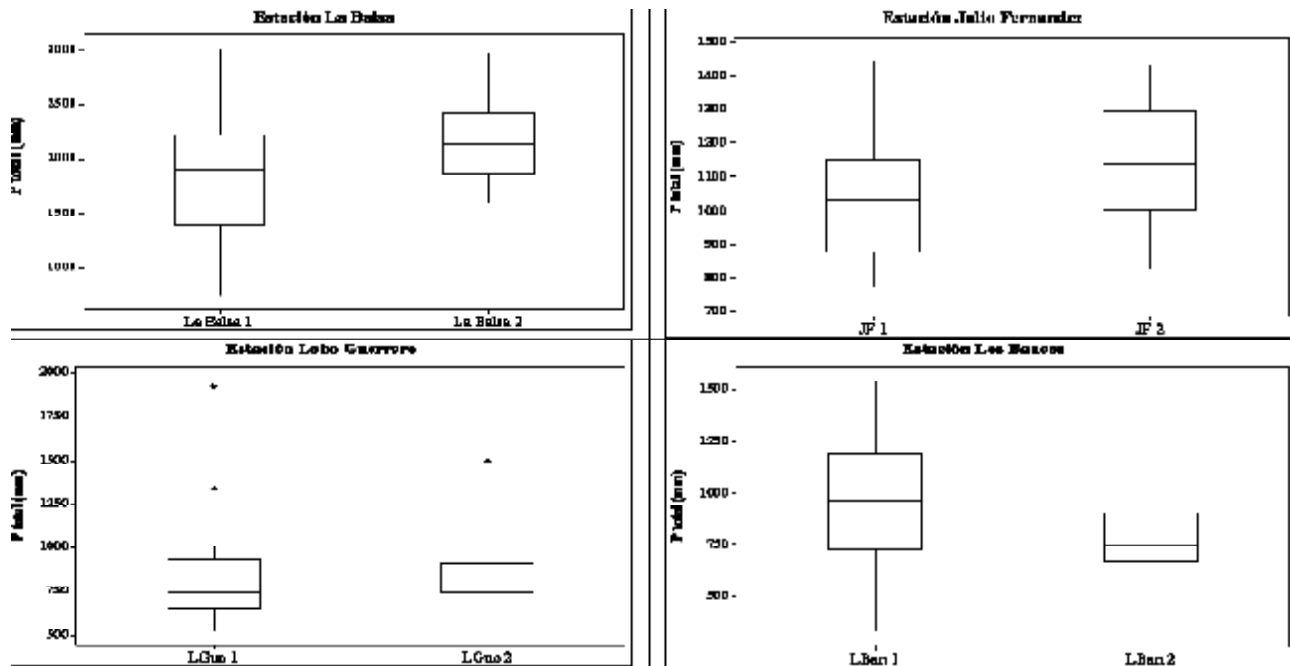


Figura 4. Diagramas de cajas de la precipitación anual en las estaciones, La Bolsa, Loboguerrero, Julio Fernández, Los Bancos



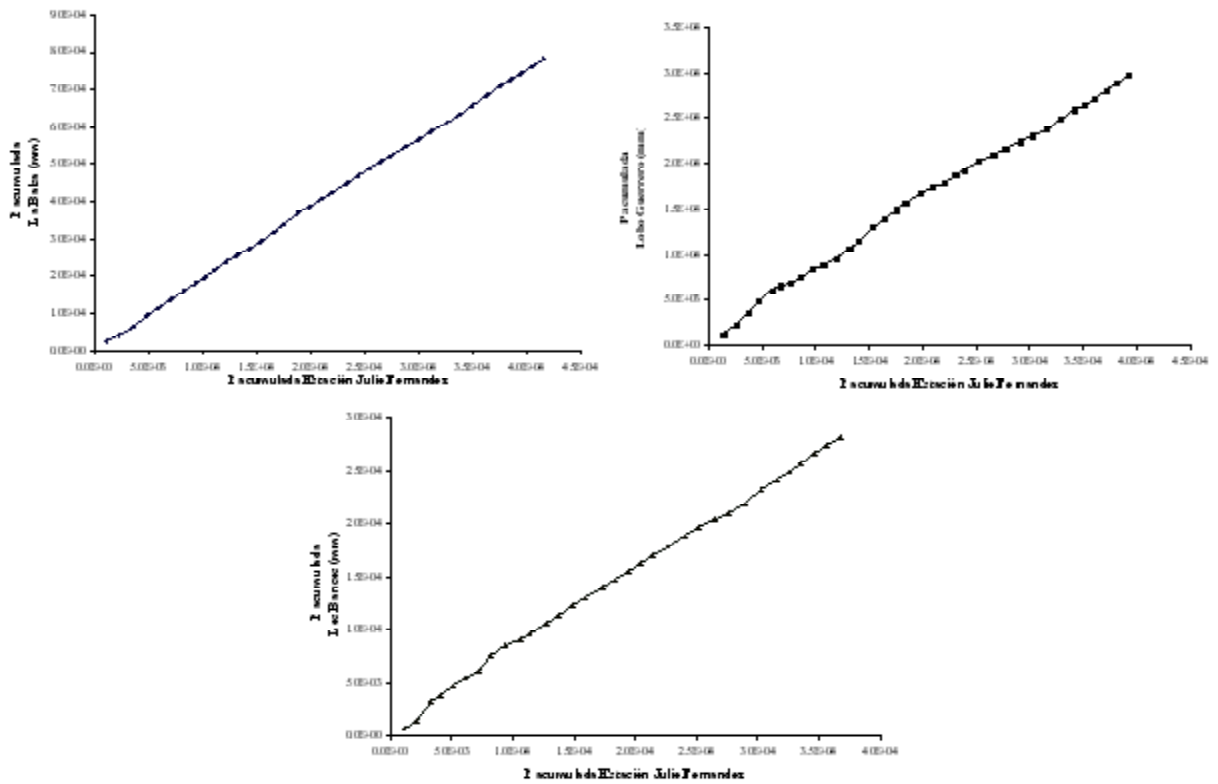


Figura 5. Curvas de Doble Masa

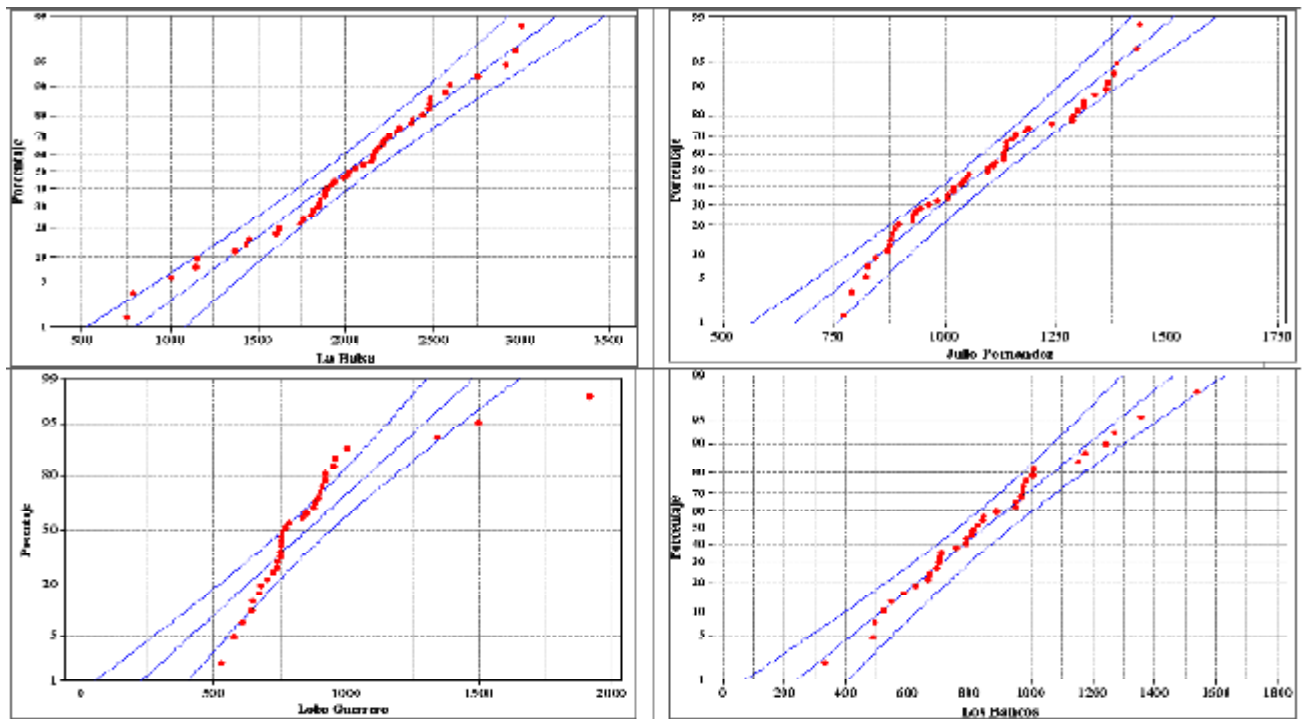


Figura 6. Curvas de probabilidad acumulada.

Tabla 2. Pruebas estadísticas de carácter confirmatorio.

Test	La Balsa	Julio Fernández	Lobo Guerrero	Los Bancos
<b>Normalidad</b>				
Shapiro Wilk	p-value > 0.10	p-value > 0.10	p-value < 0.010	p-value > 0.10
Smirnov Kolmogorov	p-value > 0.15	p-value > 0.15	p-value < 0.010	p-value > 0.15
<b>Tendencia</b>				
Coefficiente de correlación de Spearman	NR	NR	NR	NR
<b>Estabilidad en la varianza</b>				
Test - F	p-value = 0.062	p-value = 0.752	p-value = 0.045	p-value = 0.016
Test Siegel - Tukey			p-value=0.848	
<b>Estabilidad en la media</b>				
Test - t	p-value = 0.02	p-value = 0.057	p-value = 0.848	p-value = 0.087
Test U Mann - Whitney	p-value = 0.065	p-value = 0.067	p-value = 0.3892	p-value = 0.0818

NR: No se rechaza la prueba

\*\*Todas las pruebas fueron realizadas para un nivel de significancia del 5%

subconjunto de la serie que parecen diferir en forma significativa una de la otra, lo que puede manifestar falta de estabilidad en la varianza.

Por otra parte, el gráfico de doble masa acumulada, que por años ha sido empleado en el análisis de consistencia de la información, permite detectar aquellas series que presentan un error sistemático asociado a la toma de datos debido al uso de diferentes instrumentos y/o técnicas de medición durante un período considerado. La serie patrón usada para la aplicación del método gráfico fue la de la estación Julio Fernández debido a que es reconocida como de alta calidad o fidelidad, es permanentemente monitoreada, presenta continuidad, no tiene datos faltantes y exhibe confiabilidad de los datos producidos. La gráfica de doble masa acumulada de la estación La Balsa muestra un leve cambio de pendiente a partir del año 1980 hasta 1984, pero en general se recomienda que el cambio de pendiente se mantenga en el tiempo, 5 años o más, antes de considerar una corrección, para eliminar pequeñas fluctuaciones que pueden producirse y que no indican una falta de consistencia en los registros (Varas, 2003). Sin embargo, las estaciones Loboguerrero y Los Bancos muestran variadas pendientes con periodos mayores a 5 años y que pueden ser corregidos aplicando un coeficiente de corrección igual a la razón entre las pendientes (Lobo, 2004).

Con respecto a las gráficas de probabilidad, se puede afirmar que únicamente los datos de la estación Loboguerrero no presentan una distribución normal, resultado coincidente con el obtenido en las pruebas

estadísticas de Shapiro Wilk y Smirnov Kolmogorov. Lo anterior indica el tipo de pruebas (parámbricas o no parámbricas) a usar para determinar la estabilidad de la varianza y de la media.

En el caso de la estación Loboguerrero deben ser de carácter no parámtrico y para el resto de las estaciones se pueden usar indistintamente las parámbricas como las no parámbricas.

De las pruebas estadísticas de carácter confirmatorio, la aplicación del coeficiente de Spearman mostró, como se observa en la Tabla 2, que ninguna de las cuatro estaciones presenta tendencia en la serie cronológica; las estaciones La Balsa y Julio Fernández no mostraron diferencias significativas en la varianza para los subconjuntos de series usadas en el análisis, pero la que sí mostró diferencias significativas entre los subconjuntos fue la estación Los Bancos, resultado congruente con el mostrado en el diagrama de cajas. La estación Loboguerrero, según el test Sigel Tukey, no presenta cambios significativos en el valor de la dispersión de los dos grupos formados. Respecto a la estabilidad de la media, ninguna de las estaciones tuvo diferencias significativas en el valor de la medida de tendencia central y se concluye que para todas ellas, los subconjuntos formados proceden de poblaciones continuas idénticas, es decir, proceden de la misma población y, por lo tanto, sus propiedades estadísticas no difieren una de la otra.

#### 4. CONCLUSIONES

La pruebas gráficas son herramientas imprescindibles

en el análisis exploratorio de datos y permiten identificar, de forma simple, si la serie se distribuye normalmente o no, si presenta tendencias y cambios, y el periodo donde éstos se producen.

Debido a la facilidad que representa realizar las pruebas gráficas, es aconsejable utilizar más de una al momento de realizar el EDA por medio gráfico, pues algunas de las gráficas, como se mostró y explicó en el análisis de los resultados, pueden fallar en detectar problemas o pueden sugerirlos, y éstos pueden o no existir realmente. Es aquí donde el criterio del analista juega un papel importante.

El análisis cuantitativo como herramienta en el EDA es valioso debido a que confirma estadísticamente la homogeneidad de la varianza y de la media, además de la estacionariedad, condiciones básicas en cualquier simulación o modelación que haga uso de series hidroclimatológicas.

Las herramientas presentadas en este documento muestran la variedad de alternativas que el analista tiene para tomar decisiones respecto a las series que está usando, pero es importante conocer claramente con qué clase de datos se está trabajando y cuáles pueden ser las causas de error en las mismas.

Cuando el análisis exploratorio de datos muestre que las series son inconsistentes, con tendencias, cambios, etc, se hará necesario la remoción de esa parte de la serie, lo que implica una disminución en la cantidad de información disponible para realizar cualquier tipo de análisis, además que la serie se convierte en una serie artificial, manipulada por el analista.

## 5. AGRADECIMIENTOS

Los autores agradecen a la Corporación Autónoma Regional del Valle del Cauca a la Federación Nacional de Cafeteros por la información climatológica brindada y al Grupo de Investigación en Ingeniería de Recursos Hídricos y Desarrollo de Suelos- IREHISA de la escuela EIDENAR de la Universidad del Valle por el apoyo con los recursos necesarios para la realización de este trabajo.

## 6. REFERENCIAS BIBLIOGRÁFICAS

Behar, R. (1997). Comprendiendo la Estadística: Usando el sentido Común. Universidad del Valle.

Box, G. y Cox, D. (1964). An analysis of transformation. J.R. Statistic Soc. B-26, pp. 211-

252.

Chavarri, E. (2005). Curso: Modelos matemáticos en Hidrología. Escuela de Postgrado Universidad Nacional Agraria La Molina. <http://tarwi.lamolina.edu.pe/~echavarri/>.

Dahmen, E. y Hall, M. (1990). Screening of Hydrological Data: Tests for Stationary and Relative Consistency. International Institute for Land Reclamation and Improvement – ILRI. Wageningen. The Netherlands.

Kottegoda, N. y Rosso, R. (1997). Statistics, Probability and Reliability for Civil and Environmental Engineers. The McGraw – Hill Companies, Inc. United States of America.

Lobo, L. (2004). Guía Metodológica para la Delimitación del Mapa de Zonas Áridas, Semiáridas y Subhúmedas Secas de América Latina y el Caribe. Centro del Agua para Zonas Áridas y Semiáridas de América Latina y El Caribe – CAZALAC – UNESCO PHI – Gobierno de Flandes. [http://www.cazalac.org/mapa\\_alc\\_guia.php](http://www.cazalac.org/mapa_alc_guia.php).

Mesa, O., Poveda, G., y Carvajal, L. (1997). Introducción al Clima de Colombia. Universidad Nacional de Colombia. Bogotá. Colombia.

Maidment, D. (1993). Handbook of Hydrology. McGRAW – HILL, INC. United States of America.

Sánchez, J. (1999). Manual de análisis estadístico de los datos. Segunda edición. Alianza Editorial S.A. Madrid.

Sheskin, D. (1997). Handbook of Parametric and Nonparametric Statistical Procedures. Western Connecticut State University. CRC Press.

Smith, R. y Campuzano, C. (2000). Análisis exploratorio para la detección de cambios y tendencias en series hidrológicas. XIV Seminario Nacional de Hidráulica e Hidrología.

Snedecor, G.W. y Cochran, W. G. (1989). Statistical Methods. Eighth Edition, Iowa State University Press.

This document was created with Win2PDF available at <http://www.win2pdf.com>.  
The unregistered version of Win2PDF is for evaluation or non-commercial use only.  
This page will not be added after purchasing Win2PDF.