XIV SEMINARIO NACIONAL
DE HIDRAULICA E HIDROLOGIA

DETECCIÓN DE PUNTOS ANORMALMENTE EXTREMOS (OUTLIERS) EN SERIES HIDROLÓGICAS

Ricardo A. Smith y Claudia Campuzano
Posgrado en Aprovechamiento de los Recursos Hidráulicos
Facultad de Minas, Universidad Nacional de Colombia, Medellín
cpcampuz@andromeda.unalmed.edu.co

RESUMEN

Se presenta una revisión de la literatura sobre las pruebas para detección de puntos anormalmente extremos en series hidrológicas. Las pruebas para detección de puntos anormalmente extremos se clasifican en dos grupos: pruebas de distribución y pruebas de cercanía. Se describen algunas de esas pruebas y se presentan resultados de sus aplicaciones a varias series hidrológicas en Colombia. Se presentan algunas conclusiones y recomendaciones.

ABSTRACT

A review of the tests proposed in the literature for detection of outliers in hydrologic series is presented. The tests for detection of outliers are classified in two groups: distribution tests and closeness to next observation. Some of the tests are described with some detail and some results are presented with applications to some Colombian hydrologic series. Some conclusions and recommendations are also presented.

1. INTRODUCCION

Un punto anormalmente extremo u outlier es una observación que queda anormalmente lejos del comportamiento general de las observaciones en una serie. Los outliers puede ser causados por errores de medida, errores de transcripción, averías de los instrumentos, y problemas de calibración. También pueden indicar mayor variabilidad espacial o temporal que la esperada. Los outliers dan la impresión de ser muestras de una distribución distinta al resto de las observaciones, parecen no ser representativos de la muestra. En los procesos de estimación estadística los outliers a menudo llevan a resultados sesgados y escogencia de distribuciones inadecuadas. Se acostumbra entonces en hidrología, antes de los procesos de estimación de las distribuciones, detectar si la muestra disponible tiene outliers, y si así es, proceder a la remoción de los mismos.

Diferentes pruebas han sido propuestas para detectar outliers y se encuentran referenciadas en algunos libros de estadística aplicada a la hidrología (Gilbert, 1987; Gibbons, 1994; Kottegoda y Rosso, 1997; Helsel y Hirsch, 1992) y otros en la literatura sobre el tema (Tietjen y Moore, 1972). Todas esas pruebas han sido desarrolladas básicamente para probar la hipótesis nula de que todas las observaciones fueron tomadas de poblaciones igualmente distribuidas. Si no se rechaza la hipótesis nula, entonces se asume que la muestra no tiene outliers. Las pruebas propuestas en general se pueden clasificar en dos grupos:

- Basados en distribuciones
- Basados en cercanía

Las pruebas basadas en distribuciones prueban si la distribución de la muestra con y sin el punto o los puntos supuestamente outliers es la misma. Si es la misma se acepta la hipótesis de que la muestra no tiene outliers. Las pruebas basadas en cercanías prueban si el punto supuestamente outlier esta lo suficientemente alejado de la siguiente observación más cercana a ese punto. Si no lo esta se acepta la hipótesis de que la muestra no tiene outliers.

A continuación se presentan varias pruebas para detección de outliers. Algunas de estas pruebas son no paramétricas, parea las cuales no hay que hacer suposiciones acerca de la forma de la distribución de la población de la que se tomo la muestra. Otras pruebas fueron desarrolladas bajo la suposición de normalidad. En este caso si se prueba que las observaciones no vienen de una población normalmente distribuida podrían ser transformadas para garantizar el cumplimiento de esta condición.

2. PRUEBAS DE DETECCION DE OUTLIERS BASADAS EN CERCANIAS

Como se anoto anteriormente estas pruebas se basan en comparar el punto que se sospecha es un outlier con la siguiente observación más cercana a ese punto. Si el punto esta muy alejado de la siguiente observación se acepta la hipótesis de que ese punto es un outlier. Algunas de estas pruebas solo pueden detectar un outlier en la muestra y algunas de ellas pueden detectar varios outliers.

2.1 Prueba de Dixon

La prueba de Dixon puede ser usada para probar si existen outliers en un grupo de observaciones. Esta es una prueba propuesta para un grupo de observaciones con un número pequeño de observaciones que se sospecha son outliers. La prueba está basada en una comparación de la observación más alejada con la próxima observación. Se recomienda aplicar esta prueba con cuidado para evitar el enmascaramiento de un outlier por otro. En este sentido se recomienda que primero se identifique el grupo de observaciones que se sospecha pueden ser outliers. Si M representa el número de observaciones que se sospecha pueden ser outliers, entonces la prueba de Dixon podría ser aplicada para todos los M outliers. Una vez se identifica un outlier usando la prueba este es removido del grupo de observaciones y se puede proceder a probar si hay otro outlier. Esta prueba puede usarse tanto para detectar outliers en la dirección de los máximos, como para detectar outliers en la dirección de los mínimos. A continuación se describe esta prueba.

Sea Y_j, $j=1,...,N$ un grupo de N observaciones y W_j, $j= 1,...,N$ son las mismas observaciones pero organizadas en orden creciente de magnitud $(W_1 < W_2 < .... < W_N)$. La prueba estadística de Dixon está definida de acuerdo al rango del tamaño de la muestra N y probando si el valor más alto o más bajo en el grupo de observaciones es un outlier de acuerdo con la siguiente Tabla 1 (Gibbons, 1994, p.254).

Tabla 1. Estadístico para la Prueba de Dixon

N	Highest Value	Lowest Value
3-7	$D = \frac{W_N - W_{N-1}}{W_N - W_1}$	$D = \frac{W_2 - W_1}{W_N - W_1}$
8-10	$D = \frac{W_N - W_{N-1}}{W_N - W_2}$	$D = \frac{W_2 - W_1}{W_{N-1} - W_1}$
11-13	$D = \frac{W_N - W_{N-2}}{W_N - W_2}$	$D = \frac{W_3 - W_1}{W_{N-1} - W_1}$
14-25	$D = \frac{W_N - W_{N-2}}{W_N - W_3}$	$D = \frac{W_3 - W_1}{W_{N-2} - W_1}$

La hipótesis nula de que la observación más alta o más baja es un outlier se rechaza si

$$d \leq C\,(N, \alpha) \qquad (1)$$

en donde d es una estimado de la prueba estadística D dado en la Tabla anterior y $C\,(N, \alpha)$ es el valor crítico de la prueba para una muestra de tamaño N y un nivel de significancia α. Tablas para determinar $C(N, \alpha)$ están disponibles en Gibbons (1994, p .253).

2.2 Pruebas Estadísticas E_M y L_M

Las pruebas estadísticas E_M y L_M pueden usarse para probar la existencia de *M* outliers en un grupo de observaciones que se asume fue tomado de una población normalmente distribuida con media y varianza desconocidas. En la aplicación de estas pruebas se recomienda primero identificar el grupo de *M* observaciones sospechosas de ser outliers, pudiendo aplicarse la prueba para diferentes valores de *M*. La prueba estadística L_M es una prueba de hipótesis de dos lados y se puede usar para identificar outliers por encima o por debajo de la media. La prueba estadística E_M puede usarse para detectar si las *M* observaciones más extremas (por encima o por debajo de la media) en una muestra son outliers. A continuación se describen estas dos pruebas.

Sea Y_j, $j=1,...,N$ un grupo de *N* observaciones. Para calcular el estadístico para la prueba L_M, inicialmente la serie se organiza en orden creciente de magnitud (si se están detectando outliers por encima de la media) o en orden decreciente de magnitud (si se están detectando outliers por debajo de la media). W_j, $j=1,...,N$ representa la serie ordenada. La media de la muestra puede calcularse como

$$\hat{\mu}_Y = \frac{1}{N}\sum_{j=1}^{N} Y_j \tag{2}$$

La media de la serie sin las M observaciones más extremas puede ser calculada como:

Para outliers por encima de la media:

$$\hat{\mu}_Y^{N-Mi)} = \frac{1}{N-M}\sum_{j=1}^{N-M} Y_j \tag{3a}$$

Para outliers por debajo de la media:

$$\hat{\mu}_Y^{N-Mi)*} = \frac{1}{N-M}\sum_{j=M+1}^{N} Y_j \tag{3 b}$$

La estadístico para la prueba L_M está definida como (Tiejten y Moore, 1972):

Para outliers por encima de la media:

$$L_M = \frac{\sum_{j=1}^{N-M}\left(Y_j\ \hat{\mu}_Y^{(N-M)}\right)^2}{\sum_{j=1}^{N}\left(Y_j\ \hat{\mu}_Y\right)^2} \tag{4 a}$$

Para outliers por debajo de la media:

$$L_M = \frac{\sum_{j=M+1}^{N} \left(Y_j \; \hat{\mu}_Y^{(N-M)*}\right)^2}{\sum_{j=1}^{N} \left(Y_j \; \hat{\mu}_Y\right)^2} \qquad (4\ b)$$

La hipótesis nula de que el grupo de *M* observaciones en el grupo de observaciones son outliers se rechaza al nivel de significancia α si (Tietjen y Moore, 1972, p .586):

$$l_M \leq L\,(N, M, \alpha) \qquad (5)$$

en donde l_M es una estimado del estadístico L_M de la ecuación (4), *L(N, M,* α*)* es el valor crítico de la prueba dado en función del tamaño de la muestra *N*, del número *M* de outliers que se esta probando, y nivel de significancia α. Las Tablas para determinar *L(N, M,* α*)* se encuentran en Tiejten y Moore (1972, p .587).

Para determinar el estadístico para la prueba E_M inicialmente se determina la serie de las desviaciones con respecto a la media en valor absoluto dadas como:

$$\tilde{Y}_j = \left| Y_j - \hat{\mu}_Y \right| \quad , \quad j = 1,.......,N \qquad (6)$$

Luego se define la serie W_j dada por la serie $\tilde{Y}_j$ ordenada en orden creciente de magnitud $(W_1 < W_2 < .... < W_N)$. Una nueva serie se define entonces como:

$$Z_i = Y_j \quad \text{Si} \quad W_i = \tilde{Y}_j \qquad (7)$$

Es decir, Z_i es igual a la *Y-ava* observación cuyo $\tilde{Y}_j$ representa el i-ésimo valor más grande.

Usando la serie Z_j se calcula la media de la muestra usando todas las observaciones y usando las primeras *N- M* observaciones:

$$\hat{\mu}_Z = \frac{1}{N} \sum_{j=1}^{N} Z_j \qquad (8)$$

$$\hat{\mu}_Z^{N-Mi)} = \frac{1}{N\,M} \sum_{j=1}^{N-M} Z_j \qquad (9)$$

El estadístico para la prueba E_M se define entonces como (Tietjen y Moore, 1972, p .586):

$$E_M = \frac{\sum_{j=1}^{N-M} \left(Z_j \; \hat{\mu}_Z^{(N-M)}\right)^2}{\sum_{j=1}^{N} \left(Z_j \; \hat{\mu}_Z\right)^2} \qquad (10)$$

La hipótesis nula de que el grupo de *M* observaciones en el grupo de observaciones son outliers se rechaza al nivel de significancia α si (Tietjen y Moore, 1972, p .587):

$$e_M \leq E\ (N, M, \alpha) \qquad (11)$$

en donde e_M es un estimado del estadístico E_M de la ecuación (10) y $E\ (N, M, \alpha)$ es el valor crítico de la prueba en función del tamaño de la muestra *N*, del número *M* de outliers y del nivel de significancia α. Las Tablas para determinar $E\ (N, M, \alpha)$ se encuentran igualmente en Tietjen y Moore (1972, p .591).

3. PRUEBAS DE DETECCION DE OUTLIERS BASADAS EN DISTRIBUCIONES

Como se anoto anteriormente las pruebas basadas en distribuciones prueban si la distribución de la muestra con y sin el punto o los puntos supuestamente outliers es la misma. Si es la misma se acepta la hipótesis de que la muestra no tiene outliers. La distribución utilizada para realizar este tipo de pruebas es la distribución Normal.

3.1 Prueba de la Distribución Normal

Cuando se supone que los datos vienen de una distribución normal y hay outliers presentes, parecería como si los datos se hubieran tomado de otra distribución. Pruebas de normalidad realizadas utilizando grupos de observaciones con outliers en general rechazan la hipótesis de que los datos fueron tomados de una población normalmente distribuida. Una prueba para detectar outliers puede ser entonces probar repetidamente la hipótesis de normalidad cuando las observaciones que se sospechan son outliers se remueven de manera secuencial del grupo de observaciones. Cualquiera de las pruebas de normalidad existentes en la literatura se puede usar con este propósito.

Sea Y_j, $j = 1., N$ un grupo de *N* observaciones. Se puede entonces realizar una prueba de normalidad utilizando esta muestra y si no se rechaza la hipótesis de normalidad, entonces se rechaza la hipótesis nula de outliers en el grupo de observaciones. Si se rechaza la hipótesis de normalidad, entonces se acepta la hipótesis de que hay outliers en el grupo de observaciones. Las pruebas de normalidad se aplican comenzando con $i=1$. Si no se rechaza la hipótesis de normalidad sobre las $N-1$ observaciones restantes, entonces hay solo un outlier en el grupo de observaciones. Si se rechaza, hay más de un outlier en el grupo de observaciones y en este caso las pruebas de normalidad se aplican considerando $i=2$. Si no se rechaza la hipótesis de normalidad sobre las N-2 observaciones restantes, entonces hay dos outliers en el grupo de datos. Este procedimiento continúa hasta que no se rechaza la hipótesis de normalidad. Por lo tanto, si no se rechaza la prueba de normalidad para $i = k$ significa que hay *k* outliers en el grupo de observaciones.

3.2 Prueba Studentized Deviates

La prueba Studentized Deviates es una prueba de detección de multioutlier. Se asume que el grupo de observaciones se tomó de una población normalmente distribuida. La hipótesis nula es que la distribución de la población de donde se tomaron las observaciones que se sospecha son outliers es la misma distribución normal de la población de donde se obtuvo el resto de las observaciones de la muestra. La prueba requiere que se especifique, antes de realizar la misma, un número superior M de outliers potenciales en el grupo de observaciones. A continuación se describe esta prueba.

Sea Y_j, $j = 1,\ldots, N$ un grupo de N observaciones y se sospecha que hay M outliers dentro del grupo de datos. Primero el grupo de observaciones se organiza en orden creciente de magnitud y se define una nueva serie W_j tal que $W_1 < W_2 < \ldots < W_N$. $\hat{\mu}_W^{(N-i+1)}$ y $\hat{\sigma}_W^{(N-i+1)}$ representan la media y la desviación estándar del grupo W estimadas usando las primeras $N-i+1$ observaciones, y dadas como:

$$\hat{\mu}_W^{(N-i+1)} = \frac{1}{N-i+1}\sum_{j=1}^{N-i+1} W_j \quad , \qquad i = 1, \ldots\ldots\ldots, m \qquad (12)$$

$$\hat{\sigma}_W^{(N-i+1)} = \sqrt{\frac{1}{N-i+1}\sum_{j=1}^{N-i+1}\left(W_j - \hat{\mu}_W^{(N-i+1)}\right)^2} \quad , \qquad i = 1, \ldots\ldots\ldots, m \qquad (13)$$

El estadístico para la prueba se define como (Kottegoda y Rosso, 1997, p .322)

$$S_j = \frac{W_{N-i+}\,\hat{\mu}_W^{(N-i+1)}}{\hat{\sigma}_W^{(N-i+1)}} \qquad (14)$$

La hipótesis nula de que hay i outliers en el grupo de M observaciones que se sospechan que son outliers del grupo de observaciones se rechaza sí:

$$|s_j| \leq B_{M,i}(N, \alpha) \qquad (15)$$

en donde s_j es un estimado de la prueba estadística S_j de la ecuación (14) y $B_{M,i}$ (N, α) es el valor crítico para la prueba del i-ésimo outlier en el grupo de M observaciones que se sospechan que son outliers, dado en función del tamaño de la muestra N, y del nivel de significancia α para la prueba. Tablas para encontrar $B_{M,i}$ (N, α) se encuentran en Kottegoda y Rosso (1997, p .699). La prueba Studentized Deviates descrita es para outliers por encima de la media. La prueba también puede ser usada para identificar outliers por debajo de la media. En este caso $\hat{\mu}_W^{(N-i+1)}$ y $\hat{\sigma}_W^{(N-i+1)}$ de las ecuaciones (12) y (13) se calculan usando los $N-i+1$ valores de la serie W_j. La prueba estadística S_j de la ecuación (14) se define usando W_i en lugar de W_{N-i+1}, y la prueba es desarrollada usando la ecuación (15) en la misma manera como se expone arriba.

3.3 La Prueba de Rosner

Esta es una prueba de las llamadas mutioutliers ya que permite detectar hasta diez outliers en un grupo de datos. La prueba es válida para tamaños de muestra de 25 observaciones o mayores. El procedimiento asume que las observaciones se tomaron de una población normalmente distribuida. La prueba trata de evitar el enmascaramiento de un outlier por otro cuando están relativamente cercanos. Algunas pruebas de detección de outliers como la de Dixons se basan en comparar la mayor observación con la próxima, así si dos outliers tienen valores cercanos esta prueba no detectará ningún outlier en el grupo de observaciones.

Antes de realizar la prueba se necesita especificar un número M de outliers potenciales presentes en el grupo de observaciones. La hipótesis nula es que todo el grupo de observaciones representa una muestra tomada de una distribución normal y la hipótesis alternativa es que hay M outliers, o M - *1* outliers, o....., o *1* outlier (Gilbert, 1987, p .188). La prueba de Rosner es una prueba de hipótesis de dos lados e identifica outliers por encima o por debajo de la media.

Sea Y_j, $j = 1....,N$ un grupo de observaciones donde N es el número total de observaciones. Se sospecha que hay M outliers en el grupo. Sean $\hat{\mu}^{(i)}$ y $\hat{\sigma}^{(i)}$ la media estimada y la desviación estándar de las $N - i$ observaciones del grupo de observaciones después de que las i observaciones más extremas (en la dirección que interesa hacer la prueba) han sido removidas (Gilbert, 1987, p .188; Gibbon, 1994, p .247).

$$\hat{\mu}^{(i)} = \frac{1}{N-i}\sum_{j=1}^{N-i} Y_j \quad , \qquad i = 1, ..., M \qquad (16)$$

$$\hat{\sigma}^{(i)} = \sqrt{\frac{1}{N-i}\sum_{j=1}^{N-i}\left(Y_j - \hat{\mu}^{(i)}\right)^2} \quad , \qquad i = 1, ..., M \qquad (17)$$

$Y^{(i)}$ representa la observación más extrema (la más alejada de la media $\hat{\mu}^{(i)}$) en el grupo de observaciones después de que las *i* observación más extrema han sido removidas de ese grupo. El estadístico para la prueba Rosner se define entonces como:

$$R_{i+1} = \frac{Y^{(i)} - \hat{\mu}^{(i)}}{\hat{\sigma}^{(i)}} \qquad (18)$$

La hipótesis nula de que hay i +1 outliers en el grupo de observaciones se rechaza sí:

$$|r_{i+1}| \leq \lambda_{i+1}(N, \alpha) \qquad (19)$$

en donde r_{i+1} es un estimado de la prueba estadística R_{i+1} de la ecuación (18) y $\lambda_{i+1}(N, \alpha)$ es el valor crítico para la prueba en función del número de outliers $i+1$, el tamaño de la muestra N, y el nivel de significancia α. Las Tablas para determinar $\lambda_{i+1}(N, \alpha)$ están disponibles en Gilbert (1987, p .268) o Gibbon (1994, p .248).

4. CASOS DE APLICACIÓN

En la Tabla 2 se muestran los ríos a los cuales se les aplicaron las diferentes pruebas presentadas y que se sospecha tienen observaciones anormalmente extremas.

Tabla 2. Ríos utilizados para pruebas de Outliers

Nombre	Periodo	Tamaño Muestra	Tipo de serie
Río Chaquenoda	1980 – 1991	12	Máximos Instantáneos Anuales
Quebrada La Mosca	1957 - 1977	21	Máximos Instantáneos Anuales
Río Ovejas	1982 - 1992	11	Máximos Instantáneos Anuales
Río San Carlos (Balseadero)	1964 - 1981	18	Máximos Instantáneos Anuales
Río Buey	1955 - 1980	26	Máximos Instantáneos Anuales
Río Grande	1965 -1992	28	Máximos Instantáneos Anuales
Río Riachón (Suribios)	1982 – 1992	11	Máximos Instantáneos Anuales
Río Tapia	1971 – 1994	23	Máximos Instantáneos Anuales
Río Supia	1976 – 1994	19	Máximos Instantáneos Anuales
Río Tigui	1973 – 1991	17	Mínimos Medios Diarios Anuales
Quebrada La Marinilla	1979 – 1990	12	Mínimos Medios Diarios Anuales

A manera de ejemplo se presentan a continuación las gráficas de algunas de las series de tiempo descritas en la tabla anterior. En la Figura 1 se presenta el río Chaquenoda, en la Figura 2 la quebrada La Mosca, en la Figura 3 el río Grande y en la Figura 4 el río Tapia.

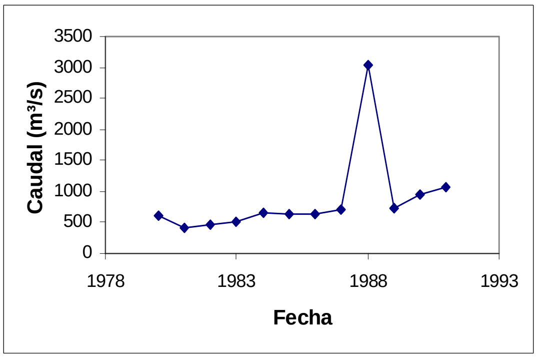

Figura 1. Río Chaquenoda

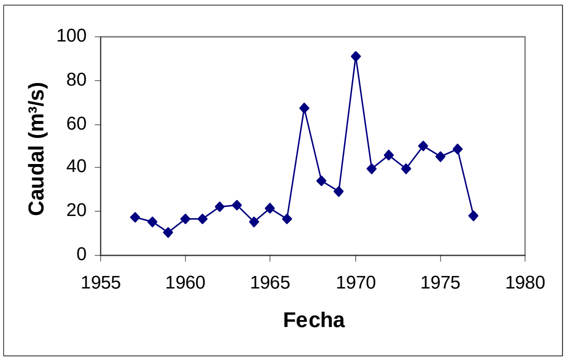

Figura 2. Quebrada La Mosca

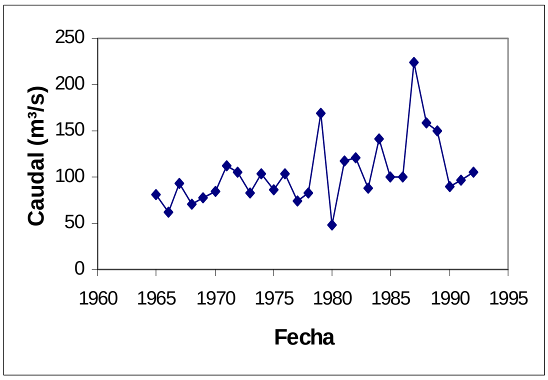

Figura 3. Río Grande

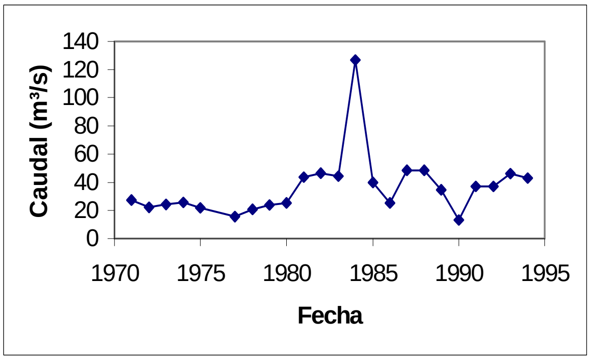

Figura 4. Río Tapias.

Los resultados obtenidos con las pruebas se muestran en la Tabla 3. En esta tabla NA significa No Aplicable, es decir, la prueba no se pudo usar al no cumplirse con algunos de los requisitos para su aplicación.

Tabla 3. Resultados de las Pruebas de Outliers.

Nombre Río	Número sospech.	Número Detectado de Outliers					
		Student.	Roesner	Normalidad	Dixon	Em	Lm
Chaquenoda	1	NA	NA	1	1	1	1
La Mosca	2	0	NA	1	1	0	0
Ovejas	1	NA	NA	NA	NA	2	0
San Carlos	1	NA	NA	1	1	1	1
Buey	1	1	1	1	1	1	1
Grande	2	1	1	1	1	1	1
Riachón	1	NA	NA	2	2	2	2
Tapia	1	1	NA	1	1	1	1
Supia	2	NA	NA	2	2	2	2
Tigui	1	NA	NA	2	1	2	2
La Marinilla	1	NA	NA	1	1	1	1

Como se puede ver de los resultados hay pocos casos en que se pudo usar la prueba de Roesner. Algo similar pero en menor dimensión ocurre con la prueba Studentized. Las pruebas en general son consistentes en los resultados existiendo una tendencia a dar los mismos resultados en todas ellas. La prueba que consistentemente dio los mejores resultados fue la prueba de Dixon. A continuación se presentan algunas conclusiones sobre la aplicación de estas pruebas a las series anteriores.

5. CONCLUSIONES Y RECOMENDACIONES

De acuerdo con las características de las pruebas presentadas y con los resultados obtenidos se pueden hacer las siguientes observaciones y conclusiones:

- Existen varias pruebas para detección de outliers con diferentes exigencias de información para su aplicación, y diferentes suposiciones sobre la distribución de donde se asume se tomo la muestra (parmetricas y no parametricas).

- En el caso de Colombia las series son de relativa poca longitud y por lo tanto esta situación tiende a favorecer las pruebas que son poco exigentes en el tamaño de la muestra, que generalmente son las pruebas de menor potencia estadística.

- En los casos utilizados generalmente las pruebas identificaron un número de outliers igual al número sospechado, dando en general un resultado aceptable y consistente entre todas ellas. Debido a las diferencias en las exigencias sobre el tamaño de la muestra no fue posible comparar empíricamente de una manera más adecuada las diferentes pruebas.

- En los casos presentados la prueba de Dixon fue la prueba que mejores resultados arrojo. Esta prueba coincidió con el número sospechado de outliers en todos los casos menos uno.

- La prueba de normalidad dio los mismos resultados cuando se usaron diferentes pruebas de normalidad. Aunque esta prueba es una prueba parametrica (asume normalidad de la población), tiene la gran ventaja que es aplicable para cualquier tamaño de la muestra. Los resultados obtenidos con esta prueba fueron bastante aceptables.

- Las pruebas de Roesner y Studentized solo se pudieron usar en algunos casos debido a las limitaciones en los tamaños de las muestras utilizados.

- Las pruebas de Em y Lm tienen una tendencia general a dar los mismos resultados.

Como recomendación se propone el uso extensivo de las pruebas de detección de puntos anormales, de tal manera que se puedan identificar y ser removidos de los análisis posteriores que se hagan con las series. La no remoción de estos puntos significa trabajar con muestras que vienen de poblaciones distintas, lo cual puede tener consecuencias no apropiadas sobre los análisis que se hagan.

6. REFERENCIAS

Gibbon, R.D., 1994. Statistical Methods for Groundwater Monitoring. John Wiley and Sons Inc., New York.

Gilbert, R.O., 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold Company, New York.

Helsel, D.R. and Hirsch, R.M., 1992. Statistical Methods in Water Resources. Elsevier, Amsterdam, p. 522.

Kottegoda, N.T. and Rosso, R., 1997. Probability, Statistics and Reliability for Civil and Environmental Engineers. McGraw Hill Book Co., New York.

Tietjen, G.L and Moore, R.H, 1972. Some Grubbs-Type statistics for the detection of several outliers. Technometrics, 14(3): 583-597.