

# BOXPLOTS Y MEDIA MÓVIL: MÉTODOS SENCILLOS DE INTERPRETACIÓN DE SERIES TEMPORALES DE PRECIPITACIÓN EN LA CUENCA DEL PASTAZA

<sup>1</sup>Julia Calahorrano, <sup>1</sup>Mery Manzano, <sup>1</sup>Dolores Gualli, <sup>2</sup>Fabián Silva, <sup>3</sup>Mario Audelo

<sup>1</sup>Ingeniería Ambiental, <sup>2</sup>Ingeniería Industrial, Universidad Nacional de Chimborazo, Riobamba, Ecuador. <sup>3</sup>Facultad de Mecánica, Escuela Superior Politécnica de Chimborazo, Riobamba, Ecuador.

## Resumen

Región de estudio: Cuenca del río Pastaza, Ecuador.

El análisis del comportamiento temporal de la precipitación es importante para el manejo sustentable del recurso hídrico en una cuenca hidrográfica determinada. En este estudio, la variabilidad, tendencia, estacionalidad y atipicidad en la precipitación, fueron evaluados utilizando la red del Instituto Nacional de Meteorología e Hidrología (INAMHI) compuesta por 23 pluviómetros (1964-2011). Las medidas de tendencia central y dispersión; media móvil y boxplots fueron aplicadas para observar las variables mencionadas. Los resultados evidencian: (a) la precipitación en la cuenca muestra variabilidad con valores que fluctúan entre los 100 mm y 4500 mm anuales. b) el 60.8 % de estaciones estudiadas presentan tendencia al incremento o decremento en la precipitación. (c) 82.6 % de los casos revela estacionalidad, con bimodalidad, registrándose dos periodos lluviosos: febrero - abril y, octubre - noviembre; y unimodalidad, cuyo periodo oscila entre abril y junio. (d) la totalidad de estaciones registraron atipicidad (outliers). Estos resultados proporcionan información para entender el comportamiento de la precipitación a nivel regional en zonas de montaña, aportando a la construcción de modelos de cambio climático global.

**Palabras claves:** Series temporales, Precipitación, Cuenca del Pastaza, Tendencia, Variabilidad, Outliers.

## Abstract

Study region: Basin Pastaza, Ecuador river.

The analysis of the temporal behavior of precipitation is important for the sustainable management of water resources in a given watershed. In this study, variability, trend, seasonality and unusualness in precipitation were evaluated using the network of the National Institute of Meteorology and Hydrology (INAMHI) composed of 30 rain gauges (1964-2011). Measures of central tendency and dispersion; moving average and boxplots were applied to observe the variables. The results show: (a) precipitation in the basin shows variability with values ranging between 100 mm and 4,500 mm annually. b) 60.8% of stations studied have a tendency to increase or decrease in precipitation. (C) 82.6% of cases reveals seasonality, with bimodality, registering two rainy periods: from February to April and from October to November; and unimodality, whose period is between April and June. (D) all stations recorded atipicidad (outliers). These results provide information for understanding the behavior of precipitation at the regional level in mountain areas, contributing to building models of global climate change.

**Keywords:** Temporal series, Precipitation, Pastaza basin, Trend, Variability, Outliers.

## INTRODUCCIÓN

El clima tiene una acción decisiva sobre el mundo mineral, vegetal y animal: meteorización de las rocas, génesis y erosión de los suelos, distribución geográfica de la flora,

fauna y asentamientos humanos, crecimiento de plantas y cultivos, influencia sobre la fisiología y salud del hombre, fuente de los mayores beneficios o de las

peores catástrofes naturales (1). Una de las variables más importantes que influyen tanto espacial como temporalmente en el clima y en la disponibilidad del recurso hídrico es la lluvia (2).

Las medidas de tendencia central y dispersión pueden dar pautas acerca de este comportamiento espacio-temporal. La media móvil aporta al establecimiento de ciclos estacionales, porque reduce la variabilidad interanual de las series de tiempo (3) y; los boxplots son herramientas útiles para identificar outliers (4).

Muchos estudios han investigado el comportamiento espacial y temporal y las tendencias de la precipitación histórica en diversas partes del mundo. Así, las tendencias de las series de precipitación en México registran cambios importantes durante el período considerado (1920-2004), dándose a escalas regionales y consistentes. En términos geográficos, el incremento de precipitación se registra especialmente en regiones áridas y semiáridas, mientras que la disminución de lluvia se hace más evidente en el centro y costas del Golfo de México. Estacionalmente, el verano es la época del año que registra el mayor porcentaje de estaciones con tendencias estadísticamente significativas de incremento y decremento de precipitación ( $p \leq 0.1$ ). (5). En la cuenca media-alta del río Cauca, en Colombia, se detectaron cambios significativos sobre la precipitación media anual en 24 (22 % del total) estaciones de medición. Al considerar estos registros, se determinó una reducción promedio -0.7% por decenio (equivalente a -7,7 mm/década) en el monto total de la precipitación media anual entre 1975-2006 (6). El análisis de tendencias de las precipitaciones anuales en la cuenca del lago de Cuitzeo en la región de Chapal, México, revela que en los últimos siete años todas las estaciones tienden a incrementar sus precipitaciones. En general los comportamientos son muy variables para todas las estaciones a lo

largo de los años y esta variabilidad interanual puede corresponder con las variaciones que se presentan en las variables año tras año (7).

Los objetivos del presente estudio fueron: (a) determinar la variabilidad en la precipitación interanual (b) estimar la tendencia temporal en el periodo de estudio (c) investigar la estacionalidad de la precipitación sobre el área de estudio y (d) identificar valores atípicos (outliers) en las series de precipitación de las estaciones de la cuenca del río Pastaza.

## MATERIALES Y MÉTODOS

### Área de estudio

La cuenca del río Pastaza en Ecuador, se encuentra localizada en la región oriental del país, en las coordenadas 77°50' 00" W latitud a 79°0'0.0" W y desde 0° 35'0.0" S a 2° 0' 0.0" S longitud, comprende una superficie de 33.280 km<sup>2</sup> con elevaciones desde los 700 a 5.200 m.s.n.m. (Fig 1). Para el estudio se utilizó 23 estaciones de la red meteorológica del Instituto Nacional de Meteorología e Hidrología, con un registro de datos de precipitación diarios de entre 24 y 48 años en el periodo 1964 - 2011.

### Metodología

Este estudio investiga existencia de variabilidad, tendencia, estacionalidad y atipicidad (outliers) de la precipitación en la cuenca del río Pastaza en Ecuador utilizando técnicas de estadística descriptiva e inferencial.

La tendencia y la variabilidad de una serie temporal pueden ser descritas en forma general a través de las medidas de posición y de dispersión (8). Estas medidas son ampliamente utilizadas para resumir conjuntos de datos y para medir la variabilidad de las observaciones climatológicas (9). Las series temporales de precipitación, fueron evaluadas en este estudio, con la media [1] y acumulada anual; se calculó la desviación estándar [2], el coeficiente de variación anual [3]. Se estableció las medias móviles [4] sobre las acumuladas anuales con el propósito de observar una correlación suavizada entre la precipitación y el tiempo. La media móvil es un procedimiento convencional para reducir la variabilidad interanual de las series de tiempo. (3). La investigación busca observar si las series muestran ciclos estacionales, por lo que se utiliza una media móvil simple con periodos de 7 años como herramienta para eliminar la fluctuación periódica, aunque raramente un ciclo perfectamente regular es difícil observar.

Media	$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$	[1]
Desviación Estándar	$\sigma = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}}$	[2]
Coefficiente de Variación	$C.V = \frac{\sigma}{\bar{X}} * 100\%$	[3]
Media Móvil	$\bar{X}_t = \frac{\sum_{t=1}^n \bar{X}_{t-1}}{n}$	[4]

La variabilidad fue comprobada a través de pruebas como el Análisis de Varianza (ANOVA) que permite comparar las medias de grupos. El ANOVA compara varios tratamientos para determinar la igualdad de los promedios. Para las pruebas del análisis de varianza se usa la distribución F. Esta distribución no es simétrica, sino sesgada, es decir, oblicua hacia la derecha. (10). La hipótesis nula para cada estación es  $H_0 =$  las medias anuales de precipitación son iguales y  $H_1 =$  las medias de precipitación son diferentes.

La ANOVA, se establece mediante las siguientes ecuaciones:

$$SCE = \sum_{j=1}^r n_j (\bar{x}_{.j} - \bar{x}_{..})^2$$

$$SCD = \sum_{j=1}^r \sum_{i=1}^n (\bar{x}_{ij} - \bar{x}_{.j})^2$$

$$SCT = \sum_{j=1}^r \sum_{i=1}^n (\bar{x}_{ij} - \bar{x}_{..})^2$$

La visualización cualitativa de los datos se integró con el uso de técnicas estadísticas para identificar valores que pueden estar ocultos en una serie numérica; en este caso se utilizó los boxplots (diagramas de cajas), porque suavizan los detalles numéricos conservando información importante acerca de la distribución (11). El diagrama de caja utiliza la mediana, los cuartiles aproximados y los puntos de datos mínimo y máximo para expresar la curtosis, amplitud y simetría de una distribución de valores. Los boxplots son ampliamente utilizados como una manera simple pero poderosa visualmente para interpretar la distribución de un conjunto de números (12). (Fig. 2). Los boxplots permiten además revelar la presencia de outliers. Un outlier es una observación que se desvía mucho de las otras observaciones como para despertar las sospechas de que ha sido generado por un mecanismo diferente. (13). Los outliers pueden provenir de errores en el registro, cambios en el entorno de la estación, alteración de equipos, etc o pueden responder a eventos particulares o raros de precipitación. (14). Estos aspectos no se pudieron establecer en este estudio puesto que no se dispone de los historiales de las estaciones y no se constituye en el objetivo de la investigación.

La existencia de ciclos estacionales predominantes en

las series de precipitación de las estaciones de la cuenca del Pastaza se determinó mediante los boxplots. Una serie es estacional cuando podemos observar en ella un patrón sistemático que se repite periódicamente (cada año, cada mes, etc.) dependiendo de las unidades de tiempo en que vengan recogidos los datos. (15)

## RESULTADOS Y DISCUSIÓN

Los parámetros estadísticos (media, desviación estándar, coeficiente de variación) de la precipitación anual y mensual han sido analizados para cada estación, que se complementó con un ANOVA.

### Variabilidad interanual de la precipitación

El comportamiento de las estaciones asentadas en la cuenca del Pastaza, muestra que en las unidades hidrográficas Patate y Chambo se tiene precipitaciones anuales que fluctúan entre los 100mm y los 3.500 mm pero con una variabilidad interanual mayor que se evidencia a través de la desviación estándar y el coeficiente de variación; mientras en la unidad hidrográfica Pastaza la precipitación anual es más elevada oscilando entre los 800 mm a los 4.500 mm, pero con una variabilidad menos marcada.

La prueba ANOVA, aplicada a todas las estaciones, confirma que las medias interanuales son diferentes a lo largo de los periodos de estudio. Como ejemplo se ha tomado tres casos con el mismo número de años de registro. (Tabla 1). El primero que corresponde a una estación ubicada en la unidad hidrográfica Patate, la estación Pillaro; el segundo a la unidad hidrográfica Pastaza, la estación Baños y el tercero a la unidad hidrográfica Chambo, la estación Guaslán. En todos los casos se determina que se rechaza la hipótesis nula, por lo tanto se ratifica que existe una elevada variabilidad interanual, es decir, las medias son distintas

a lo largo de periodo de estudio.

### **Tendencia temporal de la precipitación anual**

El uso de las medias móviles es una manera de suavizar la gran dispersión de datos de las precipitaciones y observar así la tendencia general y los ciclos. La tendencia de las precipitaciones anuales en la cuenca del Pastaza, presenta un comportamiento diverso (Tabla 2), así en la unidad hidrográfica Patate, donde existen 12 estaciones, la tendencia al incremento en la precipitación se presenta en 6 casos, decremento en 4 casos y en dos no se presenta tendencia. Por su parte, en la unidad hidrográfica Pastaza, con 4 estaciones pluviométricas, no se observa tendencia, salvo en el caso de la estación Sangay donde se apunta a un decrecimiento. Finalmente en la unidad hidrográfica Chambo, con 7 pluviómetros, se advierte tendencia al decrecimiento en 3 casos, sin tendencia en otros 3 y un caso con tendencia al crecimiento. Estas diferencias obedecen a la amplitud del periodo de datos, al peso que tienen en la regresión lineal los diversos periodos húmedos o secos, y a la ubicación geográfica de las unidades hidrográficas de estudio.

En la figura 3, se describen tres casos seleccionados para la unidad hidrográfica Patate. En la estación Cotopaxi -Clirsen se muestra una línea de tendencia hacia el decrecimiento; en la estación P. Fermín Cevallos la línea de tendencia muestra un incremento, mientras que en la estación Huambalo la línea de tendencia no muestra ninguna disposición ni al incremento ni al decremento.

En cuanto a la presencia de ciclos estacionales, la media móvil de siete años (en la que cada valor anual viene definido por el promedio de la década anterior), describe la presencia de periodos de aproximadamente 14 años. (Fig.4). No obstante, el distinto periodo temporal que abarcan las medias móviles dificulta posibles comparaciones entre

las estaciones.

### **Estacionalidad en la cuenca del Pastaza**

Una herramienta útil para completar el análisis del nivel y estacionalidad de una serie es la aplicación de diagramas boxplot. En este caso se utilizó para el reconocimiento de los ciclos estacionales predominantes en las series de precipitación de las estaciones de la cuenca del Pastaza. En la figura 5, se muestran ocho estaciones con sus respectivos boxplots mensuales, que corresponden a las tres unidades hidrográficas estudiadas. Las estaciones Patate, Puyo y Alao muestran ciclos unimodales, con un máximo bien definido en el mes de junio. Las estaciones Aereopuerto, Sangay y Guaslán, revelan ciclos bimodales, con dos máximos anuales en los meses de abril y octubre. Las estaciones de Tisaleo y Licto, no muestran estacionalidad y corresponden a un nivel circunstancial, con picos en meses variados, eventos que se presentan únicamente en estas dos estaciones.

### **Valores atípicos (Outliers) en los registros de precipitación**

Los boxplots de cada estación identificaron la existencia de valores atípicos (outliers) La figura 6, muestra tres histogramas que resumen el porcentaje de valores atípicos en las estaciones de Cotopaxi-Clirsen, Cusubamba y Puyo, cuyos porcentajes son variados, con máximos de 20% en agosto; 23,13% en enero y 8,49% en agosto, respectivamente. Estos valores no necesariamente pueden ser atribuidos a eventos de precipitación, pues podrían corresponder a errores de registro, errores de manipulación de equipos, etc.

## **CONCLUSIONES**

Este artículo resume el estudio de la variabilidad, tendencia, estacionalidad y atipicidad presente en 23 estaciones de la cuenca del Pastaza, durante el periodo 1964 a 2011, utilizando técnicas de estadística descriptiva e inferencial. Las conclusiones de estudio son las siguientes:

- Los resultados estadísticos muestran que en todas las estaciones se registra variabilidad temporal a lo largo del periodo de estudio, que se manifiesta con mayor incidencia en las unidades hidrográficas Patate y Chambo, donde la topografía presenta mayor irregularidad, lo que tiene efecto sobre los montos pluviométricos anuales, que pueden ser de importancia para la planificación hídrica.
- La tendencia es negativa en el 34,8% de los casos; positiva en el 30,4% y sin tendencia en 34,7%. El análisis sugiere ciclicidad de alrededor de 14 años en promedio. La



media móvil es una técnica de estimación de la tendencia y ciclicidad muy útil, que puede complementarse con la utilización de test no paramétricos.

- La estacionalidad sugiere que los regímenes pluviales históricos se mantienen. En la franja norte de la cuenca del Patate se evidencia bimodalidad, es decir, dos periodos lluviosos en abril y noviembre; en la parte suroccidental un periodo en los meses entre abril y junio. Las estaciones ubicadas en la parte sureste de la cuenca del Chambo muestran comportamiento unimodal con un periodo lluvioso entre abril y junio; mientras que aquellas ubicadas en la zona interandina y occidental presentan bimodalidad con un periodo lluvioso entre febrero y abril y otro entre octubre y noviembre. En la unidad hidrográfica Pastaza en cambio el nivel es inestable sin tendencia.

- La atipicidad está presente en todas las estaciones. La presencia de valores atípicos puede provenir de errores en el registro, cambios en el entorno de la estación, al-

teración de equipos, etc o, responder a eventos particulares o raros de precipitación. El establecimiento del origen de estos valores no ha sido el objetivo de este estudio, que de ser el caso, requeriría del historial de las estaciones.

**AGRADECIMIENTOS:**

Agradecimientos al Instituto Nacional de Meteorología e Hidrología (INAMHI) por facilitar los registros diarios de la red de estaciones meteorológicas de la cuenca del Pastaza, para el presente estudio. Al H. Gobierno provincial de Tungurahua a través de la Dirección de Recursos Hídricos y el programa PACT por facilitar información respecto de la cuenca.

**ANEXOS**

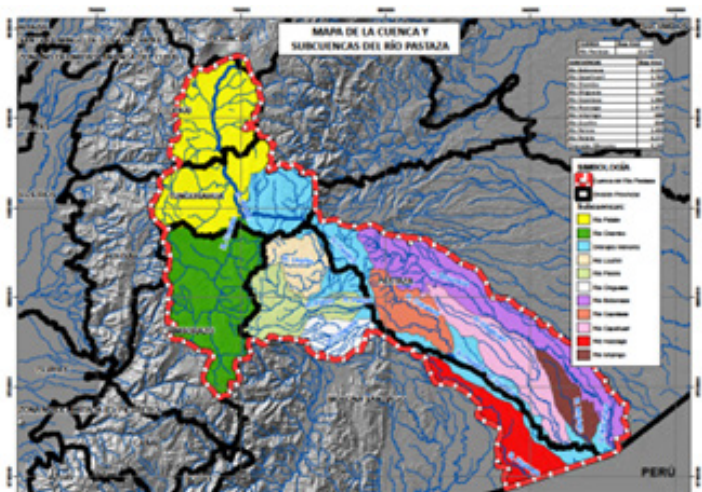


Figura 1. Cuenca del río Pastaza en Ecuador

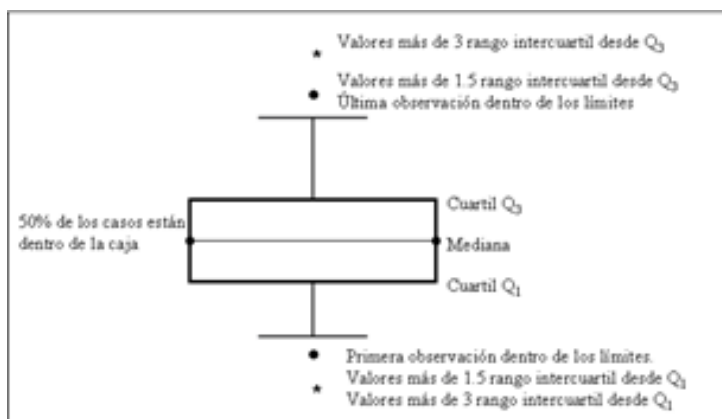


Figura 2. Estructura de un Boxplot

Estación	Media global	Media Desviación Estándar	SCE	SCD	SCT	Valor de F calculado	Valor de F crítico $\alpha = 0.05$ $n = 48$ $r = 12$	Observación
<b>Unidad hidrográfica Patate</b>								
Pillaro	1,70	2,70	4.279,61	1.538,08	2.594,05	9,10	2,05	Se rechaza Ho
<b>Unidad hidrográfica Pastaza</b>								
Baños	3,92	5,62	10.648,67	7.789,17	10.202,87	4,47	2,05	Se rechaza Ho
<b>Unidad hidrográfica Chambo</b>								
Guastán	1,73	3,20	7.676,73	2.701,08	4.331,98	9,30	2,05	Se rechaza Ho

Tabla 1. ANOVA de la precipitación mensual de tres estaciones seleccionadas.

Estación	Media mm	STD	CV %	Línea de tendencia
<b>Unidad hidrográfica Patate</b>				
Cotopaxi-Clirsen	1109	269,9	34,4	Decremento
Rumipamba	541	111,2	20,5	Incremento
Pastocalle	859	381,2	44,3	Decremento
Saquisilí	244	269,8	34,4	Incremento
Patate	2.524	125,2	5	Decremento
Pillaro	604	137,6	22,8	Decremento
Huambaló	835	461,7	55,9	Sin tendencia
Aeropuerto	550	113,2	20,6	Sin tendencia
Querochaca	595	104,8	17,5	Incremento
P. F. Cevallos	518	125,3	24,2	Incremento
Tisaleo	732	259,9	35,5	Incremento
Pilahuin	695	125,2	17,4	Incremento
<b>Unidad Hidrográfica Pastaza</b>				
Puyo	4.500	421,6	9,4	Sin tendencia
Sangay	3.843	421,8	10,7	Decremento
Río Verde	2.968	423,7	14,27	Sin tendencia
Baños	1.401	416,4	29,7	Sin tendencia
<b>Unidad Hidrográfica Chambo</b>				
Alao	1.018	368,2	36,14	Decremento
Licto	1.218	621,1	51	Decremento
Guano	475	121,2	25,62	Decremento
Cebadas	503	212,8	42,3	Incremento
Guaslán	635	180,04	28,32	Sin tendencia
Cusubamba	567	184,3	32,5	Sin tendencia
San Juan	690	154,57	22,38	Sin tendencia

Tabla 2. Tendencia calculada a través de las medidas de tendencia central y de dispersión para todas las estaciones de la cuenca del Pastaza

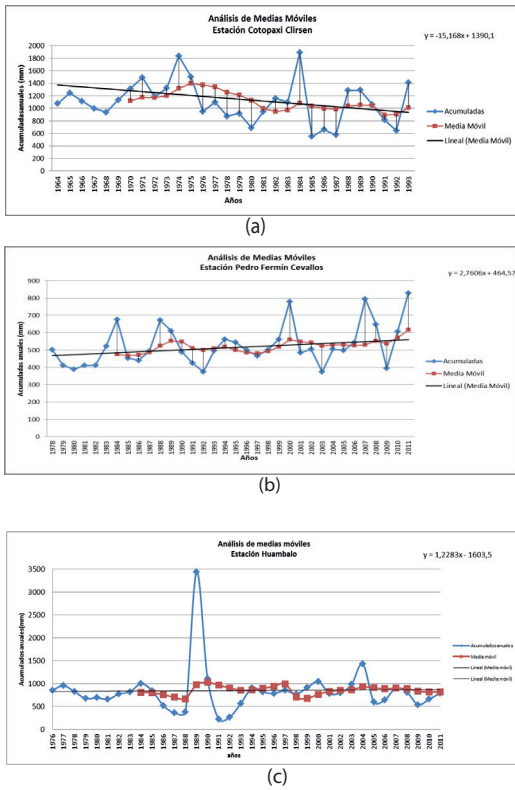


Figura 3. Media móvil anual. (a) estación Cotoxaxi-Clirsen; (b) estación Pedro F. Cevallos; (c) estación Huambaló

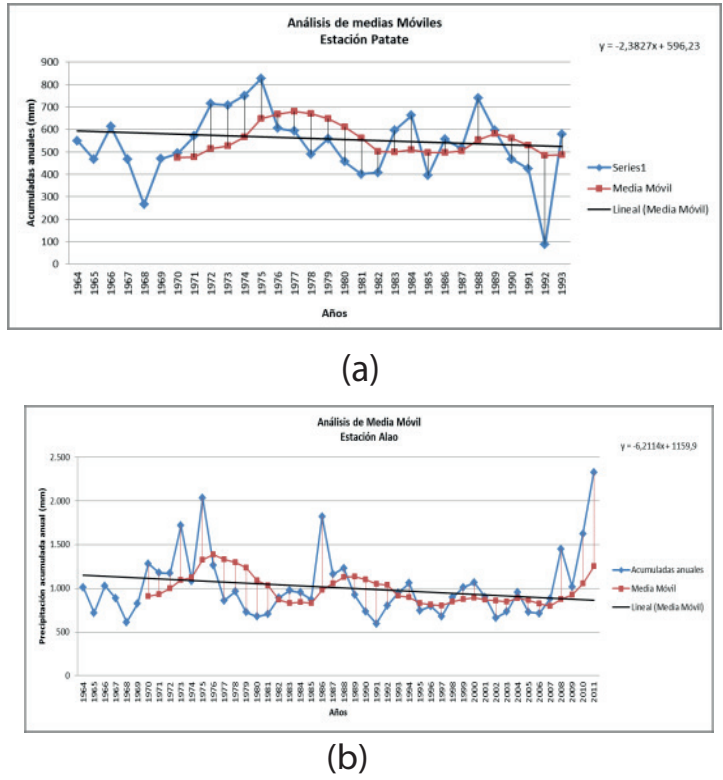


Figura 4. Ciclos estacionales para dos estaciones seleccionadas (a) estación Patate, (b) estación Alao

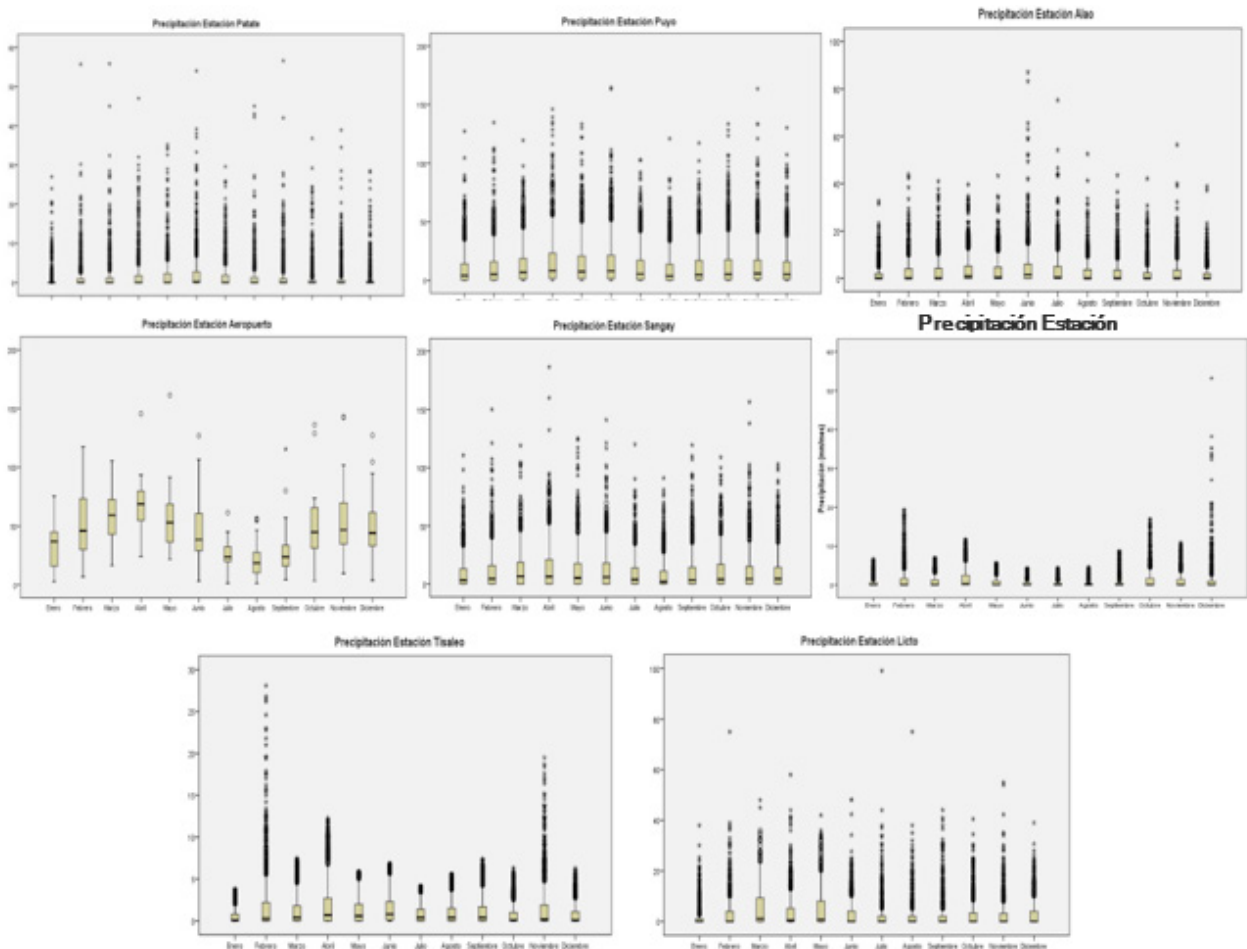
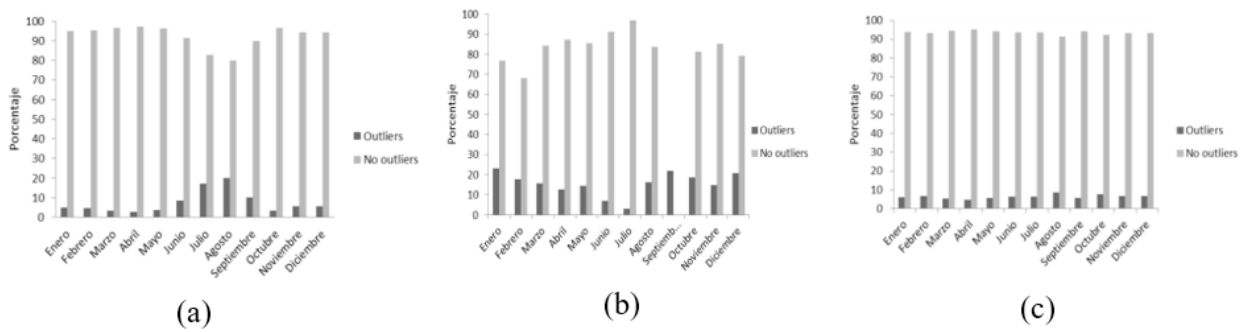


Figura 5. Boxplots mensuales para estacionalidad de ocho estaciones seleccionadas



**Figura 6. Porcentaje de valores atípicos de tres estaciones seleccionadas. (a) estación Cotopaxi-Clirsen; y (b) estación Cusubamba; (c) estación Puyo.**

## Bibliografías

1. Pourrut P. Los climas del Ecuador-Fundamentos Educativos. MAG-ORSTOM, Quito. 1983.
2. González J, De Luis M, Raventós J, Sánchez J, Cortina J. Aproximación metodológica al análisis de la estructura de las tendencias de lluvia. Valencia : Geographicalia, 1999; 37: 53-72.
3. Sneyers R. On the use of statistical analysis for the objective determination of climate change. s.l. Meteorologische Zeitschrift; 1992. p.247-256.
4. Guenni L, Bardossy A. A two steps disaggregation method for highly. Alemania y México: Stochastic Environmental Research and Risk Assessment, 2002, Vol. 16.
5. González J, et.al. 2007. Análisis de tendencias de precipitación (1920-2004) en México. México: Investigaciones Geográficas, Boletín del Instituto de Geografía, UNAM.
6. Puertas O, Carvajal Y, Quintero M. Study of monthly rainfall trends in the upper and middle Cauca river basin. Colombia. DYNA. 2011; 169: 112-120.
7. Carlon T, Mendoza M. Análisis hidrometeorológico de las estaciones de la cuenca del lago de Cuitzeo. Chapala: Boletín del Instituto de Geografía de la UNAM. 2007; 3: 56-76.
8. Gutiérrez J, Rodríguez V, Santos J. Técnicas Cuantitativas (Estadística Básica). Barcelona : Oikos-tau, 1995.
9. OMM. Guía de prácticas climatológicas. 2011.
10. Quevedo H. Métodos estadísticos para la ingeniería ambiental y la ciencia. Juárez : Biblioteca Virtual de la UNAM [Internet] 2006. Disponible en <http://bivir.uacj.mx/LibrosElectronicosLibres/UACJ/ua00001.pdf>
11. Williamson D, Parker R, Kendrick J. The box plot: A simple visual method to interpret data. Annals of internal medicine. 1989; 916-921.
12. Lavado C, Silvestre E, Pulache W. Tendencias en los extremos de lluvia cerca a la ciudad del Cuzco y su relación con las inundaciones de enero del 2010. Revista Peruana Geo-Atmósferica. 2010; 2: 89-98.
13. Hawkins, D.M. Identification of outliers. London : Chapman and Hall, 1980.
14. Seo, S. A review and comparison of methods for detecting outliers in univariate data sets. Pittsburgh: Pittsburgh University, 2002.
15. Ezequiel U, Amado P. Introducción al análisis de series temporales. Madrid : Paraninfo. 2000.