APLICABILIDAD DEL MODELO DE TRANSFORMADA RÁPIDA DE FOURIER EN SERIES DE CAUDALES DEL RÍO CAUCA EN LA CUENCA ANDINA

CESAR AUGUSTO GARCÉS OROZCO
JORGE IVÁN CARDONA LOAIZA

UNIVERSIDAD EAFIT
ESCUELA DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA CIVIL
ÁREA DE PROYECTOS
MEDELLÍN
2013

APLICABILIDAD DEL MODELO DE TRANSFORMADA RÁPIDA DE FOURIER EN SERIES DE CAUDALES DEL RÍO CAUCA EN LA CUENCA ANDINA

CESAR AUGUSTO GARCÉS OROZCO

JORGE IVÁN CARDONA LOAIZA

Proyecto de grado para optar al título de Ingeniero Civil

Asesor: Heber A. Escobar P.
Ingeniero Civil

UNIVERSIDAD EAFIT

ESCUELA DE INGENIERÍA

DEPARTAMENTO DE INGENIERÍA CIVIL

ÁREA DE PROYECTOS

MEDELLÍN

2013

Nota de aceptación:

_____

_____

_____

_____

Presidente del Jurado

_____

Jurado

_____

Jurado

Medellín, Abril de 2013

AGRADECIMIENTO

Al finalizar este último gran paso de nuestro sueño de convertirnos en ingenieros, es inevitable e indispensable agradecerles a las personas que de cierta manera hicieron posible esto. Por tal motivo les agradecemos inmensamente a nuestras familias ya que sin ellos este sueño sería solo eso, un sueño; nuestros padres han sido el motor principal de este camino y los que nos han estado con nosotros incondicionalmente frente a las dificultades que se nos presentaron a lo largo de la carrera.

De igual modo también damos las sinceras gracias a los hermanos, amigos, compañeros y profesores, que ayudaron de manera directa o indirecta en este proyecto, ellos son testigos del arduo trabajo que se realizo durante este.

Es importante agradecerle a INTEGRAL S.A. por brindarnos la información base que permitió realizar este estudio.

Finalmente y no siendo menos importante, un infinito agradecimiento a Heber A. Escobar quien se desempeño como asesor de este proyecto, convirtiéndose en una guía durante este camino, muchas gracias.

"Los hombres construimos demasiados muros y no suficientes puentes."

Isaac Newton

TABLA DE CONTENIDO

1	INTRODUCCIÓN	10
2	OBJETIVOS	12
2.1	OBJETIVO GENERAL	12
2.2	OBJETIVOS ESPECÍFICOS	12
3	MARCO TEÓRICO	13
3.1	ANÁLISIS DE OUTLIERS EN SERIES DE TIEMPO	13
3.1.1	Basados en distribuciones	14
3.1.2	Basados en cercanía	17
3.2	ANÁLISIS DE LA MEDIA EN SERIES DE TIEMPO	19
3.2.1	Media móvil	19
3.2.2	Media fija	20
3.2.3	Prueba de las carreras	20
3.2.4	Wilcoxon suma de rangos	22
3.3	TRANSFORMADA RÁPIDA DE FOURIER APLICADA A SERIES DE TIEMPO	23
4	DESCRIPCIÓN DE LA ZONA DE ESTUDIO	25
5	INFORMACIÓN BASE	27
6	ANÁLISIS Y SELECCIÓN INICIAL DE ESTACIONES	28
7	BÚSQUEDA Y ELIMINACIÓN DE OUTLIERS EN LAS ESTACIONES SELECCIONADAS	30
8	ANÁLISIS DE LA MEDIA EN LAS ESTACIONES SELECCIONADAS	34
8.1	MEDIA MÓVIL	34
8.2	MEDIA FIJA	40
8.3	PRUEBA DE LAS CARRERAS	43
8.4	WILCONXON SUMA DE RANGOS	44
9	FRECUENCIA CARACTERÍSTICAS EN LAS ESTACIONES SELECCIONADAS	46
10	FENÓMENOS MACRO CLIMÁTICOS DE LA ZONA DE ESTUDIO	54
10.1.1	Índice Multivariado del ENSO (MEI)	54
10.1.2	Índice Oceánico del Niño (ONI)	55
10.1.3	Oscilación de Madden-Julian (MJO)	56

11 VALIDACIÓN DEL MODELO DE LA FFT EN LA ZONA DE ESTUDIO . 58
12 METODOLOGÍA PARA COMPLETAR SERIES DE CAUDALES 64
13 METODOLOGÍA PARA TRASLADAR SERIES DE CAUDALES 70
14 CONCLUSIONES 73
BIBLIOGRAFÍA 75

LISTADO DE FIGURAS

Figura 4—1 Cuenca drenaje del rio Cauca en la zona de estudio. 26
Figura 7—1 Serie ejemplo con outlier identificado 31
Figura 7—2 Serie ejemplo con outlier suavizado 31
Figura 7—3 Cantidad de Outliers eliminados en cada estación 33
Figura 8—1 Medias anuales de las estaciones analizadas 35
Figura 8—2 Caudal medio anual de las estaciones analizadas 36
Figura 8—3. Media móvil cada 2 años 38
Figura 8—4. Media móvil cada 3 años 39
Figura 8—5. Media fija cada 2 años 41
Figura 8—6. Media fija cada 3 años 42
Figura 8—7 Resultados Wilcoxon suma de rangos 45
Figura 9—1 Espectro de Fourier estación Pintada 47
Figura 9—2 Espectro de Fourier estación Pescadero 48
Figura 9—3 Espectro de Fourier estación Valdivia 48
Figura 9—4 Espectro de Fourier estación Apaví 49
Figura 9—5 Espectro de Fourier estación La Coquera 50
Figura 9—6 Espectro de Fourier estación Margento 50
Figura 9—7 Espectro de Fourier estación Las Flores 51
Figura 9—8 Espectros de Fourier de las estaciones analizadas 53
Figura 10—1 Índice MEI (NOAA) 55
Figura 10—2 Índice ONI (NOAA) 56
Figura 10—3 Índice MJO (NOAA) 57
Figura 11—1 Frecuencias características del MEI con la Pintada 1 58
Figura 11—2 Frecuencias características del ONI con la Pintada 1 59
Figura 11—3 Frecuencias características del MJO con la Pintada 1 59
Figura 11—4 Frecuencia característica entre 1,9 y 2,3 años identificada en las estaciones 60
Figura 11—5 Frecuencia característica entre 1,9 y 2,3 años identificada en los índices climáticos 61
Figura 11—6 Frecuencia característica entre 5 y 6 años identificada en las estaciones 62

Figura 11—7 Frecuencia característica entre 5 y 6 años identificada en los índices climáticos 62
Figura 11—8 Frecuencia característica de 0,5 años identificada en las estaciones 63
Figura 12—1 Serie original de caudales medios diarios estación La Coquera 64
Figura 12—2 Año pivote de 1986 65
Figura 12—3 Normalización de serie y función de la estación La Coquera 66
Figura 12—4 Serie normalizada rellenada 66
Figura 12—5 Serie de caudales rellenada con armónicos 67
Figura 12—6 Serie definitiva de caudales medios diarios 68
Figura 12—7 Validación metodología para complementar información **¡Error! Marcador no definido.**
Figura 13—1 Función armónica normalizada 70
Figura 13—2 Validación metodología traslado de información 72

LISTADO DE TABLAS

Tabla 3-1 Estadístico para la Prueba de Dixon 14
Tabla 5-1 Estaciones analizadas 27
Tabla 6-1 Años de registro de las estaciones analizadas 28
Tabla 6-2 Numero de datos reales registrados y su equivalente en años de las estaciones analizadas 29
Tabla 7-1 Outliers identificados por estación 32
Tabla 7-2 Cantidad de Outliers eliminados en cada estación 32
Tabla 8-1 Estadísticos Prueba de las Carreras 43
Tabla 9-1 Periodos continuos con mayor longitud de registro para las estaciones de la zona de estudio 46
Tabla 9-2 Frecuencias características de las estaciones analizadas 52
Tabla 12-1 Errores en m^3/s de la metodología para completar caudales 69
Tabla 13-1 Desviación estándar y media de las estaciones de estudio 71

1 INTRODUCCIÓN

Actualmente la ejecución de proyectos de infraestructura de transporte, social e industrial, aunado a los episodios de inviernos intensos, han hecho que los análisis hidrológicos tomen gran importancia, esto debido al riesgo al que se ven expuestos, ya sea por la superficialidad de los análisis que alteran el diseño de obras hidráulicas para la protección de inundación, el control de erosión de las riveras, definición de gálibos de obras de cruce sobre corrientes hidráulicas, o incluso en el peor escenario, la no previsión de las mismas.

Es indispensable destacar que en los estudios hidrológicos, se debe dar suma importancia a la implicación de los fenómenos macroclimáticos en el régimen de precipitaciones y caudales, con el fin de realizar de manera adecuada el dimensionamiento de las obras civiles o desarrollar planes de emergencia.

Debido a que desastres naturales como inundaciones, deslizamientos, avalanchas, vendavales, y tormentas eléctricas han causando pérdida de vidas y daños en viviendas, zonas agrícolas y en la infraestructura pública o privada; de otro lado, también se han presentado intensas sequías las cuales causan pérdida de cultivos, obligan a la población a someterse a corte o racionamiento de agua potable e incluso de la energía eléctrica, la cual en nuestro país es en su mayoría generada por centrales hidroeléctricas.

El problema es que para poder efectuar un análisis hidrológico en una zona determinada, se debe contar con información de las variables hidroclimáticas de una región. Sin embargo, es claro que Colombia cuenta con una red de instrumentación hidrometeorológica deficiente, la cual no alcanza a cubrir de manera adecuada todo su territorio, en consecuencia no existe información de estas variables en algunos lugares, y sí existe, esta información presenta una problemática adicional, debido a que la estaciones son víctimas de actos vandálicos, los equipos no se encuentran calibrados, o tienen vacíos de información a causa de mantenimientos no programados, por lo cual se hace necesario recurrir a metodologías matemáticas y/o estadísticas para trasladar información de variables hidroclimáticas a zonas donde por algún motivo no se cuenta con la información apropiada para determinar las variables hidrológicas pertinentes.

Adicionalmente, resulta claro que se requiere una metodología práctica, la cual funcione con herramientas informáticas de fácil acceso, de dominio público y que permita la aplicación de procedimientos que consigan tener claridad en la estimación de variables hidrológicas, como niveles, caudales, y precipitación, insumos claves para los dimensionamientos hidráulicos, para así diseñar y construir obras civiles que puedan proteger a una población de los efectos del clima, o simplemente para prever las obras necesarias que puedan suplir las necesidades básicas o comodidad de la población.

En el medio, existe una gran variedad de metodologías que logran representar el comportamiento generado por los fenómenos macroclimáticos en los caudales de un río, pero estas no son aplicadas por el desconocimiento y por la creencia que estos

estimativos consumen demasiado tiempo, o inversiones altas de dinero en licencias de software especializados.

Estas metodologías son muy características de determinada zona en particular, en este caso, se propone analizar estadísticamente varias estaciones de registro y buscar la aplicabilidad del modelo matemático de la transformada rápida de Fourier (FFT, por sus siglas en ingles Fast Fourier Transform) a series de caudales del río Cauca en la cuenca andina, para así proveer un modelo en una plataforma asequible y amigable para el usuario, como lo es Microsoft Excel. Este modelo deberá tener en cuenta la inferencia de fenómenos macroclimáticos, y trasladar la información de manera confiable a zonas de la misma cuenca donde no hayan registros propios de caudales, y de esta manera contar con los datos apropiados para un estudio hidrológico conveniente en el diseño y la construcción de obras civiles, que reduzcan la vulnerabilidad que tienen las poblaciones frente a estos fenómenos.

Adicionalmente, en el análisis hidrológico de series de caudales para el desarrollo de los estudios de diseño de los principales proyectos de infraestructura del país, existen diversas metodologías para detectar, corregir los errores de mediciones y determinar posibles cambios en la tendencia de los datos

Estos análisis se hacen con el fin de tener herramientas que permitan estimar las condiciones hidrológicas características en las zonas de los proyectos.

Entre las metodologías más comunes, se destacan la búsqueda e eliminación de datos anormalmente extremos (Outliers), y el análisis de cambios en la media en series de tiempo, los cuales hacen parte primordial del modelo planteado en el presente proyecto de grado.

Durante la investigación de los temas relacionados con la búsqueda de Outliers, se encontraron algunos artículos provenientes del postgrado “En aprovechamiento de los recursos hidráulicos” de la facultad de minas en la Universidad Nacional de Colombia sede Medellín, en los que se destaca los estudios realizados por Ricardo A. Smith y Claudia P. Campuzano.

Por otra parte, con respecto al análisis de la media en series cronológicas, se encontraron antecedentes donde estas metodologías son utilizadas en los estudios de proyectos hidroeléctricos, dado la importancia que tiene la estimación del caudal para determinar la productividad de este tipo de proyectos; entre los antecedentes que encontraron se destaca los estudios realizados por INTEGRAL S.A. para verificar la factibilidad de este tipo de obras.

2 OBJETIVOS

2.1 OBJETIVO GENERAL

Analizar la aplicabilidad del modelo matemático de la transformada rápida de Fourier (FFT), en series de caudales del río Cauca en la cuenca andina, para asociar estos comportamientos encontrados a una posible intervención de fenómenos macroclimáticos.

Adicionalmente, si este modelo matemático representa de manera confiable esta intervención, se crearan metodologías que ayuden a completar y generar series de caudales en un lugar que carezca de esta información, y de esta manera garantizar que se tenga lo necesario para poder realizar un estudio hidrológico e hidráulico adecuado, en cualquier lugar ubicado dentro de la cuenca andina del río Cauca.

2.2 OBJETIVOS ESPECÍFICOS

Los objetivos específicos del proyecto de grado que busca aplicabilidad del modelo de Transformada Rápida de Fourier a series, se presentan a continuación:

- Seleccionar los registros de las estaciones a las que se tiene acceso, que cumplan con las condiciones adecuadas para realizar los análisis pertinentes.
- Verificar la confiabilidad de series de caudales mediante el análisis de valores extremadamente anómalos (Outliers).
- Estudiar el comportamiento de series de caudales mediante el análisis de variación de la media, para así identificar si existe un cambio en la tendencia de los caudales.
- Estimar frecuencias características mediante la transformada rápida de Fourier (FFT) de las series de caudales del río Cauca en la cuenca andina.
- Determinar que fenómenos macroclimáticos tienen influencia en la zona de estudio y buscar los índices que lo representen según la National Oceanic and Atmospheric Administration (NOAA).
- Encontrar una posible asociación de las frecuencias características de las series de caudales, con los índices que representan los fenómenos macroclimáticos que tienen intervención en la zona.
- Proponer una metodología que se base en el modelo de la FFT para completar datos de caudales en estaciones que en un lapso de tiempo determinado no se tenga registro.
- Proponer una metodología que se base en el modelo de la FFT para trasladar datos de caudales a lugares donde no se cuenten con estos registros.
- Desarrollar un programa en Microsoft Excel con su respectivo manual de usuario que tenga la capacidad de realizar análisis de la media, buscar y eliminar valores extremadamente anómalos y estimar las frecuencias características.

3 MARCO TEÓRICO

Basados en la literatura técnica que existe en el medio, se extrajo la metodología y los procedimientos en los cuales se apoya el presente proyecto de grado. Inicialmente se presenta la metodología utilizada en la detección de outliers, posteriormente se explica los métodos de análisis de la media aplicados y finalmente se explica resumidamente los conceptos básicos que componen la transformada rápida de Fourier, con el fin de suministrar las bases necesarias para entender y poder replicar el análisis propuesto en el presente proyecto de grado.

3.1 ANÁLISIS DE OUTLIERS EN SERIES DE TIEMPO

Un Outlier es una observación que queda anormalmente lejos del comportamiento general de las observaciones en una serie. Los outliers pueden 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 (Smith y Campuzano, 2001). 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 está 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 tomó 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 (Smith y Campuzano, 2001).

3.1.1 Basados en distribuciones

Como se anotó 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 está 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 (Smith y Campuzano, 2001).

3.1.1.1 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 (Smith y Campuzano, 2001).

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 Yj, j=1,...,N un grupo de N observaciones y Wj, j= 1,...,N son las mismas observaciones pero organizadas en orden creciente de magnitud (W1<W2<....<WN). 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 Tabla 3-1(Smith y Campuzano, 2001)

Tabla 3-1 Estadístico para la Prueba de Dixon

N	Valores por encima	Valores por debajo
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	$D = \frac{W_N - W_{N-2}}{W_N - W_3}$	$D = \frac{W_3 - W_1}{W_{N-2} - W_1}$

Nota: Gibbons, 1994

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

$$d < C(N, \alpha) \quad (1)$$

En donde d es estimado de la prueba estadística D dado en la Tabla 3-1y C (N, α) es el valor crítico de la prueba para una muestra de tamaño N y un nivel de significancia α (Smith y Campuzano, 2001). Las Tablas para determinar C(N, α) están disponibles en el anexo 3.

3.1.1.2 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:

$$\mu_y = \frac{1}{N}\sum_{j=1}^{N} Y_j$$

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

$$\mu_y^{(N-M)} = \frac{1}{N-M}\sum_{j=1}^{N-M} Y_j$$

Para outliers por debajo de la media:

$$\mu_y^{(N-M)*} = \frac{1}{N-M}\sum_{j=m+1}^{N} Y_j$$

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

Para outliers por encima de la media:

$$L_M = \frac{\sum_{j=1}^{N-M}\left(Y_j \mu_y^{(N-M)}\right)^2}{\sum_{j=1}^{N}\left(Y_j \mu_y\right)^2}$$

Para outliers por debajo de la media:

$$L_M = \frac{\sum_{j=M+1}^{N-M} (Y_j \mu_y^{(N-M)^*})^2}{\sum_{j=1}^{N} (Y_j \mu_y)^2}$$

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):

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

En donde l_M es una estimado del estadístico L_M de la ecuación anterior, $L(N, M, \alpha)$ 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 (Smith y Campuzano, 2001). Las Tablas para determinar $L(N, M, \alpha)$ se encuentran en el Anexo 3

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:

$$\overline{Y}_j = \left| Y_j - \mu_Y \right| \qquad j=1,.....,N$$

Luego se define la serie W_j dada por la serie ~ Y_j ordenada en orden creciente de magnitud $(W_1 < W_2 < .... < W_N)$. Una nueva serie se define entonces como:

$$Z_i = Y_j \quad Si \quad W_i = \overline{Y}_j$$

Es decir, $Z_i$es igual a la *Y-ava* observación cuyo $\overline{Y}_j$ 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:

$$\hat{\mu}_z = \frac{1}{N} \sum_{j=1}^{N} Z_j$$

$$\hat{\mu}_z = \frac{1}{N-M} \sum_{j=1}^{N-M} Z_j$$

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

$$E_M = \frac{\sum_{j=1}^{N-M} (Z_j \hat{\mu}_z^{(N-M)})^2}{\sum_{j=1}^{N} (Z_j \hat{\mu}_z)^2}$$

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):

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

En donde e_M es un estimado del estadístico E_M de la ecuación anterior 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,* α*)* se encuentran igualmente en el Anexo 3

3.1.2 Basados en cercanía

Como se anotó 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 (Smith y Campuzano, 2001).

3.1.2.1 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 (Smith y Campuzano, 2001).

Sea Yj, 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 (Smith y Campuzano, 2001).

3.1.2.2 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 (Smith y Campuzano, 2001).

Sea Yj, j= 1,...., 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 Wj tal que $W1<W2<....<W_N$.

$\hat{\mu}_w^{(N-i+1)}$ 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:

$$\hat{\mu}_W^{(N-i+1)} = \frac{1}{N-i+1}\sum_{j=1}^{N-i+1} W_j \quad \textbf{\textit{i= 1,……m}}$$

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

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

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

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:

$$\left|S_j\right| \leq B_{M,i}(N, \alpha)$$

En donde Sj es un estimado de la prueba estadística y $B_{M,i}$ (N, α) es el valor crítico para la prueba del i-ésimooutlier en el grupo de M observaciones que se sospechan que son outliers, dado en función del tamaño de la muestra N(Smith y Campuzano, 2001), las tablas para encontrar $B_{M,i}$ (N, α) y del nivel de significancia α para la prueba se encuentran en el Anexo 3.

3.1.2.3 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). La prueba de Rosner es una prueba de hipótesis de dos lados e identifica outliers por encima o por debajo de la media.

Sea Yj, 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 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; Gibbon, 1994).

$$\hat{\mu}^{(i)} = \frac{1}{N-i}\sum_{j=1}^{N-i} Y_j \quad \textbf{\textit{i= 1,……m}}$$

$$\hat{\sigma}^{(i)} = \sqrt{\frac{1}{N-i}\sum_{j=1}^{N-i}\left(Y_j - \hat{\mu}^{(i)}\right)^2} \quad \textbf{\textit{i= 1,……m}}$$

Y(i) representa la observación más extrema (la más alejada de la media 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:

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

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

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

en donde r i +1 es un estimado de la prueba estadística R i +1 de la ecuación anterior y i +1 (N, α) es el valor crítico para la prueba en función del número de outliersi+1, el tamaño de la muestra N, y el nivel de significancia α. Las tablas para determinar (N, α) están disponibles en el anexo 3

3.2 ANÁLISIS DE LA MEDIA EN SERIES DE TIEMPO

A continuación se presentan las diferentes metodologías que se utilizaron para analizar cambios en la tendencia de la media y dependencia del tiempo de las series.

3.2.1 Media móvil

En estadística es un método utilizado para analizar un conjunto de datos en modo de puntos para crear series de promedios. Así las medias móviles son una lista de números en la cual cada uno es el promedio de un subconjunto de los datos originales.

Por ejemplo, si se tiene un conjunto de 100 datos el primer valor de la serie de medias móviles podría ser el promedio de los primeros 25 términos, luego el promedio de los términos 2 al 26, el tercer elemento de los términos 3 al 27 y así, hasta por último el promedio de los últimos 25 números del 76 al 100.

Una serie de medias móviles puede ser calculada para cualquier serie temporal. Se usa para demanda estable, sin tendencia ni estacionalidad; suaviza las fluctuaciones de plazos cortos, resaltando así las tendencias o ciclos de plazos largos

Una media móvil simple (Moving Average) es la media aritmética de N los datos anteriores. En esta técnica elemental de predicción, mientras más grande sea N, mayor será la influencia de los datos antiguos. En contrapartida, si se selecciona un N bajo, se tendrán en cuenta datos más recientes para nuestra predicción.

De acuerdo a lo enunciado con anterioridad, concluimos que la elección de N influenciará decisivamente nuestra predicción. Dependiendo del tipo de datos de serie temporal analizados podremos adaptar eficazmente nuestra predicción a los mismos. Así, si se elige un N bajo, nuestra predicción tendrá una alta capacidad para responder rápidamente ante fluctuaciones o variaciones significativas en los datos de un período a otro. Sin embargo, la predicción en este caso estará altamente influenciada por efectos aleatorios. Por otro lado, la elección de un N muy alto provocará que, aunque se filtre la existencia de efectos aleatorios, nuestras predicciones presenten una adaptación lenta ante fluctuaciones significativas en los datos de períodos más recientes, pues dicha predicción estará teniendo en cuenta el valor de datos antiguos.

3.2.2 Media fija

Al igual que la media móvil en estadística la media fija es un método utilizado para analizar un conjunto de datos en modo de puntos para crear series de promedios. Así las medias fijas son una lista de números en la cual cada uno es el promedio de un subconjunto de los datos originales. A diferencia de la media móvil, si se tiene un conjunto de 100 datos el primer valor de la serie de medias fijas sería el promedio de los primeros 25 términos, el segundo correspondería al promedio de los siguiente 25 términos, y así hasta completar la serie.

Del mismo modo de la media móvil, una serie de medias fijas puede ser calculada para cualquier serie temporal. Se usa para demanda estable, sin tendencia ni estacionalidad; suaviza las fluctuaciones de plazos cortos, resaltando así las tendencias o ciclos de plazos largos.

3.2.3 Prueba de las carreras

Al estudiar series cronológicas, la primera pregunta que necesita contestarse es si la serie realmente depende del tiempo, pruebas estadísticas como la de carreras pueden usarse para responder esta pregunta.

Si es cierto que la serie depende del tiempo, el siguiente paso es el de estimar la naturaleza de esta dependencia, puesto que solamente son de interés principal los grandes movimientos de la serie a lo largo del tiempo. Los métodos de regresión pueden usarse para obtener esta estimación; sin embargo, otro método, basado en los promedios móviles se emplea comúnmente para llevar a cabo esta estimación (Hoel, 1964).

Frecuentemente es claro, por medio de una simple inspección ocular de sus graficas, que gran parte de las series de tiempo cambian de manera regular con el tiempo.

Por ejemplo, si la serie consiste de mediciones de la altura en periodos semanales para una planta en crecimiento, es obvio que las mediciones serán dependientes del tiempo ya que aumentan con él. La gráfica del poder adquisitivo del dólar por los últimos cincuenta años sería de esperarse que mostrara una fuerte dependencia sobre el tiempo, debido que, como cualquier persona de suficiente edad puede decirlo, el valor del dólar ha ido aumentando en este periodo de tiempo. Sin embargo, si la serie consistiera de los rendimientos de acciones ordinarias, producción de carbono o porcentaje de fuerza humana empleada, sumados todo ellos anualmente, durante los últimos cincuenta años, entonces ya no resulta tan claro el que se presente una dependencia definida por el tiempo(Hoel, 1964).

La dificultad con la inspección de la grafica de una serie cronológica y por lo tanto el tratar de determinar si la serie se comporta como una muestra al azar de alguna población estable o si depende del tiempo es que la variación de una serie al azar frecuentemente engaña al observador, haciendo creer que es dependiente del tiempo. La mayor parte de las personas que observan una serie cronológica, tienden a ver la presencia de movimientos regulares en la serie, independiente de que tal regularidad exista o no. Por ejemplo, si se deja a la imaginación en libertad, indudablemente se verán algunos movimientos bastantes regulares en las series. Para series que no son claramente dependientes del tiempo, es necesario aplicar alguna prueba estadística antes de justificarse el procedimiento de análisis de la serie (Hoel, 1964).

Cuando se han tomado datos repartidos en un periodo de tiempo, y existe razón para creer que las observaciones pueden no ser un conjunto de observaciones al alzar, es aconsejable el aplicar una prueba que compruebe la hipótesis la hipótesis de azar. Esta prueba, que es bastante útil y fácil de aplicar se basa en carreras (Hoel, 1964).

El primer paso para aplicar la prueba de carreras para un grupo de mediciones, es encontrar la mediana del conjunto. Ahora bien, cada medida en el grupo se substituye por la letra a si es mayor que la mediana y por la letra b si es inferior.

Una secuencia de letras idénticas precedida y seguida por una letra distinta (o sin letra si se encuentra al principio o al final de toda la secuencia) se llama carrera. La longitud de la carrera está determinada por el número de letras idénticas en ella.

La prueba que se habrá de presentar aquí, depende solamente del número total de carreras en toda la secuencia y por lo tanto no trata de las longitudes de las carreras. Si se emplea la letra u para denotar el número total de carreras, entonces se supone que la secuencia constituye una muestra al azar para alguna población, y puede obtenerse la distribución de la variable u por métodos matemáticos relativamente simples. Esta distribución dependerá de n1 y n2, que denotan el numero de a's y b's en el conjunto. Puesto que existe generalmente un número igual de a's y b's cuando se obtienen determinando si las mediciones se encuentran arriba o debajo de la mediana del grupo, podría aparecer que n1 y n2, solas serian suficientes. Sin embargo, la prueba se aplica también a problemas en las que las a's y b's se obtienen por métodos diferentes y en algunos de estos problemas, n1 y n2 pueden ser también bastante diferentes entre sí (Hoel, 1964).

Luego se aplica una prueba unilateral donde la región crítica contiene solamente valores pequeños de u. Puesto que la aproximación normal de muestra larga se representa por la formulas,

$$\mu_u = \frac{2n_1 n_2}{n_1 + n_2} + 1$$

$$\sigma_u = \sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 + (n_1 + n_2 - 1)}}$$

Se sigue que la región critica para esta prueba deberá consistir de aquellos valores de z que satisfacen la desigualdad $z < -1.64$, en la cual

$$z = \frac{u - \mu_u}{\sigma_u}$$

Si los valores de *z* son menores que -1.64 la serie cronológica dependería del tiempo, de lo contrario se comportaría como una serie a azar.

3.2.4 Wilcoxon suma de rangos

Los cambios y tendencias comúnmente encontrados en series temporales en recursos hidráulicos e ingeniería del medio ambiente pueden ser el resultado de factores naturales o intervenciones humanas. Por ejemplo, eventos naturales como el cambio climático global, quemas forestales, explosiones volcánicas y derrumbes, pueden inducir cambios y tendencias, mientras que las variaciones por intervenciones humanas ocurren debido a cambios en la cultura y el uso de la tierra, por deforestación, o en general, destrucción o alteración de la cobertura vegetal, pastoreo de ganado, alteración del suelo o modificación de prácticas de agricultura. Por definición el cambio en la media es un cambo abrupto (ocurrido en un periodo corto de tiempo) en el nivel o magnitud de un momento o parámetro de ciertas variables (INTEGRAL, 2012).

La prueba utilizada en el presente proyecto de grado para evaluar cambios en la tendencia de la media en series es la prueba de Wilcoxon suma de rangos o Mann-Whitney, la cual se explica a continuación:

Supóngase que se tienen dos poblaciones continuas independientes $X_1 \, y \, X_2$ con medias $\mu_1 \, y \, \mu_2$ que ambas tienen la misma forma y dispersión y que (posiblemente) difieren en sus medias. La prueba de la suma de rango puede emplearse para probar la hipótesis $H_0: \mu_1 = \mu_2$. Este procedimiento también se conoce como prueba de Mann-Whitney, aunque lo usual es expresar el estadístico de prueba de Mann-Whitney en forma diferente (Montgomery y Runger, 1996).

Sean $X_{11}, X_{12} \ldots\ldots, X_{1n_1}$, y $X_{21}, X_{22} \ldots\ldots, X_{2n_2}$ dos muestras aleatorias independientes de tamaños $n_1 \leq n_2$ tomadas de las poblaciones continuas $X_1 \, y \, X_2$ descritas anteriormente. Se desea probar la hipótesis

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 = \mu_2$$

El procedimiento a seguir es el siguiente, se acomodan todas las $n_1 + n_2$ observaciones en orden ascendente de magnitud y se les asigna un rango. Si dos o más observaciones quedan empatadas (son idénticas), entonces se emplea la media de los rangos que se le habría asignado si fuesen diferentes. Sea W_1 la suma de los rangos en la muestra más pequeña, y defínase W_2 como la suma de los rangos en la otra muestra (Montgomery y Runger, 1996). Entonces:

$$W_2 = \frac{(n_1 + n_2) + (n_1 + n_2 + 1)}{2} - W_1$$

Ahora bien, si la medias muestrales no difieren entre sí, entonces se espera que la suma de los rangos sea casi igual para las dos muestras después de ajustar el tamaño de muestra. En consecuencia, si las sumas de los rangos son muy diferentes, se concluye que las medias no son iguales (Montgomery y Runger, 1996).

Cuando $n_1\ y\ n_2$son más o menos grandes (por ejemplo, mayores que 8), la distribución de W_1 puede aproximarse de buena manera con una distribución normal con media

$$\mu_{W_1} = \frac{n_1(n_1 + n_2 + 1)}{2}$$

y varianza

$$\sigma^2{}_{W_1} = \frac{n_1 n_2(n_1 + n_2 + 1)}{12}$$

Por consiguiente, para $n_1\ y\ n_2 > 8$, puede emplearse

$$Z_0 = \frac{W_1 - \mu_{W_1}}{\sigma_{W_1}}$$

Una vez explicada la metodología se presenta el resultado de la detección de cambios en la media. La confiabilidad utilizada es del 95%, por lo tanto $Z_{1-\frac{\alpha}{2}} = 1.96$, será el estadístico de prueba.

3.3 TRANSFORMADA RÁPIDA DE FOURIER APLICADA A SERIES DE TIEMPO

Otra metodología que existe para evaluar series de tiempo, consiste en no evaluar estas series en el dominio del tiempo, sino en el dominio de la frecuencia. De tal manera una señal cualquiera definida en el dominio del tiempo, f(t), puede ser transformada en una función definida en el dominio de la frecuencia, F(ϖ), mediante el empleo de la definición de la transformada de Fourier (Botero, 2011)

$$F(\tilde{\omega}) = \int_{-\infty}^{+\infty} f(t) e^{-i\tilde{\omega}t}\, dt$$

El valor de ϖ es variable y corresponde a la frecuencia circular de vibración de los armónicos que componen la señal en estudio. De igual forma, una señal cualquiera que esté definida en el dominio de la frecuencia, F(ϖ), se puede transformar al dominio del tiempo, f(t), empleando la definición de la transformada inversa de Fourier (Botero, 2011)

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(\tilde{\omega}) f(t) e^{-i\tilde{\omega}t}\, d\omega$$

Actualmente existe un método numérico ampliamente aceptado por la comunidad científica ingenieril, el cual es conocido con el nombre de la Transformada Rápida de Fourier (Fast Fourier Transform, FFT) o de la Transformada Rápida de Fourier Inversa (FFT^{-1}) y fue propuesta por Cooley and Turkey (1965).

La Trasformada Rápida de Fourier es un método numérico muy eficiente para evaluar la trasformada discreta de Fourier y está desarrollado en base binaria. Por tal motivo, se requiere que el número de puntos discretos, N, que definen la serie, sea una función de potencia par ($N= 2^M$) donde M es un número entero cualquiera. La eficiencia del método comienza a evidenciarte cuando la señal discreta está conformada por mil o

más puntos (Clough and Penzien, 1993). Ahora bien, es poco probable que una serie cualquiera, tenga exactamente el número de puntos requerido para emplear el método. Esto se puede subsanar agregándole al registro una cantidad de puntos iguales a cero, de tal forma que se complete el número de puntos necesario. Por efectos de estabilidad numérica, se recomienda que el número total de puntos incluyendo los ceros, sea al menos un 30 por ciento mayor que el número de puntos que definen el registro original (Botero, 2011).

Al aplicar la FFT a una señal definida en el dominio del tiempo, ésta se descompone en una serie de puntos complejos, asociados a una determinada frecuencia, ϖ. Esta serie de puntos presenta una particularidad numérica interesante: la serie de la componente real es simétrica respecto a la frecuencia media de la serie conocida como Frecuencia de Nyquist, $\varpi_{Nyquist}$ y la serie del componente imaginario es la conjugada respecto a la frecuencia en mención (Botero, 2011).

El valor de la Frecuencia de Nyquist ($\varpi_{Nyquist}$)se puede calcular mediante la siguiente expresión:

$$\widetilde{\omega}_{\text{Nyquist}} = \frac{N\Delta\widetilde{\omega}}{2}$$

Donde el diferencial de frecuencia, $\Delta\varpi$, se define así:

$$\Delta\widetilde{\omega} = \frac{1}{\Delta t(N-1)}$$

El valor del diferencial de frecuencia indica el espaciamiento que deben tener los puntos de la señal compleja en el dominio de la frecuencia.

Es importa mencionar que cada una de las metodologías que componen el marco teórico anteriormente indicado, se aplicaran de tal manera que sean conveniente para analizar la información base y el tipo de variables que componen el estudio, para así evitar salirnos del contexto del problema que desea ser evaluado.

A continuación se describe la zona de estudio y se caracteriza la información encontrada en la misma.

4 DESCRIPCIÓN DE LA ZONA DE ESTUDIO

La zona de estudio se localiza sobre el rio Cauca, la segunda cuenca de la zona andina más importante del país. Este cauce fue seleccionado para la validación del modelo, gracias a que es una cuenca bien instrumentada, cuenta con una buena longitud de registros (en algunos casos existe información desde 1960), y el caudal está bien regulado, dada la extensión y geomorfología de la cuenca hidrográfica.

El río Cauca, a partir de su nacimiento, en el extremo sur-occidental del país, a unos 4.200 m.s.n.m., se dirige hacia el norte, entre las cordilleras Central y Occidental, manteniéndose sensiblemente paralelo a la costa Colombiana sobre el mar Pacífico, para desembocar, luego de un recorrido de unos 1.290 km y a los 90 msnm, en el río Magdalena, el principal río del país que, a su vez, hace parte de la vertiente Atlántica. (INTEGRAL, 1999).

A lo largo del cauce se cuenta con una vasta red de estaciones limnigráficas y limnimetricas con registros de caudales y sedimentos que datan desde la década de los sesenta, como se mencionó anteriormente, lo que hace confiables las predicciones de caudal asumidos para el diseño de las obras.

El valle formado por el rio Cauca constituye el eje de desarrollo de las diferentes actividades socioeconómicas de la región suroccidental de Colombia, como son, la generación de energía, la extracción de materiales del lecho y el abastecimiento de agua para el consumo agrícola, doméstico e industrial; sin embargo, esta situación ha traído como consecuencia el uso del río como receptor de los vertimientos generados por dichas actividades, ocasionando un deterioro en la calidad del agua lo cual genera el agotamiento del oxígeno disuelto y pérdida de biodiversidad entre otros.

La región presenta una precipitación media anual de 1597 mm (±224 mm) y 21,3°C (±0,5 °C) de temperatura media del aire. (Puertas et al, 2011).

Debido a influencia de los vientos alisios y la migración de la Zona de Convergencia Intertropical, se presenta un régimen bimodal con periodos lluviosos: en Marzo-Abril-Mayo y Septiembre-Octubre-Noviembre, así como de menor precipitación durante Diciembre-Enero-Febrero y Junio-Julio-Agosto. Además, las fases extremas del fenómeno El Niño Oscilación Sur (ENOS) y su mayor recurrencia y acentuación de sus fases extremas por efectos del cambio climático influencian la variabilidad interanual del clima. (Puertas et al, 2011).

La zona de interés del presente estudio corresponde al tramo del río Cauca comprendido entre las estaciones limnigráficas La Pintada (2618711) y Las Flores (2502727) operadas por el IDEAM, con una longitud de 386 Km como se presenta en la Figura 4—1. Este tramo del río se encuentra entre los municipios La Pintada y Nechi.

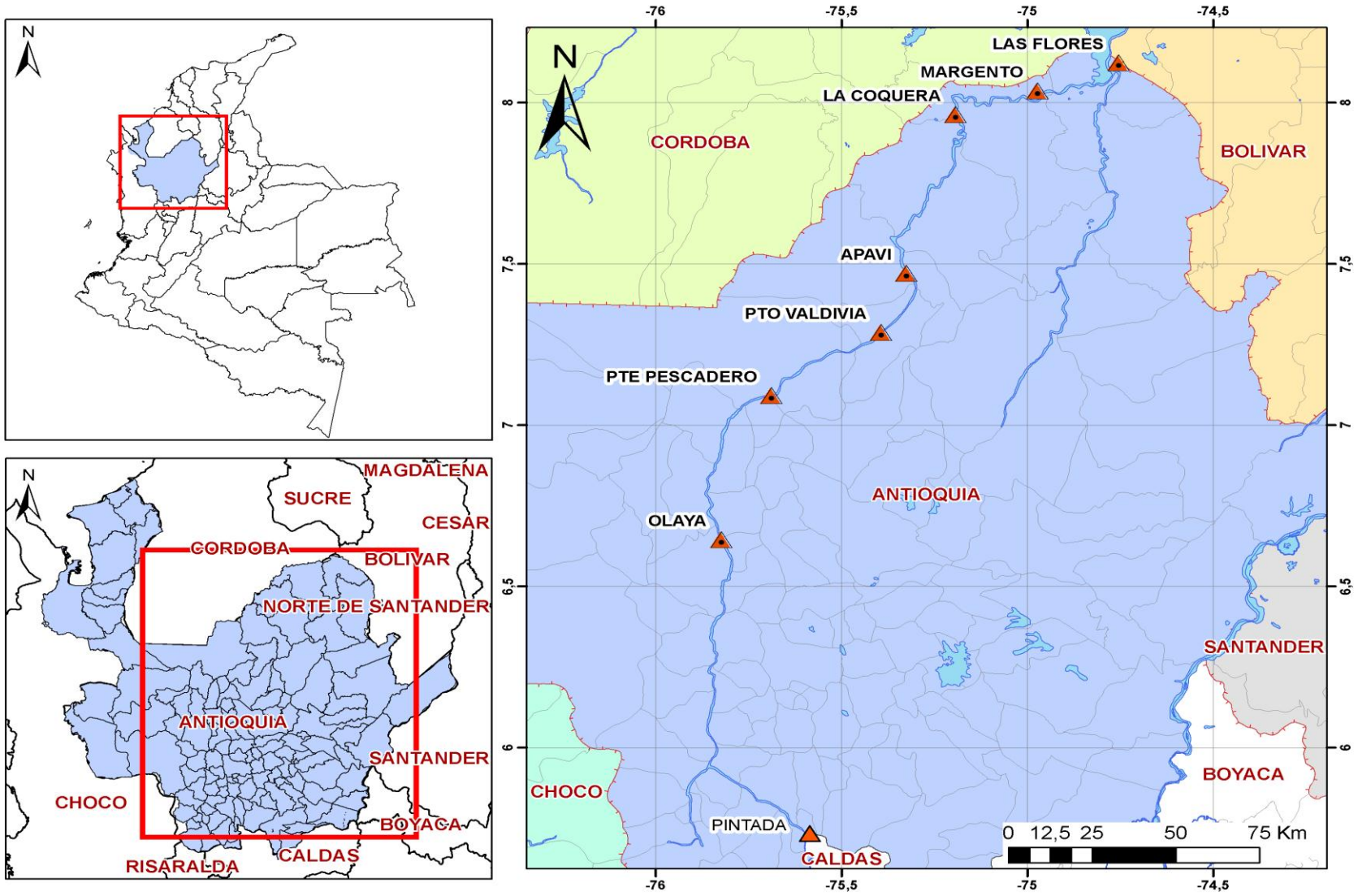


Figura 4—1 Cuenca drenaje del rio Cauca en la zona de estudio.

5 INFORMACIÓN BASE

Como información inicial para elaborar el presente estudio, se contó con registros de caudales medios diarios en 10 estaciones limnigráficas o limnimétricas ubicadas sobre el rio cauca, estas estaciones indican el inicio y el final del tramo de análisis como se indicó anteriormente.

Las estaciones limnigráficas son aquellas que permiten la determinación de caudales por medio de registro gráfico continuo de los niveles de agua, y las limnimétricas permiten determinación de caudales por medio de lecturas periódicas sobre una regla graduada llevadas a cabo por un observador, la frecuencia de estas lecturas es de dos veces por día.

En la Tabla 5-1, se muestran las estaciones utilizadas para este estudio, donde se aprecia el código de la estación, su nombre y el tipo de estación de medición. Cabe mencionar que sin importar el tipo de estación, ya sea limnigráfica o limnimétrica, el dato registrado por la misma y utilizado para el estudio es el caudal medio diario.

Tabla 5-1 Estaciones analizadas

Código	Nombre	Operador	Tipo de estación
2618711	La Pintada	IDEAM	Limnigráfica
2620703	Puente Iglesias	IDEAM	Limnimétrica
2620708	Bolombolo	IDEAM	Limnigráfica
2621705	Cañafisto	IDEAM	Limnigráfica
2623705	Pescadero	IDEAM	Limnigráfica
2623704	Valdivia	IDEAM	Limnigráfica
2624703	Apaví	IDEAM	Limnigráfica
2624702	La Coquera	IDEAM	Limnigráfica
2502705	Margento	IDEAM	Limnigráfica
2502727	Las Flores	IDEAM	Limnigráfica

6 ANÁLISIS Y SELECCIÓN INICIAL DE ESTACIONES

De cada una de las estaciones mostradas en la Tabla 5-1, se analizaron los registros obtenidos, con el objetivo de identificar que estaciones son las más apropiadas para los análisis a ser realizados.

Inicialmente de cada una de estas estaciones se buscó el año en el que inician y terminan los registros, con el fin de estimar el periodo de registro de cada una de estas, en la Tabla 6-1 y Figura 6—1se puede apreciar cada una de las estaciones inicialmente analizadas con su respectivo año de inicio de lecturas, año final de lecturas y periodo de registro de esta.

Tabla 6-1 Años de registro de las estaciones analizadas

Código	Nombre	Año inicial	Año Final	Años de registro
2618711	La Pintada	1965	1998	34
2620703	Puente Iglesias	1979	1990	12
2620708	Bolombolo	1971	1990	20
2621705	Cañafisto	1979	1998	20
2623705	Pescadero	1984	2009	26
2623704	Valdivia	1960	2011	52
2624703	Apaví	1972	2011	40
2624702	La Coquera	1966	2011	46
2502705	Margento	1966	2011	46
2502727	Las Flores	1974	2011	38

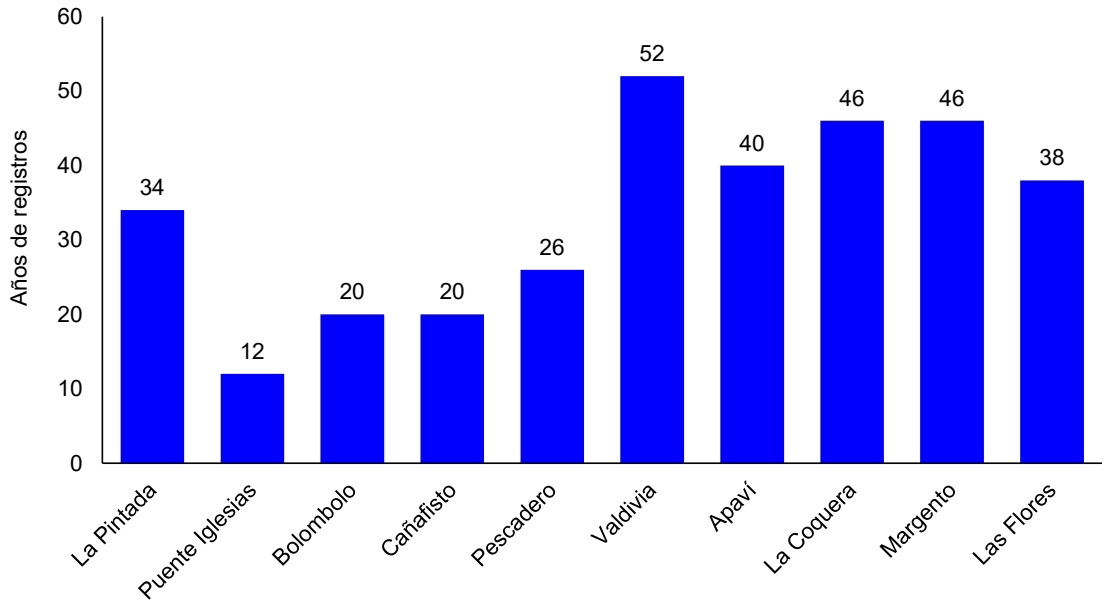


Figura 6—1 Años de registro de las estaciones analizadas

Como se observa en la Tabla 6-1, los periodos de registros oscilan entre 12 y 52 años, donde las estaciones con menores periodos de registros son Puente Iglesias con 12 años, Bolombolo con 20 años al igual que Cañafisto. Se determinó que estas estaciones no son adecuadas para la elaboración del estudio, ya que la longitud de los registros no es lo suficientemente amplia para las metodologías que serán aplicadas, y adicionalmente periodos tan cortos de registros no representan adecuadamente el comportamiento hidrológico de la zona de estudio.

Sin embargo, es importante mencionar que el periodo de registro que se indicó anteriormente, no representa realmente la cantidad de lecturas que posee cada estación, dado que estos registros presentan alguno vacíos en su serie de tiempo debido a que la estaciones en algún momento de su operación posiblemente fueron víctimas de actos vandálicos, o los equipos no se encontraban calibrados, o se presentaron mantenimientos no programados. Por tal motivo se realizó un análisis minuciosos de cada uno de las series, con el objetivo de estimar la cantidad real de registros diarios y su equivalente en años, los cuales se muestran en la Tabla 6-2 y en la Figura 6—2

Tabla 6-2 Numero de datos reales registrados y su equivalente en años de las estaciones analizadas

Código	Nombre	Datos registrados	Años reales	Años de registro
2618711	La Pintada	12189	33	34
2623705	Pescadero	6791	19	26
2623704	Valdivia	16580	45	52
2624703	Apaví	12748	35	40
2624702	La Coquera	14779	40	46
2502705	Margento	12094	33	46
2502727	Las Flores	10896	30	38

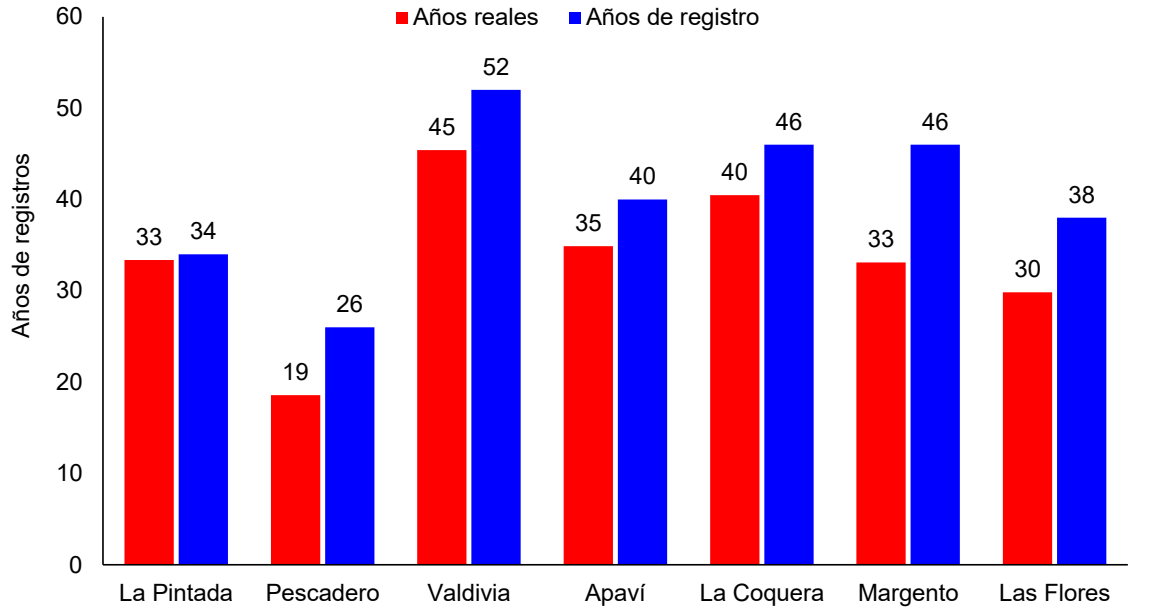


Figura 6—2 Número de datos reales registrados y su equivalente en años de las estaciones analizadas

7 BÚSQUEDA Y ELIMINACIÓN DE OUTLIERS EN LAS ESTACIONES SELECCIONADAS

Con cada una de las estaciones anteriormente mostradas se procedió a aplicar las 5 metodologías de Outliers presentadas en el marco teórico, las cuales son Dixon, Studentized Deviates, Rosner, Em y Lm. Conviene mencionar que cada una de estos procedimiento fueron programados mediante macros de Excel con el fin de facilitar y agilizar su ejecución, en el Anexo 1 se presenta el manual que indica como ejecutar dichos programas.

Dado que la aplicación de estos métodos tienen algunas condiciones, a continuación se mencionan las consideraciones generales que se tuvieron en cuenta para la ejecución de estas metodologías estadísticas:

- Para todos los métodos se consideró un nivel de significancia de 0,05.
- En todos los casos se trabajaron muestras de 25 datos, pero de manera móvil, es decir, cada valor era comparado con los 25 datos anteriores y 25 posteriores, no se consideró tomar muestras mayores para la ejecución, dado que la prueba de Dixon trabaja máximo con muestras de 25 datos.

Es importante dejar claro la forma en que se eliminarán los outliers de las series de caudales, por lo cual a continuación se presenta a manera de ejemplo la eliminación de un outlier.

Inicialmente se le aplico las metodologías empleadas para la búsqueda de outliers con cada una de las consideraciones anteriormente indicadas, después de haber identificado el valor anómalo, se tomó un tramo de 50 datos y se grafico como se muestra en la Figura 7—1, este valor fue identificado por las pruebas de Dixon, Studentized Deviates y Rosner.

Como se aprecia gráficamente el valor se encuentra muy alejado de la tendencia de los datos, y por consiguiente, este fue remplazado por el promedio entre el dato anterior y posterior a él, logrando suavizar la serie como se muestra en la Figura 7—2.

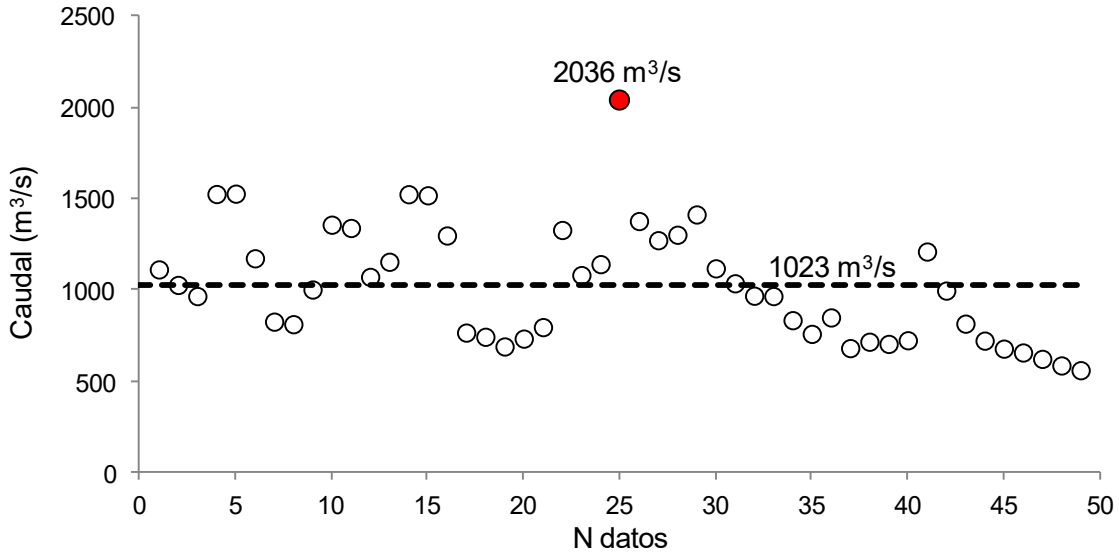


Figura 7—1 Serie ejemplo con outlier identificado

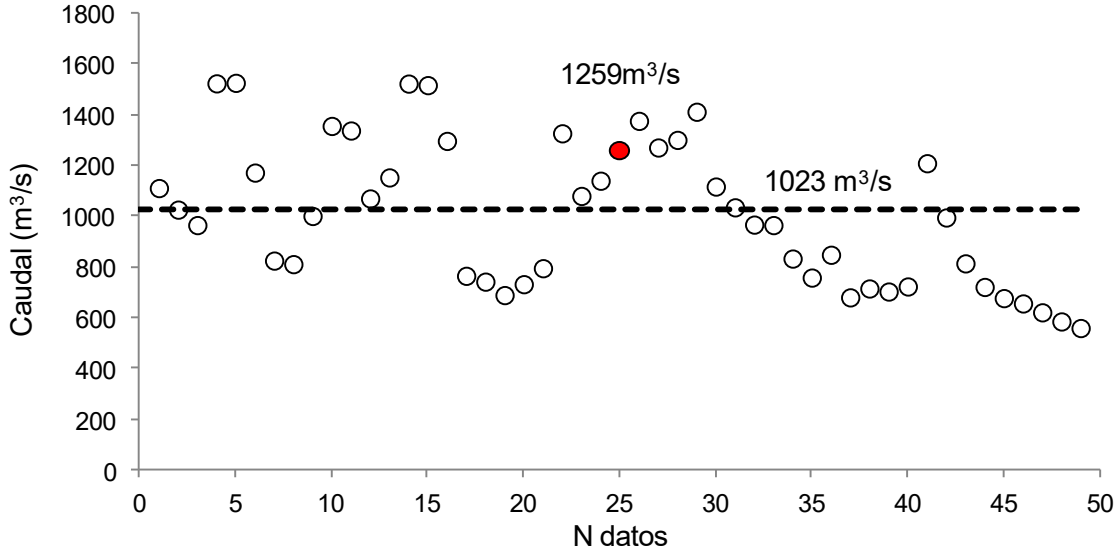


Figura 7—2 Serie ejemplo con outlier suavizado

Con cada una de estas consideraciones se ejecutaron los métodos de búsqueda de valores extremadamente anómalos, dando como resultado la identificación estadística de una gran cantidad de Outliers por estación como se muestra en la Tabla 7-1.

Tabla 7-1 Outliers identificados por estación

Nombre	Dixon	Lm	Em	Studentized Deviates	Rosner
La Pintada	379	16	9	1131	254
Pescadero	215	60	705	516	146
Valdivia	476	125	992	1317	315
Apaví	364	42	698	1033	248
La Coquera	256	125	2507	943	171
Margento	256	125	2507	943	171
Las Flores	200	171	2731	940	141

Como se puede observar en la Tabla 7-1, cada prueba identifica un numero diferente de Outliers, ya que algunas pruebas son un poco más conservadoras que otras. En el caso de la prueba denominada Em, debido a su metodología, marca los espacios en blanco de la serie como datos anormalmente extremos.

Por otro lado, la prueba de Studentized Deviates es la más conservadora, dado que marca como Outliers, valores que no están tan alejados de la muestra. Debido a esto, las dos pruebas anteriores son las que contienen mas valores marcados como Outliers. Sin embargo, es notable que la prueba que menos valores identifica como Outliers es la Lm, y por este motivo se considera la metodología menos conservadora.

Considerando los Outliers identificados por cada una de la pruebas en cada estación, se procedió a determinar cuáles de estos valores serian eliminados. Esta eliminación no se automatizó con base en las metodologías estadísticas de identificación de Outliers, aunque se considera que estas son una guía importante para determinar estos valores, sin embargo, se le debe dar mayor importancia a la experiencia en análisis de series de caudales, dado que esta variable en ríos tiene un comportamiento que depende del clima, las condiciones de la cuenca, entres otras. Por lo cual en este estudio se realizó un barrido de cada unas de las series, identificando la magnitud de los caudales, con el fin de tratar de entender el comportamiento de estos valores en el tiempo. Con base en este barrido y en las pruebas de Outliers, se seleccionaron los valores a eliminar de las series.

Finalmente, la cantidad de valores eliminados en cada una de las estaciones se muestran en la Tabla 7-2 y la Figura 7—3.

Tabla 7-2 Cantidad de Outliers eliminados en cada estación

Nombre	Outliers eliminados
La Pintada	50
Pescadero	13
Valdivia	69
Apaví	59
La Coquera	58
Margento	25
Las Flores	21

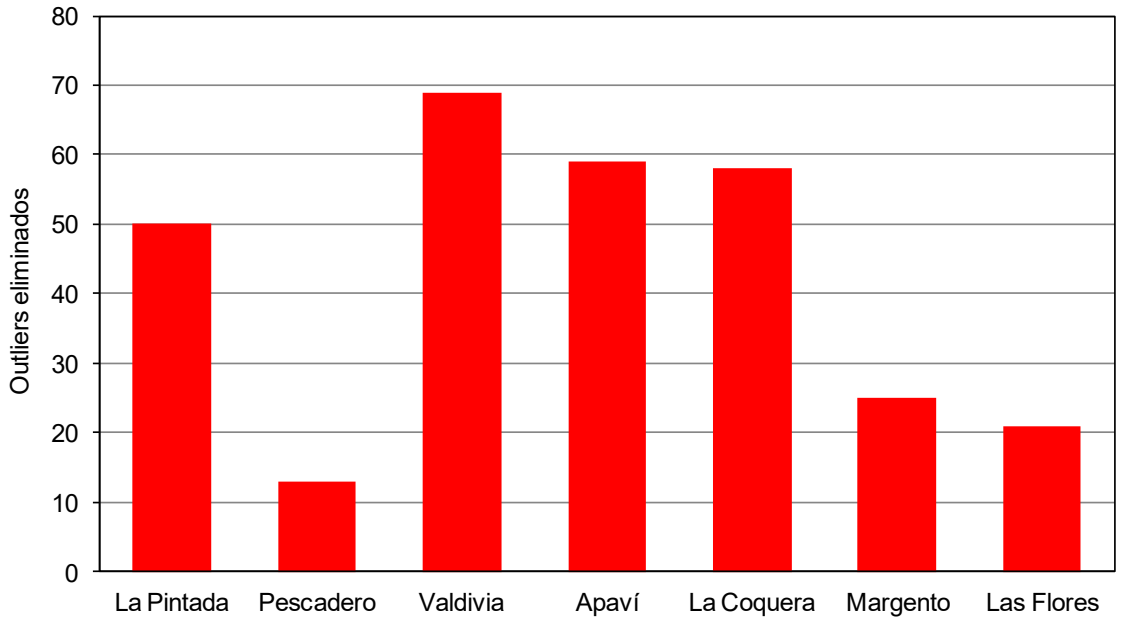


Figura 7—3 Cantidad de Outliers eliminados en cada estación

Es importante aclarar, que en lugar de eliminar estos valores, se procedió a remplazarlos por el promedio entre el dato anterior y el siguiente a él, con el fin de no generar más vacíos en la serie, debido a que estos no son convenientes para continuar con el análisis.

8 ANÁLISIS DE LA MEDIA EN LAS ESTACIONES SELECCIONADAS

Continuando con el análisis estadístico de las series de caudales de la estaciones ubicadas sobre el río Cauca, se procede con el análisis de homogeneidad de la series, es decir, un análisis de tendencia de las mismas, este procedimiento se realiza con el fin de identificar cambios en la tendencia de los datos con respecto al tiempo.

Las metodología aplicadas para la homogeneidad se componen básicamente en análisis de las medias anuales y cada N años, con base en el concepto de media móvil y media fija que se explicaron con mayor detalle en el marco teórico del presente estudio.

También se analizó la dependencia del tiempo mediante la aplicación de la prueba de carreras y se verificó que la serie conserva la misma tendencia si es partida en dos mediante la prueba de Wilcoxon suma de rangos, que al igual que las metodologías de las medias móviles y fijas, se explica con más detalle en el marco teórico.

A continuación se presenta cada una de las pruebas mencionadas y los resultados encontrados en cada una de las estaciones.

8.1 MEDIA MÓVIL

Con base en los caudales medios diarios en cada estación, se estimó el caudal promedio de cada uno de los años de registro, con los cuales, se realizaron las medias móviles cada 2 y 3 años, para así obtener un caudal medio cada 1,2 y 3 años, con el fin de suavizar la fluctuaciones de los caudales anuales para resaltar la tendencia que presentan los datos de cada estación, y de esta manera identificar si estos caudales tienden a aumentar o disminuir en el tiempo, o por lo contrario, no presentan ninguna dependencia del él.

Utilizando los conceptos básicos y la teoría de las medias, se desarrolló un programa para determinar de manera ágil y rápida los caudales medios anules de cada estación.

Estos resultados fueron recopilados, para así, lograr analizar de manera precisa la tendencia que presentan los datos de cada serie, identificando posibles cambios en la tendencia con respecto al tiempo.

Inicialmente en la Figura 8—1, se presentan los valores de los caudales medios anuales, con los cuales se determinó la media móvil.

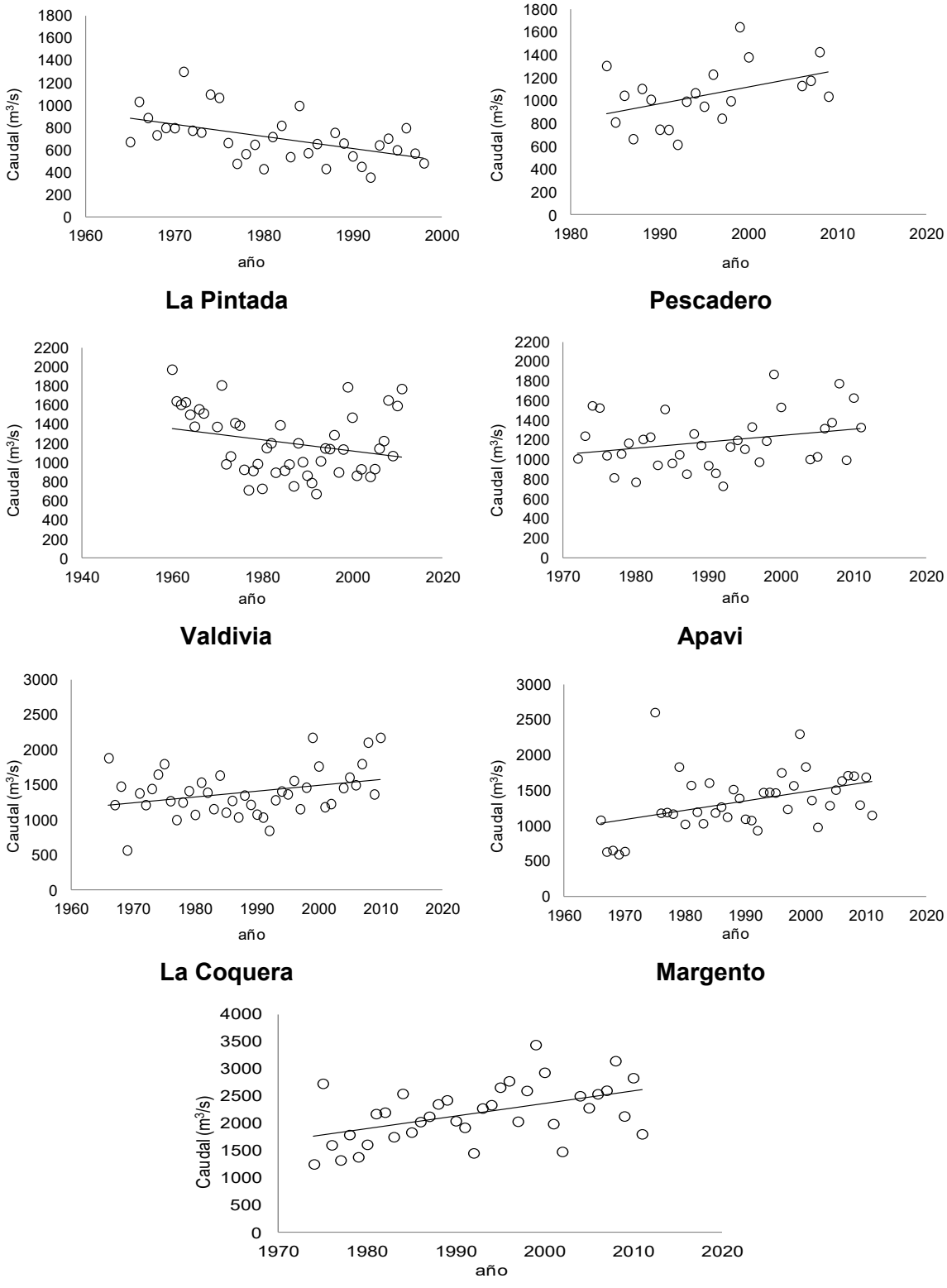


Figura 8—1 Medias anuales de las estaciones analizadas

Como se puede observar en la Figura 8—1, se presenta un comportamiento esperado, dado que el orden de magnitud de los caudales aumenta en el sentido del flujo, es decir, entre mas aguas abajo se encuentre la estación de medición de caudal, la magnitud de esta medida aumenta gradualmente.

Lo anterior se puede identificar claramente en la Figura 8—2 y si comparamos las magnitudes de los caudales medios de la primera y última estación elegidas para el estudio, La Pintada con un caudal medio de 703 m3/s y Las Flores con 2180 m3/s. Es claro que entre estas estaciones la magnitud del caudal se triplica, lo cual indica que en el tramo entre estas estaciones existe un gran aporte hídrico al río Cauca.

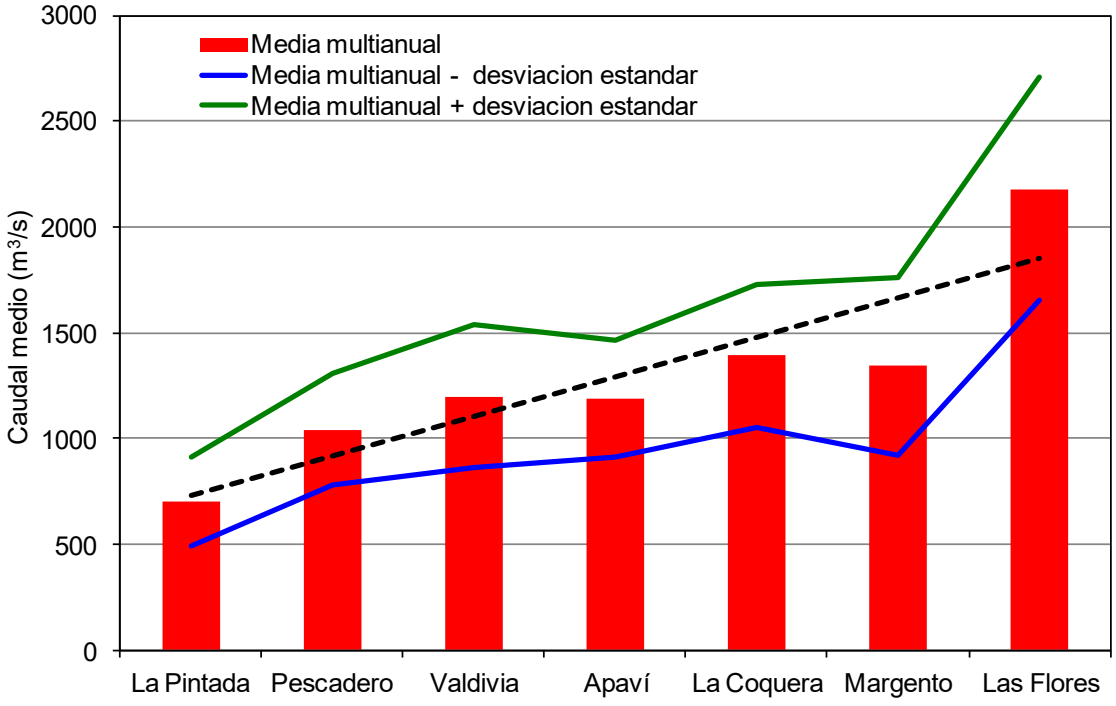


Figura 8—2 Caudal medio anual de las estaciones analizadas

Por otro lado, analizando esta información se observó que la tendencia de los caudales con respecto al tiempo aumenta, lo cual puede ser generado por las intervenciones antrópicas en la cuenca del río Cauca en los últimos años, en los cuales el país en general ha presentado un fuente incremento en la industria en general, y ha aumentado su población generando un gran crecimiento en los pueblos y ciudades, representando un cambio importante en los usos del suelo, y por lo tanto un posible aumento en la escorrentía.

Este comportamiento es claro en la mayoría de las estaciones, excepto en las estaciones Pintada y Valdivia, las cuales presentan una leve tendencia a disminuir su caudal en el tiempo.

En la Figura 8—1, se puede identificar una gran dispersión entre los caudales medios de una misma estación, sin embargo esta dispersión disminuye a medida que el caudal

aumenta en estaciones aguas abajo, esto se observa claramente en la estación Las Flores la cual presenta medias anules homogéneas entre sí.

Para disminuir esta dispersión en cada una de las estaciones se aplicó la metodología de las medias móviles cada 2 y 3 años, con el fin de suavizar la tendencia de las medias anuales anteriormente analizadas. Estos resultados se presentan gráficamente en la Figura 8—3 y la Figura 8—4 para 2 y 3 años respectivamente.

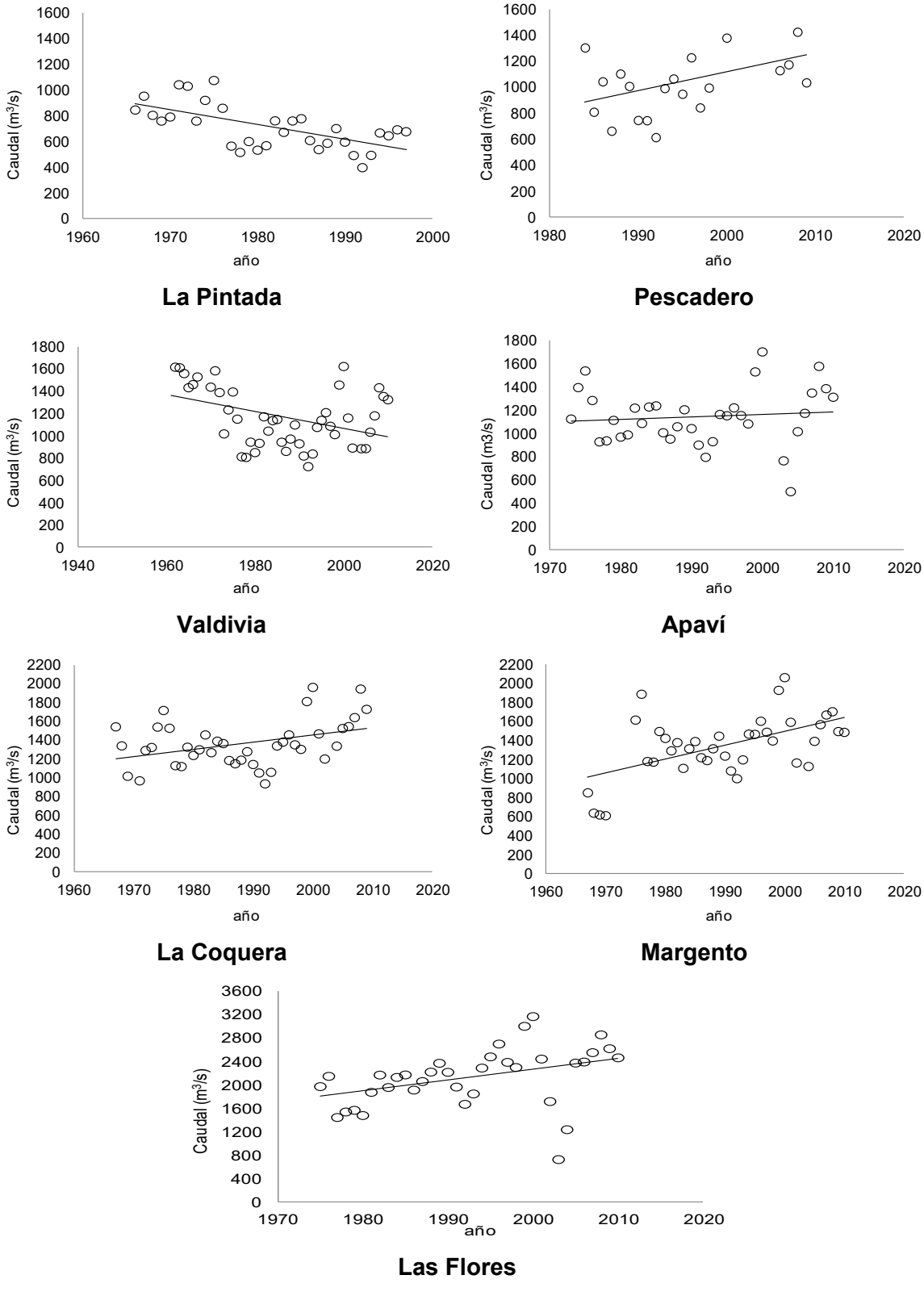


Figura 8—3. Media móvil cada 2 años

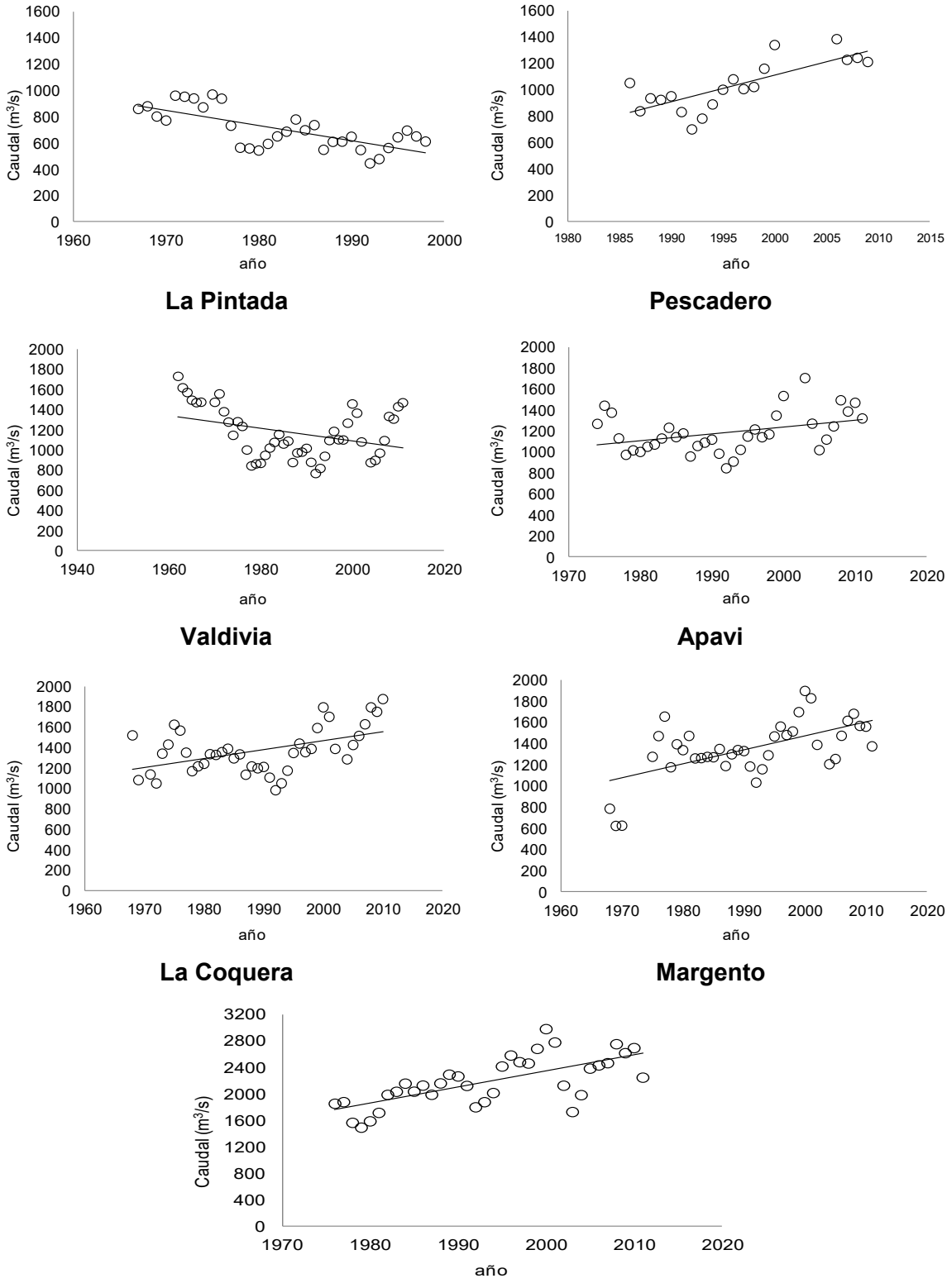


Figura 8—4. Media móvil cada 3 años

Como se ilustra en las Figura 8—3 y Figura 8—4, la dispersión de los datos disminuyo con la aplicación de las medias móviles, sin presentar mayor cambio en la tendencia de los datos, lo cual era el objetivo principal de esta metodología, disminuir dispersión sin alterar la tendencia.

Al no presentarse mayor cambio en la tendencia, se conserva el comportamiento que indica que los caudales aumentan en el tiempo en todas las estaciones excepto en La Pintada y Valdivia, como se mencionó anteriormente.

También es claro que la aplicación de las medias móviles no altera significativamente la magnitud de los caudales, por lo cual este mantiene su tendencia a aumentar en el sentido del flujo, siendo este el comportamiento esperado.

8.2 MEDIA FIJA

De igual manera que en las medias móviles, se tomaron los caudales medios diarios y se aplicó la metodología de la media fija, la cual consiste en determinar el promedio de un intervalo fijo de una serie de puntos, y de esta manera la primera media fija correspondería al promedio del caudal de los dos o tres primeros años de registro y el segundo promedio fijo correspondería a los siguientes dos o tres valores de caudales excluyendo los anteriores.

El objetivo principal de la media fija es el mismo que el de la media móvil, suavizar las fluctuaciones de los caudales anuales para resaltar la tendencia que presentan los datos de cada estación, y de esta manera identificar si estos caudales tienden a aumentar o disminuir en el tiempo, o por lo contrario, no presentan ninguna dependencia del él.

La metodología fue aplicada para 2 y 3 años en cada una de las estaciones con base a los promedios anuales, para luego ser recopilados y lograr analizarlos de una manera más precisa.

Del mismo modo se construyeron graficas con esta información para así identificar fácilmente su tendencia, las cuales se ilustran en la Figura 8—5 y Figura 8—6.

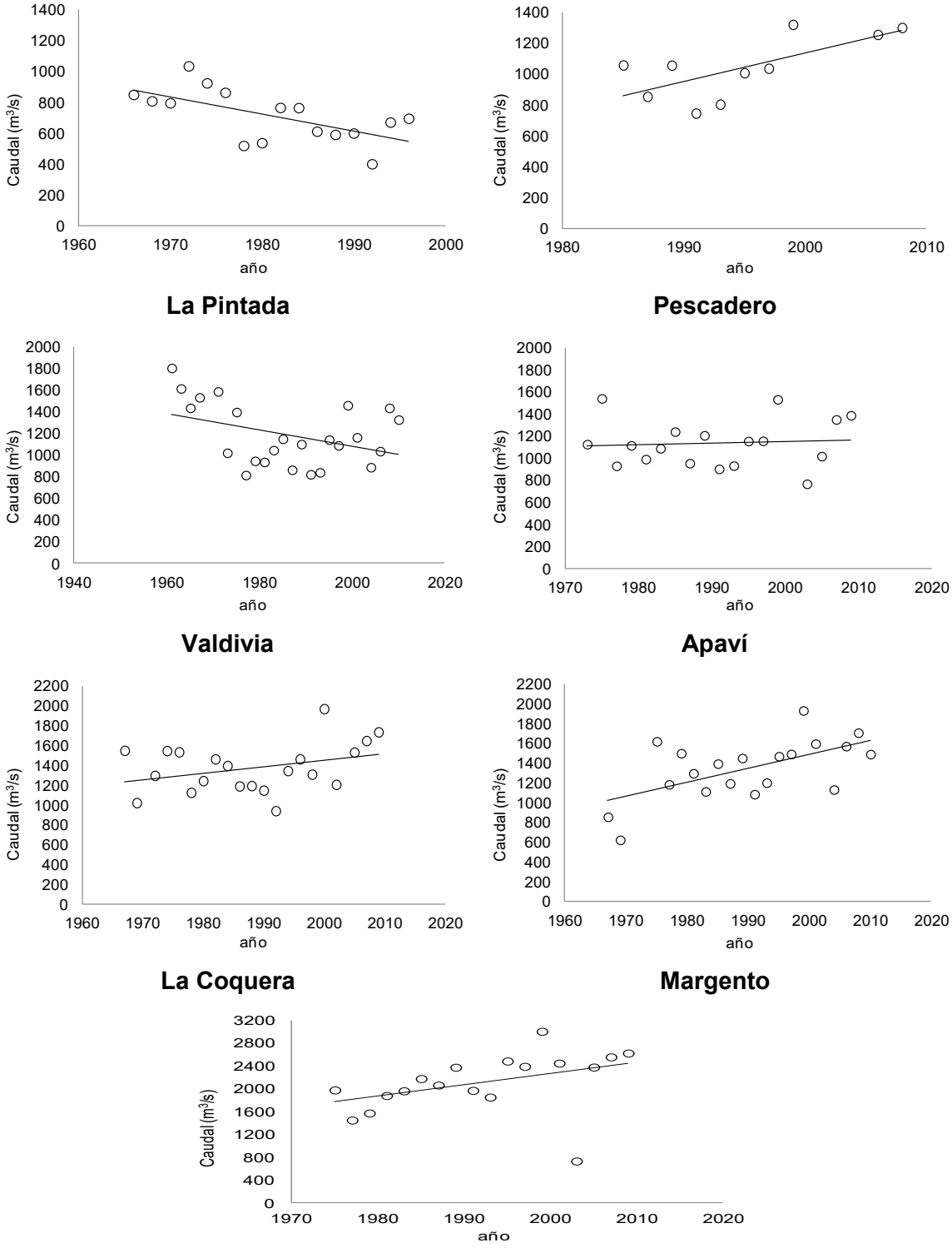


Figura 8—5. Media fija cada 2 años

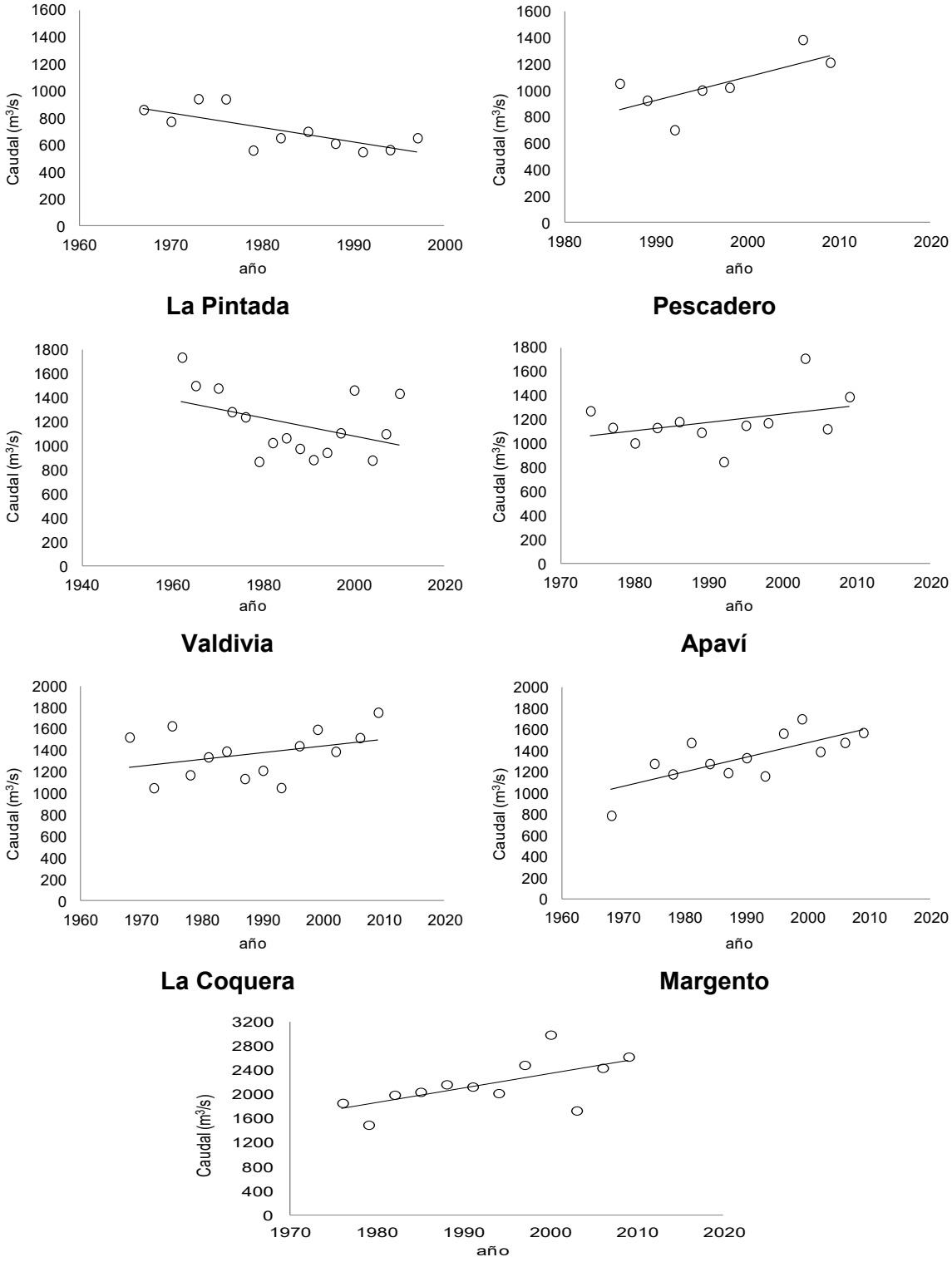


Figura 8—6. Media fija cada 3 años

Como se muestra en la Figura 8—5 y Figura 8—6, la aplicación de las medias fijas no altera de manera representativa la magnitud de los caudales, por lo cual se conserva su tendencia a aumentar en el sentido del flujo, siendo este el comportamiento esperado.

También es claro que la dispersión de los caudales medios disminuyó al aplicar las medias fijas, sin que se presente un cambio significativo en la tendencia de los caudales, logrando así disminuir dispersión sin alterar la tendencia, lo cual era el objetivo principal de esta metodología.

Como se mencionó anteriormente, al no presentarse mayor cambio en la tendencia, se conserva el comportamiento que indica que los caudales aumentan en el tiempo en todas las estaciones excepto en La Pintada y Valdivia. Sin embargo, para la media fija cada 2 años en la estación Apavi la tendencia tuvo una leve variación que llevaría a pensar que los caudales se conservan en el tiempo, pero la media fija cada 3 años de esta misma estación presenta un aumento de los caudales con respecto al tiempo.

8.3 PRUEBA DE LAS CARRERAS

Para ser consecuente con los anteriores análisis realizados con respecto a las medias tanto fijas como móviles, se aplicó una metodología que permite determinar si estas series de caudales medios anuales presentan dependencia del tiempo. Esta prueba se llama Prueba de Carreras, la cual como se mencionó en el marco teórico con base en unos conteos y unos valores estadísticos que definen la prueba, determina si un conjunto de datos presenta o no dependencia del tiempo.

Esta prueba fue aplicada para cada uno de las estaciones analizadas en el presente estudio, trayendo como resultado si los caudales medios anuales de cada una de estas dependen del tiempo ya sea de manera creciente o decreciente. Los resultados de la Prueba de Carreras se presentan en la Tabla 8-1 .

Tabla 8-1 Estadísticos Prueba de las Carreras

Estación	Pintada	Pescadero	Valdivia	Apavi	La Coquera	Margento	Las Flores
Mediana	661	1,022	1,143	1,124	1,365	1,277	2,145
U	12	11	19	20	21	17	16
n1	17	11	25	19	22	21	19
n2	16	10	24	18	21	20	18
Promedio	17.5	11.5	25.5	19.5	22.5	21.5	19.5
Desviación típica	2.82	2.23	3.46	3.00	3.24	3.16	3.00
z	-1.94	-0.21	-1.87	0.17	-0.46	-1.42	-1.16
Depende del tiempo	NO	SI	NO	SI	SI	SI	SI

Con base en los resultados presentados en la Tabla 8-1, se observa que la Prueba de las Carreras indico que todas la estaciones presentan dependencia del tiempo excepto las estaciones Pintada y Valdivia, las mismas dos estaciones que en el análisis grafico de las medias indicaban un disminución de los caudales con respecto al tiempo, esto llevaría a concluir que fue apresurado afirmar que estas estaciones presentaban una disminución de los caudales en el tiempo, dado que esta conclusión se tomó con base

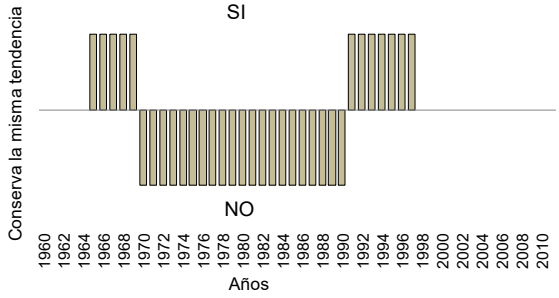
en la visualización de la línea de tendencia de las gráficas. Sin embargo, con los resultados anteriores se demuestra que era simplemente un efecto visual o un error en la determinación de la línea de tendencia.

8.4 WILCONXON SUMA DE RANGOS

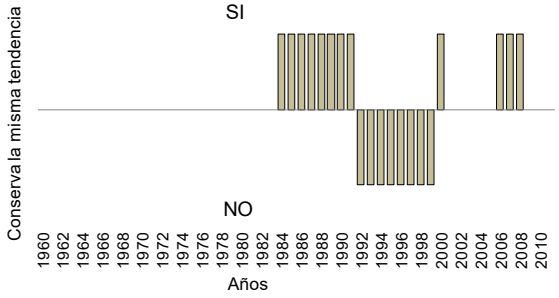
Para concluir el análisis estadístico de las series de caudales, se aplicó una última prueba que permite comparar varios segmentos de la serie y determinar si las medias de estos segmentos conservan la misma tendencia.

Para el estudio que se está realizando se aplicó esta metodología a los caudales medios anuales de cada estación. La manera en que se decidió partir las series es muy peculiar y permite lograr un análisis más profundo de los datos, ya que la seria fue segmentada en cada uno de los años, con el objetivo de determinar si la serie presenta cambio en la media entre los caudales anuales anteriores y posteriores a este año, finalmente se obtuvo un indicador para todos y cada uno de los años de registro que define que si la serie es partida en dicho año presenta cambios en la media entre los segmentos.

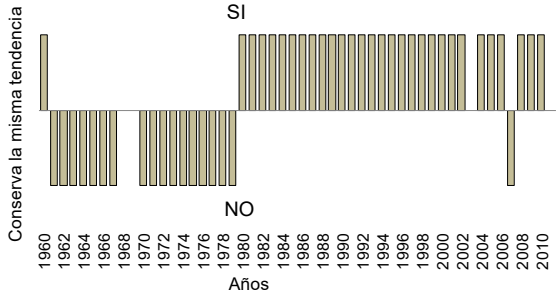
En la Figura 8—7 se presentan estos indicadores para cada uno de los años de registro y cada una de las estaciones, donde si el indicador es SI, quiere decir que si la serie es partida en este año los caudales anteriores y posteriores conserva la misma tendencia, o por el contrario, si el indicador es NO indica que entre los caudales anteriores y posteriores a este año no se conserva la misma tendencia en la medias.



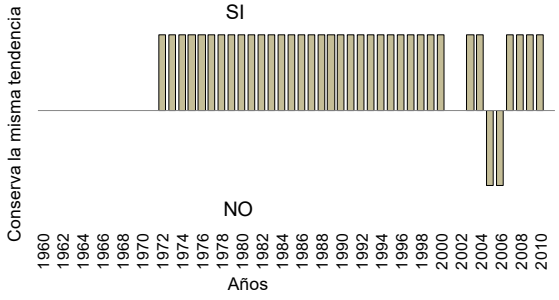
La Pintada



Pescadero



Valdivia



Apavi

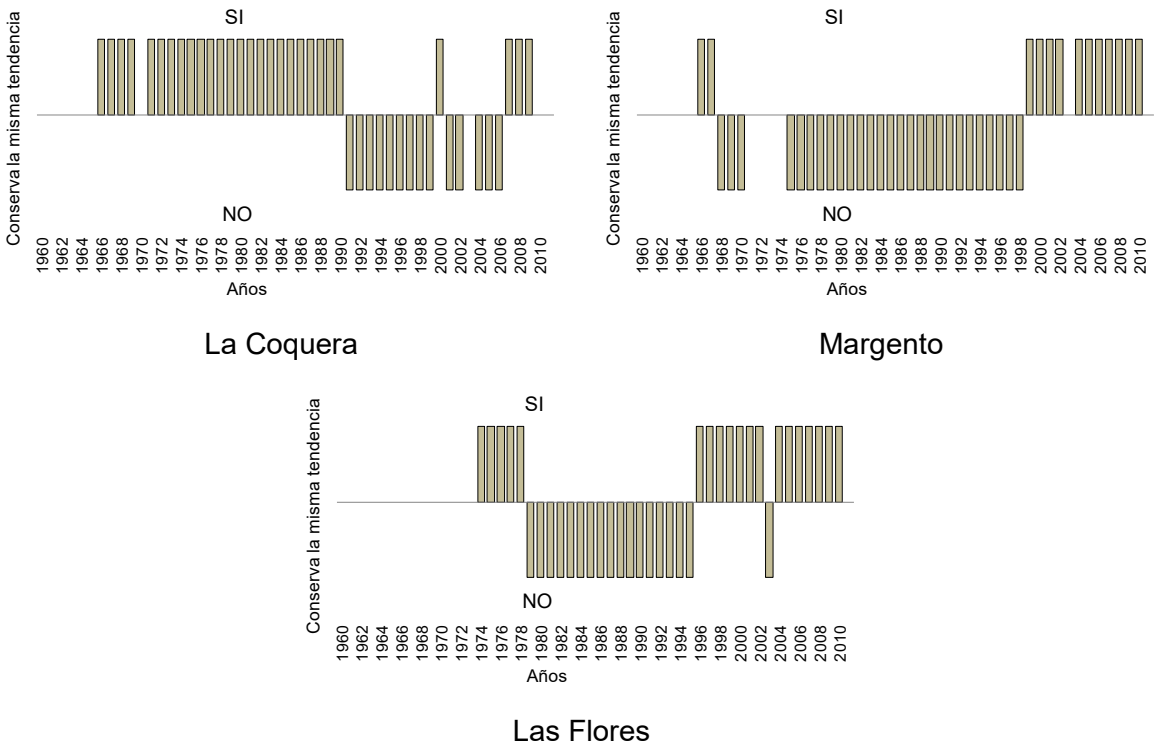


Figura 8—7 Resultados Wilcoxon suma de rangos

Con base en los resultados anteriores se pudo observar, que la estación que menos presenta cambios en la tendencia de la media es la estación Apaví, lo cual es coherente con los resultados en el análisis de media fija cada 2 años para esta estación, donde aparentemente la línea de tendencia se estabilizaba, es decir, al parecer los caudales medios de esta estación no aumentaban ni disminuían con el tiempo.

En ninguno de los años se presentan que para todas las estaciones, hay cambios en la tendencia de las medias, máximo se pudo identificar que entre los años 1992 y 1995, cuatro de las 7 estaciones presentan cambio en la tendencia de la media al igual que en 1979.

9 FRECUENCIA CARACTERÍSTICAS EN LAS ESTACIONES SELECCIONADAS

En el marco teórico se indicó el procedimiento para transformar una señal que está en el dominio del tiempo al dominio de la frecuencia, este procedimiento se logra gracias a la transformada rápida de Fourier (FFT), la cual como se mencionó anteriormente al transformar la serie de caudales, entrega N valores complejos, es decir, un componente real y uno imaginario. La raíz cuadrada de la suma de los cuadrados de estos componentes permite obtener la amplitud de Fourier.

Al graficar cada una de estas amplitudes con su respectiva frecuencia se obtiene el espectro de Fourier el cual permite identificar las frecuencias características o dominantes de las series de caudales de las estaciones elegidas para el presente estudio.

Es importante mencionar que la Transformada Rápida de Fourier solo trabaja con series continuas, por lo cual se seleccionaron los dos periodos continuos más largos de cada una de las estaciones, debido a que las series de caudales presentan vacios, los cuales no permiten que esta metodología sea aplicada a la totalidad de la serie.

Se tomó la determinación de trabajar con los dos periodos más largos ya que para poder utilizar la serie completa de cada una de las estaciones, estas series tendrían que ser rellenadas con base en alguna metodología distorsionando así la información medida en campo, lo cual no se consideró adecuado ya que lo que se pretende es validar la aplicabilidad de la transformada rápida de Fourier en series de caudales.

A continuación, en la Tabla 9-1, se presentan los dos periodos más largos seleccionados para las estaciones.

Tabla 9-1 Periodos continuos con mayor longitud de registro para las estaciones de la zona de estudio

Estación	Periodo	Fecha						Cantidad de datos
		Inicio			Fin			
		Año	Mes	Día	Año	Mes	Día	
La Pintada	1	1966	diciembre	6	1990	julio	30	8638
	2	1990	agosto	1	1998	junio	22	2883
Pescadero	1	1990	marzo	27	1996	septiembre	11	2361
	2	1996	octubre	1	2000	diciembre	31	1553
Valdivia	1	1961	septiembre	1	1967	septiembre	21	2212
	2	1970	abril	1	1987	febrero	26	6176
Apaví	1	1972	junio	1	1987	febrero	28	5386
	2	1988	octubre	20	2000	diciembre	31	4456
La Coquera	1	1971	abril	1	1979	julio	31	3044
	2	1979	octubre	1	2004	abril	29	8613
Margento	1	1983	julio	1	1989	octubre	29	2313
	2	1997	septiembre	1	2002	abril	22	1695
Las Flores	1	1983	octubre	1	1987	febrero	28	1247
	2	1987	abril	1	1991	febrero	28	1430

Con base en las series continuas indicadas anteriormente se aplicó la transformada rápida de Fourier a cada periodo obteniendo de esta manera el espectro de Fourier de cada uno de estos en las estaciones, es decir, se obtuvieron dos espectros de Fourier por cada estación, ya que como se mencionó anteriormente se seleccionaron los dos periodos más largos de cada una de estas.

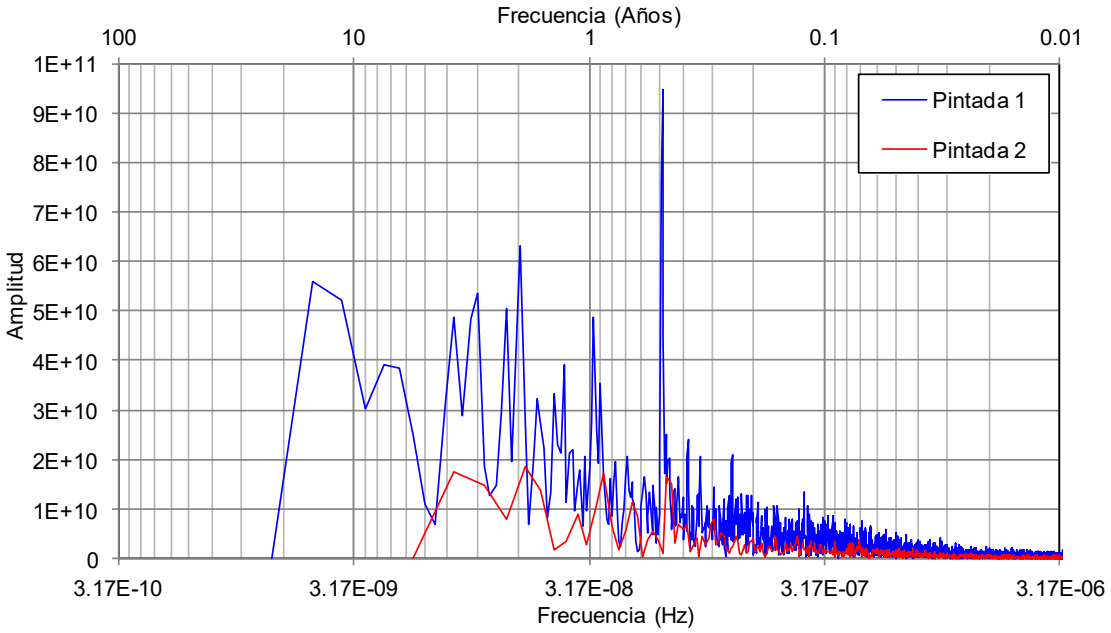


Figura 9—1 Espectro de Fourier estación Pintada

Como lo muestra la Figura 9—1, para la estación Pintada en el primer periodo más largo, en el espectro de Fourier, se logra identificar claramente cuatro frecuencias características o dominantes, donde las amplitudes de Fourier son las máximas. Estas frecuencias corresponden a periodos de 14.96 ,1.95, 0.95y 0.49 años. Igualmente para el segundo periodo más largo las frecuencias características corresponden a periodos de 3.74, 1.87, 0.86 y 0.47 años.

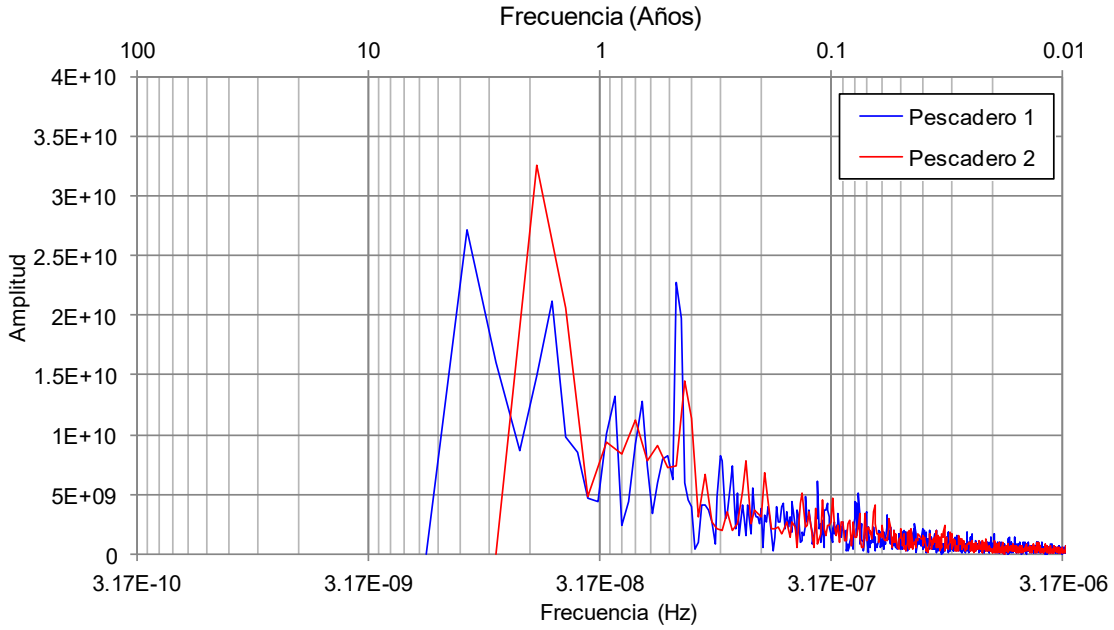


Figura 9—2 Espectro de Fourier estación Pescadero

Como lo muestra la Figura 9—2, para la estación Pescadero en el primer periodo más largo, en el espectro de Fourier, se logra identificar claramente cuatro frecuencias características o dominantes, donde las amplitudes de Fourier son las máximas. Estas frecuencias corresponden a periodos de 3.74 ,1.6, 0.86 y 0.47 años. Sin embargo, para el segundo periodo más largo solo se logra identificar tres frecuencias características, la cuales corresponden a periodos de 1.87, 0.7 y 0.43 años.

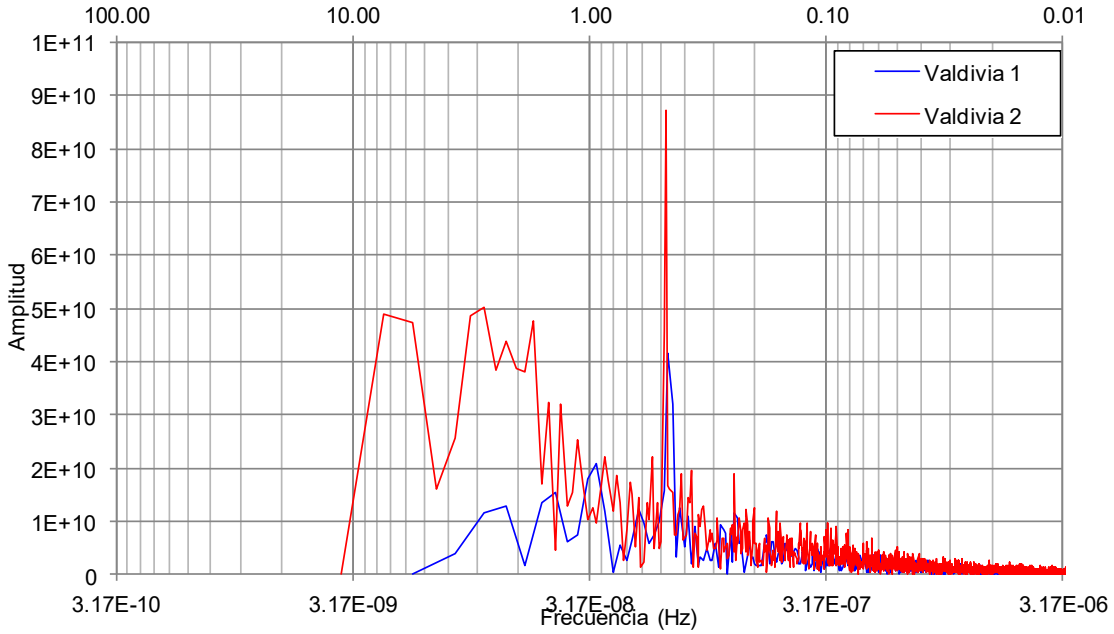


Figura 9—3 Espectro de Fourier estación Valdivia

Como lo muestra la Figura 9—3, para la estación Valdivia en el primer periodo más largo, en el espectro de Fourier, se identificaron tres frecuencias dominantes y características, donde las amplitudes de Fourier son las máximas. Estas frecuencias corresponden a periodos de 2.24, 0.93 y 0.47 años. Sin embargo, para el segundo periodo más largo se logra identificar nuevamente cuatro frecuencias características, la cuales corresponden a periodos de 7.48, 2.81, 1.73 y 0.48 años.

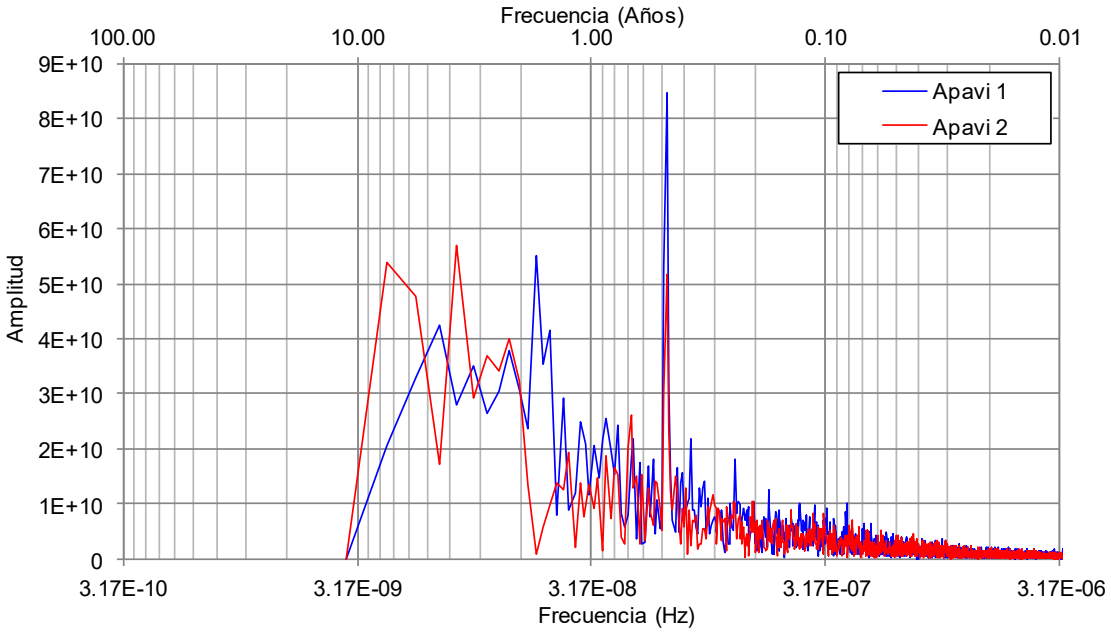


Figura 9—4 Espectro de Fourier estación Apaví

Como lo muestra la Figura 9—4, para la estación Apaví en el primer periodo más largo, en el espectro de Fourier, se logra identificar claramente cuatro frecuencias características o dominantes, donde las amplitudes de Fourier son las máximas. Estas frecuencias corresponden a periodos de 4.49 ,1.73, 1.5 y 0.48 años. Igualmente para el segundo periodo más largo las frecuencias características corresponden a periodos de 7.48, 3.74, 2.24 y 0.48 años.

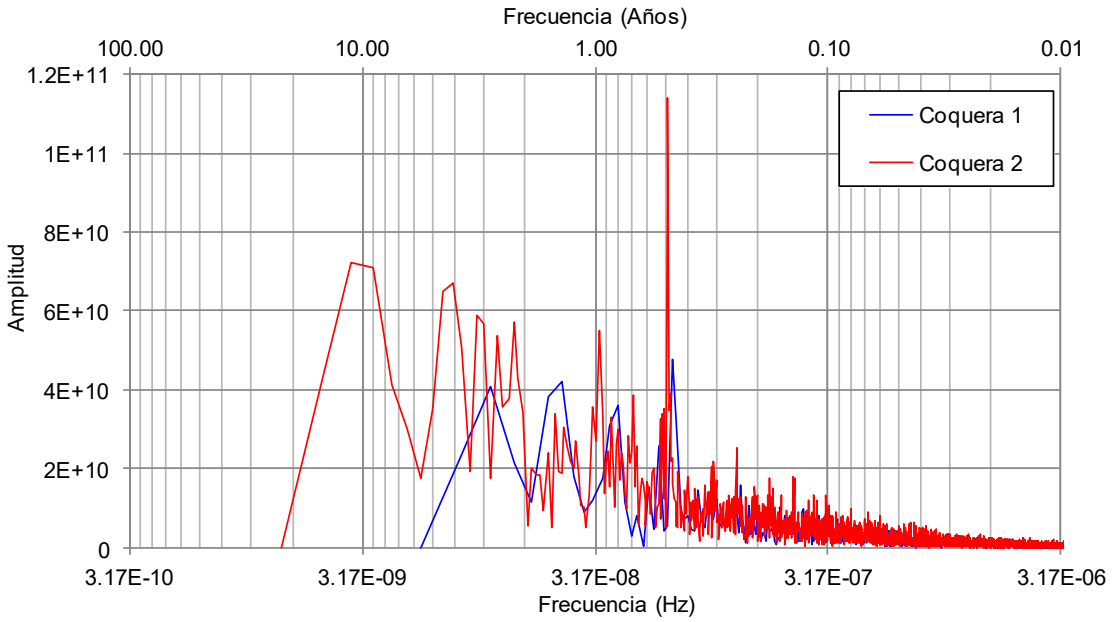


Figura 9—5 Espectro de Fourier estación La Coquera

Como lo muestra la Figura 9—5, para la estación La Coquera en el primer periodo más largo, en el espectro de Fourier, se pudieron identificar cuatro frecuencias características o dominantes, donde las amplitudes de Fourier son las máximas. Estas frecuencias corresponden a periodos de 2.8 ,1.4, 0.80 y 0.47 años.

Igualmente, para el segundo periodo más largo las frecuencias características corresponden a periodos de 11.22, 4.08, 0.95 y 0.49 años.

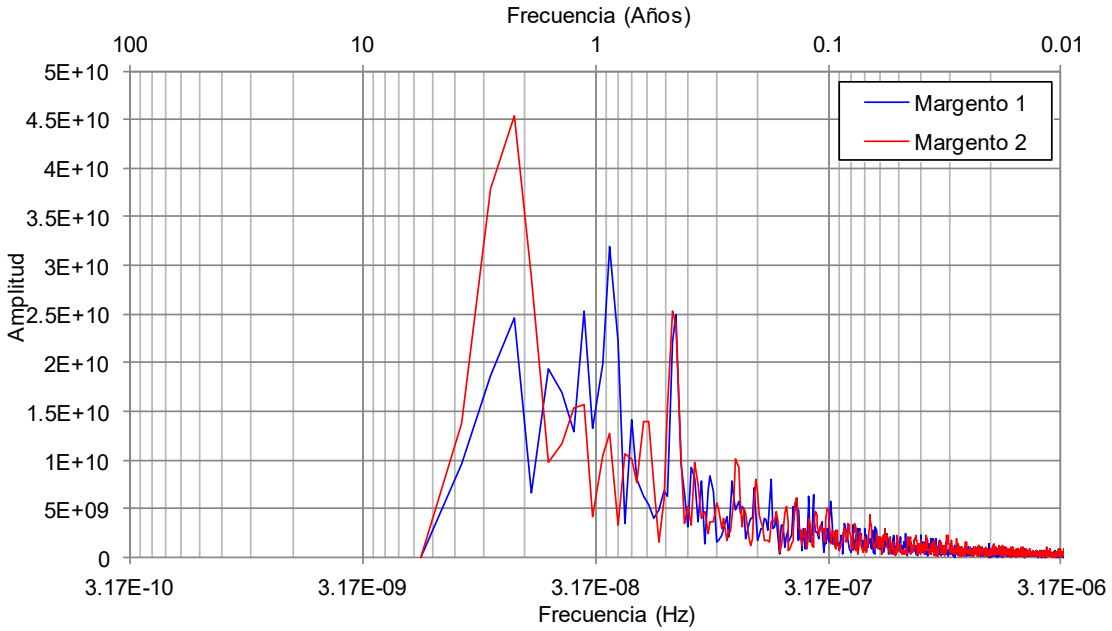


Figura 9—6 Espectro de Fourier estación Margento

Como lo muestra la Figura 9—6, para la estación Margento en el primer periodo más largo, en el espectro de Fourier, se logró identificar con claridad cuatro frecuencias características o dominantes, donde las amplitudes de Fourier son las máximas. Estas frecuencias corresponden a periodos de 2.24 ,1.12, 0.86 y 0.45 años.

De igual forma, para el segundo periodo más largo las frecuencias características corresponden a periodos de 2.24, 1.12, 0.63 y 0.47 años.

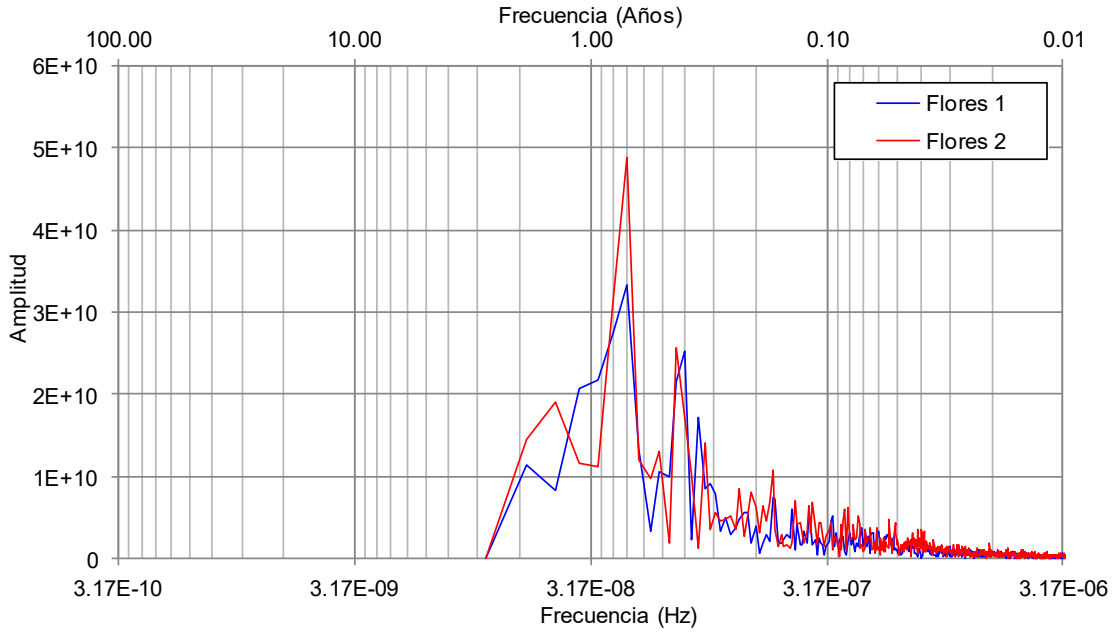


Figura 9—7 Espectro de Fourier estación Las Flores

Como lo muestra la Figura 9—7, para la estación Las Flores en el primer periodo más largo, en el espectro de Fourier, se lograron identificar claramente cuatro frecuencias características o dominantes, donde las amplitudes de Fourier son las máximas.

Estas frecuencias, corresponden a periodos de 1.87 ,0.70, 0.40 y 0.35 años. Igualmente para el segundo periodo más largo las frecuencias características corresponden a periodos de 1.4, 0.7, 0.43 y 0.33 años.

A continuación, en la Tabla 9-2, se presentan las frecuencias características identificadas en cada una de las estaciones para cada uno de los periodos.

Tabla 9-2 Frecuencias características de las estaciones analizadas

Estación	Periodo	Frecuencias características (Hz)				Periodo en años			
		1	2	3	4	1	2	3	4
Pintada	**1**	2.12E-09	1.62E-08	3.32E-08	6.50E-08	14.96	1.95	0.95	0.49
	2	8.48E-09	1.70E-08	3.67E-08	6.78E-08	3.74	1.87	0.86	0.47
Pescadero	**1**	8.48E-09	1.98E-08	3.67E-08	6.78E-08	3.74	1.60	0.86	0.47
	2		1.70E-08	4.52E-08	7.35E-08		1.87	0.70	0.43
Valdivia	**1**		1.41E-08	3.39E-08	6.78E-08		2.24	0.93	0.47
	2	4.24E-09	1.13E-08	1.84E-08	6.64E-08	7.48	2.81	1.73	0.48
Apavi	**1**	7.07E-09	1.84E-08	2.12E-08	6.64E-08	4.49	1.73	1.50	0.48
	2	4.24E-09	8.48E-09	1.41E-08	6.64E-08	7.48	3.74	2.24	0.48
La Coquera	**1**	1.13E-08	2.26E-08	3.96E-08	6.78E-08	2.80	1.40	0.80	0.47
	2	2.83E-09	7.77E-09	3.32E-08	6.50E-08	11.22	4.08	0.95	0.49
Margento	**1**	1.41E-08	2.83E-08	3.67E-08	7.07E-08	2.24	1.12	0.86	0.45
	2	1.41E-08	2.83E-08	5.04E-08	6.78E-08	2.24	1.12	0.63	0.47
Las Flores	**1**	1.70E-08	4.52E-08	7.92E-08	9.04E-08	1.87	0.70	0.40	0.35
	2	2.26E-08	4.52E-08	7.35E-08	9.61E-08	1.40	0.70	0.43	0.33

Analizando los valores recopilados en la Tabla 9-2, se logró identificar que existe una dependencia entre la longitud de los registros y las frecuencias características que se logran visualizar, esto es claro en periodos con longitud de registros alta como son La Pintada 1 y La Coquera 2 con alrededor de 8600 datos.

Analizando el espectro de Fourier de estos periodos de registro, se obtienen las frecuencia más bajas como son 14.96 y 11.22 años respectivamente. También es notable que para periodos con cantidad de registros más cortos como las Flores 1 y 2 con aproximadamente 1300 datos, las frecuencias son más altas.

Por otra parte, se observó que existen frecuencias que se logran identificar en los espectros de Fourier en la gran mayoría de los periodos de registro, como las frecuencias que se sitúan alrededor de 0.45 años, siendo estas las más concurrentes en todos los periodos de registro.

También se determinó la predominancia de frecuencias que oscilan entre 1.5 y 2 año, finalmente, con menos predominio que las anteriores se observaron frecuencias entre 3,7 y 4 años. Estos 3 intervalos de frecuencias identificados como dominantes en la mayoría de las estaciones, como se muestra en la Figura 9—8, se consideran como las frecuencias que definen el comportamiento del rio Cauca en el tramo de estudio.

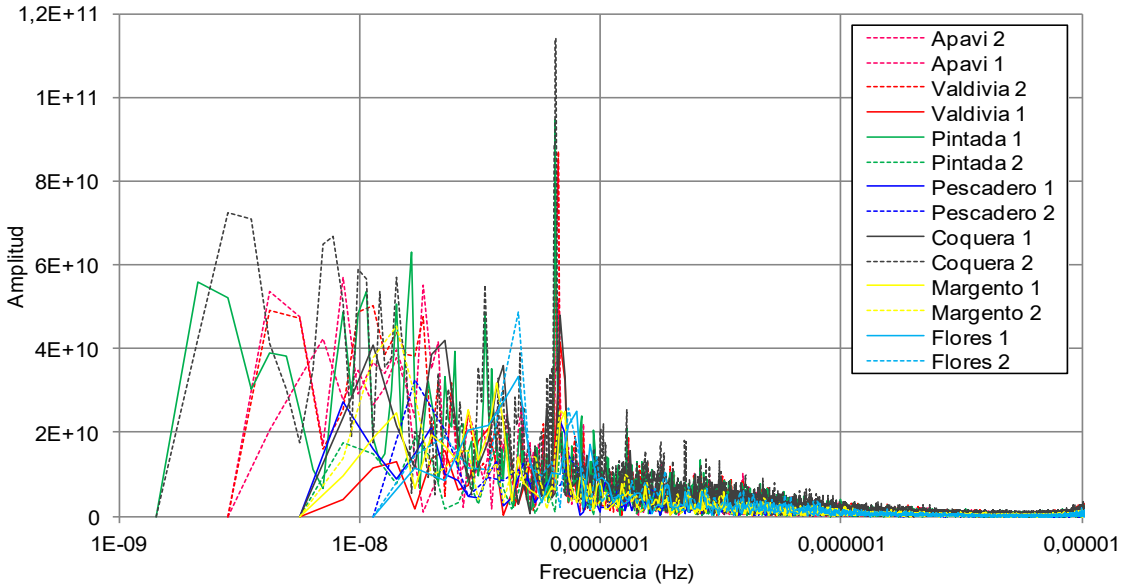


Figura 9—8 Espectros de Fourier de las estaciones analizadas

10 FENÓMENOS MACRO CLIMÁTICOS DE LA ZONA DE ESTUDIO

Para el estudio del comportamiento de los fenómenos macro climáticos y sus efectos, se han creado algunos índices estandarizados, los cuales proporcionan una medida aproximada de la situación climática global, con base en las mediciones de un conjunto de variables climáticas en diferentes lugares del planeta. Para le presente proyecto se trabajaron con 3 de estos índices, los cuales son el índice Multivariado del ENSO (MEI), índice Oceánico del Niño (ONI) y oscilación de Madden-Julian (MJO), los cuales son índices que representan fenómenos macro climáticos que tiene influencia en nuestro país.

Estos índices se analizan con el objetivo de validar la influencia de estos fenómenos y la aplicabilidad del modelo matemático de la FFT en la cuenca andina del río Cauca.

A continuación se describen cada uno de estos índices y se presentan gráficamente sus registros.

10.1.1 Índice Multivariado del ENSO (MEI)

El MEI o Índice Multivariado del ENSO (Wolter y Timlin, 1993) se puede deducir como un índice promedio de las características principales del ENSO contenidas en las siguientes seis variables: presión a nivel del mar, componentes zonal y meridional de los vientos superficiales, temperatura superficial del mar, temperatura del aire superficial y cantidad total de nubosidad.

El MEI se calcula por separado para cada uno de los doce pares de meses (Dic/Ene, Ene/Feb) y a partir del comportamiento de las seis variables observadas combinadas. Esto se realiza normalizando la variación total de cada variable y luego realizando la extracción de la componente principal en la matriz de covarianza de las variables combinadas.

Los valores negativos del MEI representan la fase fría de ENSO, correspondiente a La Niña, mientras que los valores positivos de MEI representan la fase caliente de ENSO o EL Niño. En la Figura 10—1 se muestra el comportamiento del ENSO desde al año 1950 hasta el año 2013, mediante el Índice Multivariado del ENSO (MEI).

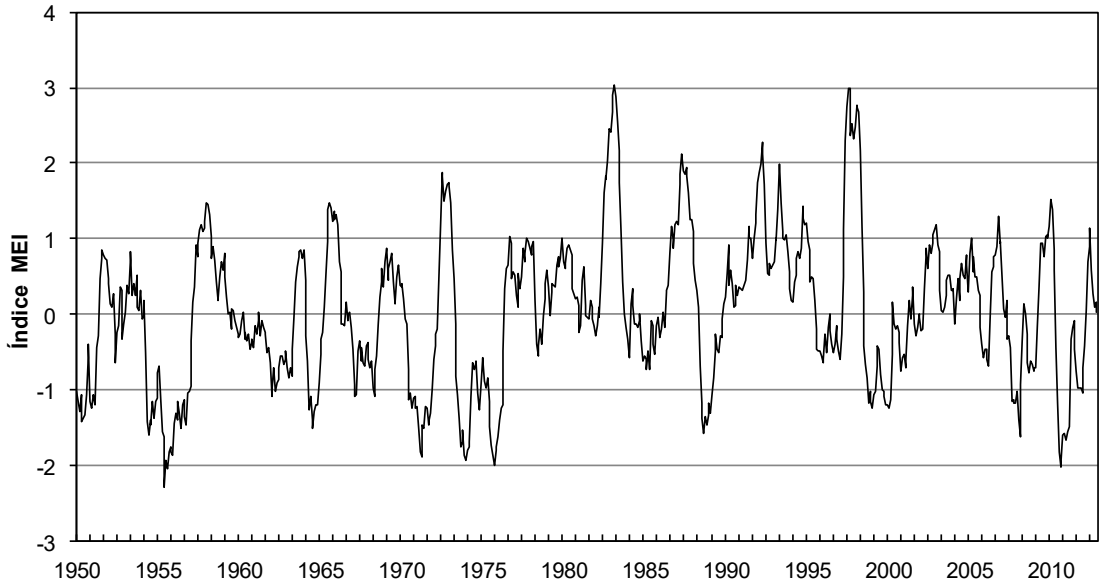


Figura 10—1 Índice MEI (NOAA)

10.1.2 Índice Oceánico del Niño (ONI)

El ONI o Índice Oceánico del Niño se obtiene a partir de valores estandarizados de SST, en la región niño 3.4. Las temperaturas por encima o por debajo de la media se denominan “anomalías”. El ONI se calcula promediando los tres últimos meses de SST; para valores +0.5 durante 5 o más valores trimestrales consecutivos se considera la condición “cálida”. La Figura 10—2 muestra la serie histórica del Índice ONI.

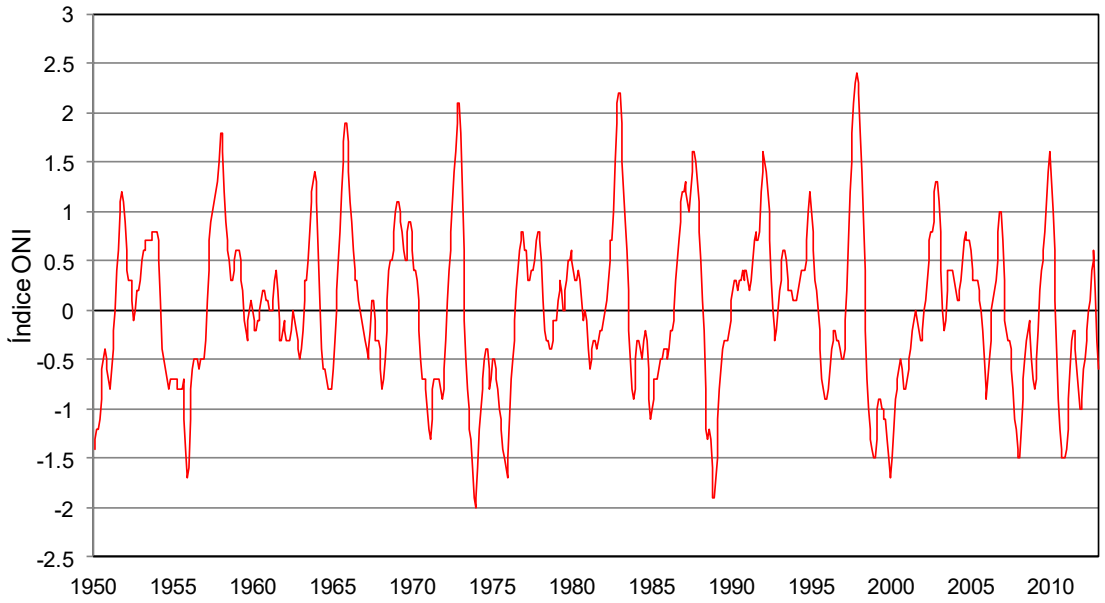


Figura 10—2 Índice ONI (NOAA)

10.1.3 Oscilación de Madden-Julian (MJO)

La oscilación de Madden-Julian, también conocida como la oscilación de 30-60 días, oscilación de 40-50 días, u oscilación intra-estacional, es la principal fluctuación anual que explica la variación climática en los trópicos, afecta toda la troposfera tropical pero se hace aún más evidente en el océano Indico, así como en el oeste del océano Pacífico.

Esta oscilación implica variaciones en la magnitud y dirección de la velocidad del viento, temperaturas superficiales del mar, precipitación y formación de nubes, y es más fuerte aún su influencia sobre la radiación de onda larga reflejada en la tierra.

La oscilación de Madden-Julian interactúa tanto con la intensidad y periodicidad de los monzones australianos y asiáticos como con el fenómeno del Niño, esto le da la relevancia suficiente para considerar su efecto en el modelo de pronóstico aquí planteado (NOAA).

Hay fuertes evidencias de los efectos de la fase de la oscilación intraestacional en la hidro-climatología de Colombia (Arias, 2005). La Figura 10—3 muestra la serie histórica del Índice MJO.

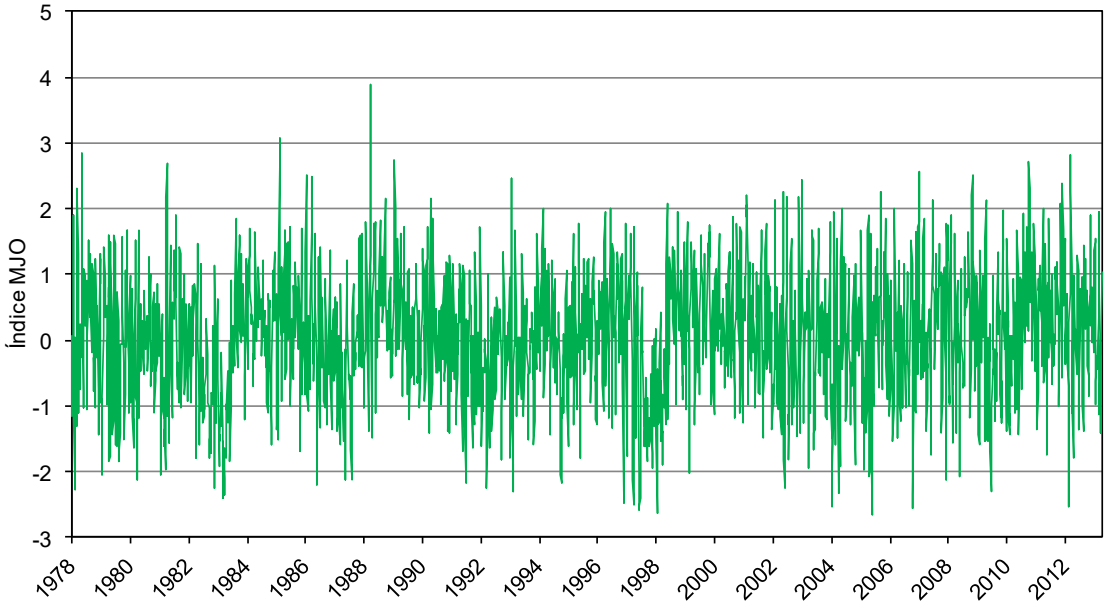


Figura 10—3 Índice MJO (NOAA)

11 VALIDACIÓN DEL MODELO DE LA FFT EN LA ZONA DE ESTUDIO

Con el objetivo de validar que la Transformada Rápida de Fourier puede representar de manera simple y precisa los fenómenos macroclimáticos que tienen influencia en la cuenca andina del rio Cauca, se procedió a comparar la frecuencias características de cada uno de los índices climáticos presentados en el capitulo anterior con la frecuencias dominantes que se encontraron en las estaciones de medición de caudal.

Adicionalmente, se encontraron algunos fenómenos que no tienen un índice claro de medición, pero que posiblemente tienen una influencia muy clara en los caudales analizados.

A continuación en la Figura 11—1, Figura 11—2 y Figura 11—3 se presentan las frecuencias características de cada uno de los fenómenos, comparadas con las frecuencias características de la serie de La Pintada 1 la cual se seleccionó para este análisis, dado que es una estación que representa de manera clara las frecuencias dominantes que se identificaron en la gran mayoría de estaciones.

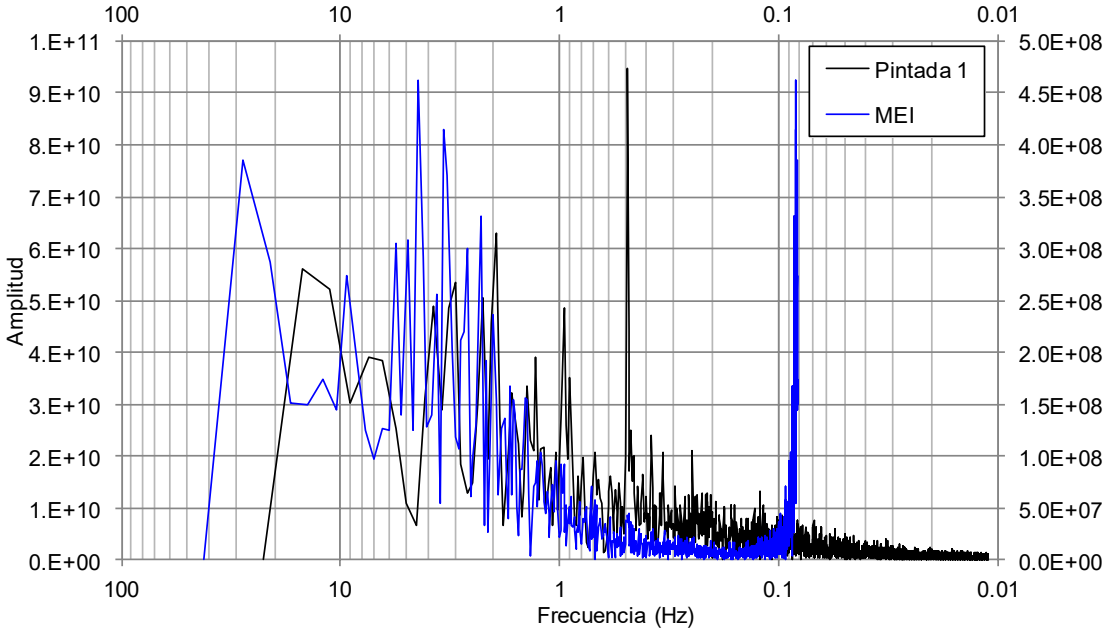


Figura 11—1 Frecuencias características del MEI con la Pintada 1

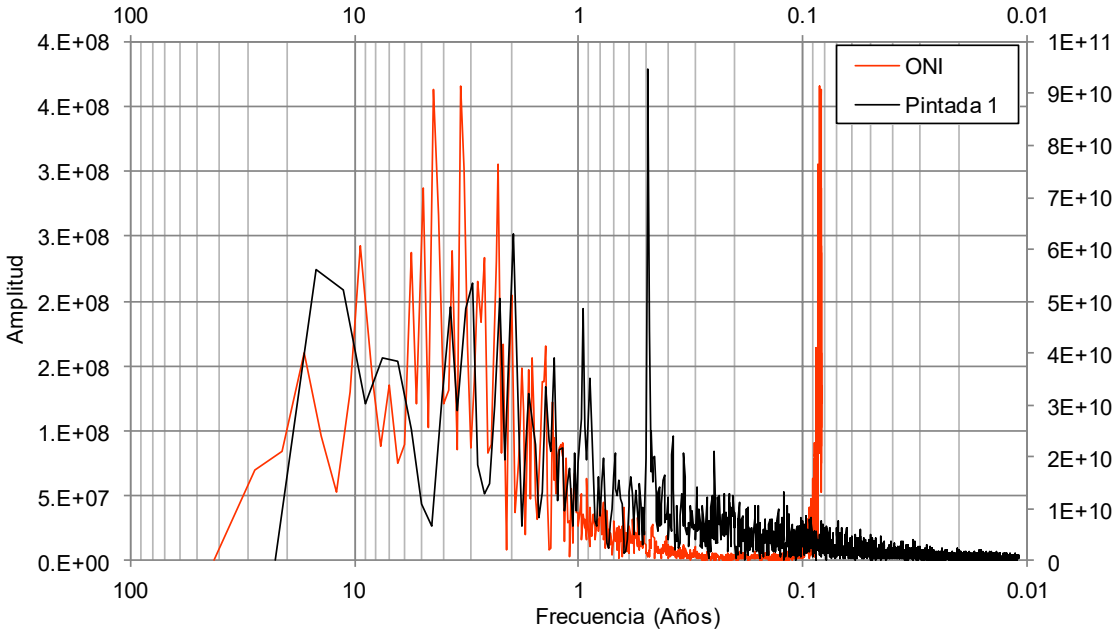


Figura 11—2 Frecuencias características del ONI con la Pintada 1

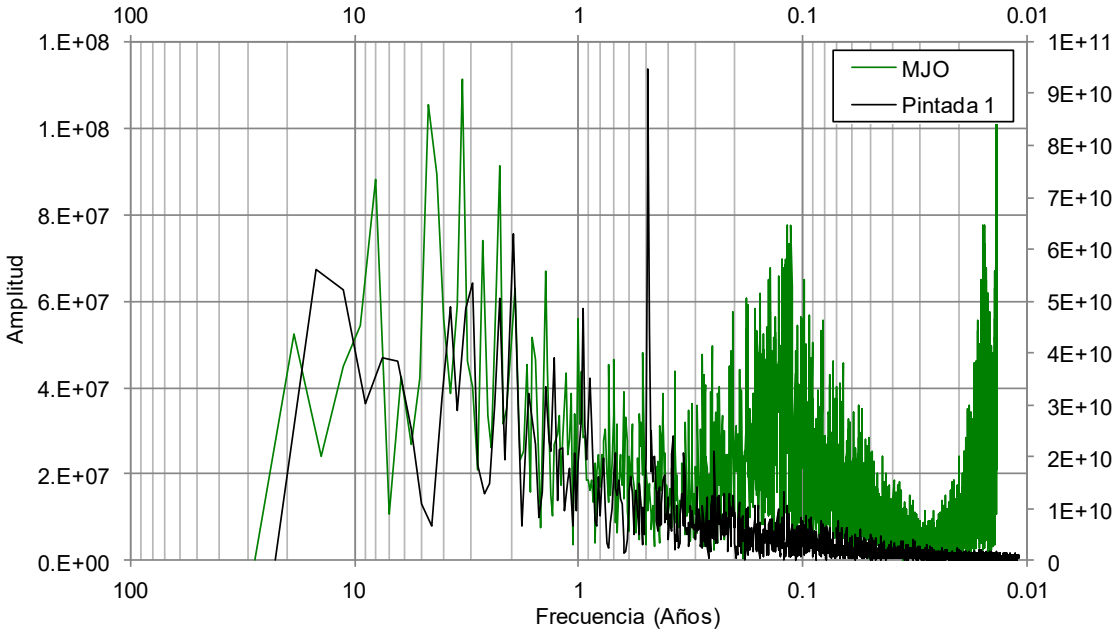


Figura 11—3 Frecuencias características del MJO con la Pintada 1

En las figuras anteriores es fácil identificar que los índices climáticos presenta un pico dominante que coincide con un pico característico en la mayoría de las estaciones analizadas, este pico se detecta con un frecuencia entre 1,9 y 2,3 años, como se muestra en la Figura 11—4 y Figura 11—5, donde se muestra claramente estas frecuencias tanto en las estaciones como en los índices climáticos respectivamente.

Estas frecuencias podrían ser asociadas a la llamada oscilación cuasi-bienal (QBO), fenómeno que tiene una periodicidad entre 24 y 30 meses (2 y 2,5 años). Algunos de los aspectos más relevantes, y ampliamente conocidos de la QBO son: (Holton, 1972)

- La oscilación cuasi-bienal (QBO) es un fenómeno atmosférico que presenta una periodicidad casi perfecta, se trata de una oscilación en la dirección de los vientos zonales y en la temperatura, esto es, una inversión regular de los vientos en la estratosfera sobre el ecuador.
- Los vientos zonales presentan un patrón alternante entre dirección oeste y este, con una periodicidad alrededor de 24-30 meses.
- Los regímenes sucesivos aparecen primero por encima de los 30 km (~10 hPa) pero se propagan hacia niveles inferiores de la atmósfera a una velocidad de 1 km/mes.
- La propagación hacia niveles bajos ocurre sin perder la amplitud entre los 30 y 23 km, pero decrece por debajo de los 23 km.
- La oscilación es simétrica respecto al ecuador con una amplitud máxima de alrededor 20 m/s

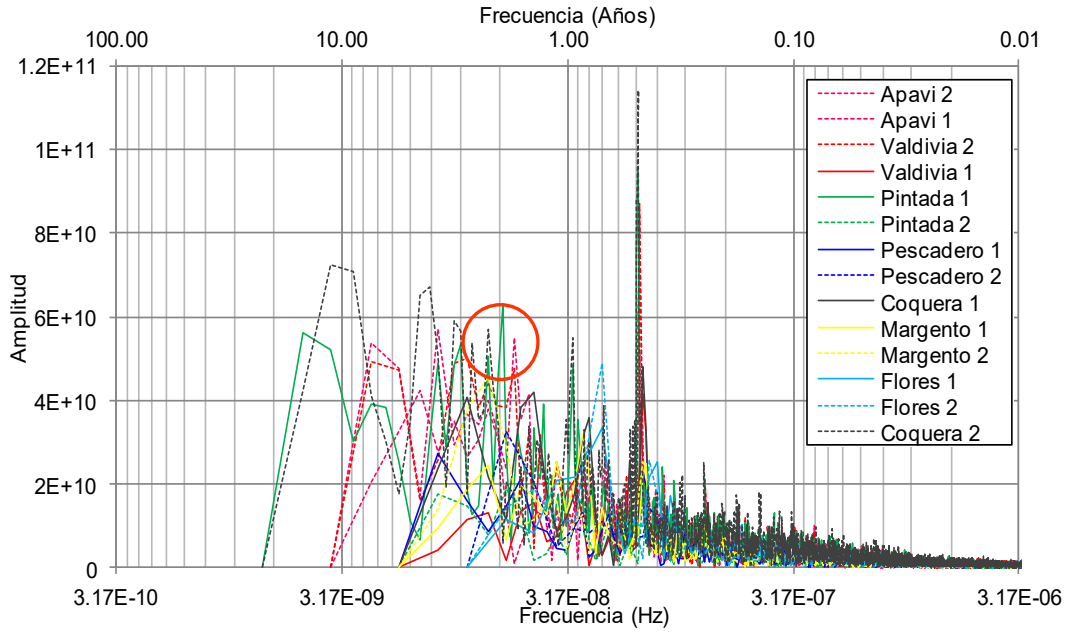


Figura 11—4 Frecuencia característica entre 1,9 y 2,3 años identificada en las estaciones

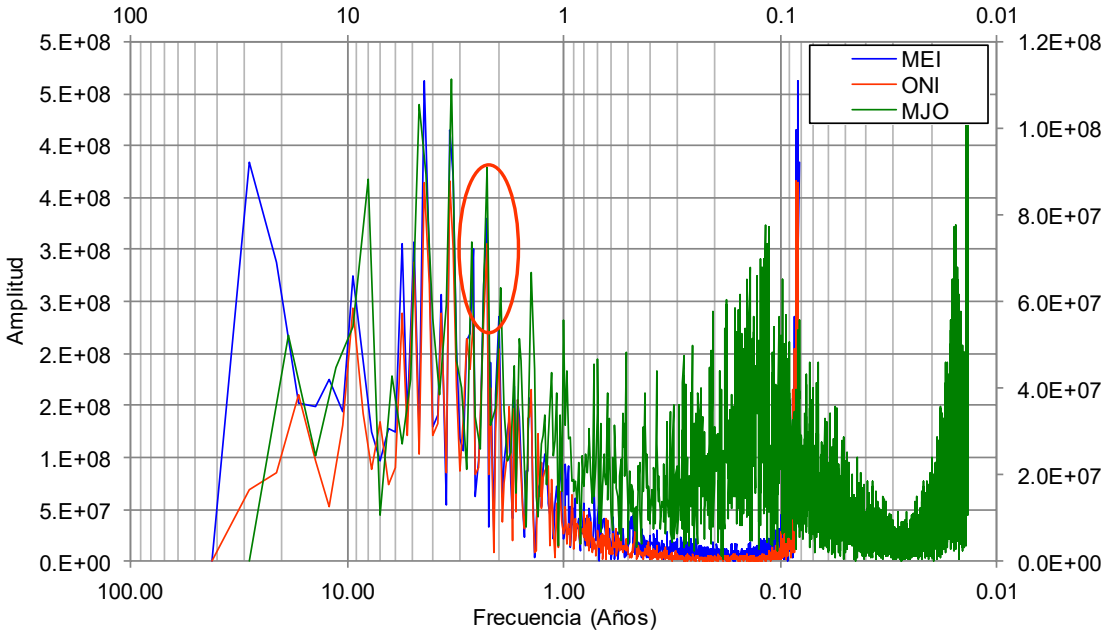


Figura 11—5 Frecuencia característica entre 1,9 y 2,3 años identificada en los índices climáticos

Por otra parte, se observó que existen frecuencias características entre los índices climáticos los cuales están inversamente relacionados, es decir, las estaciones presentan amplitudes bajas en las frecuencias entre 5 y 6 años, mientras que estas mismas frecuencias los índices climáticos presenta picos característico con una alta amplitud, como se muestra en la Figura 11—6 y la Figura 11—7.

Estas frecuencias pueden estar asociadas a las anomalías hidroclimáticas durante las dos fases del fenómeno El Niño/Oscilación del Sur (ENSO). Dado que en Sur América tropical incluyendo Colombia se presentan anomalías negativas en la precipitación, y en los caudales de los ríos durante El Niño, y generalmente lo contrario durante La Niña, lo cual se observa claramente la Figura 11—6 y la Figura 11—7.

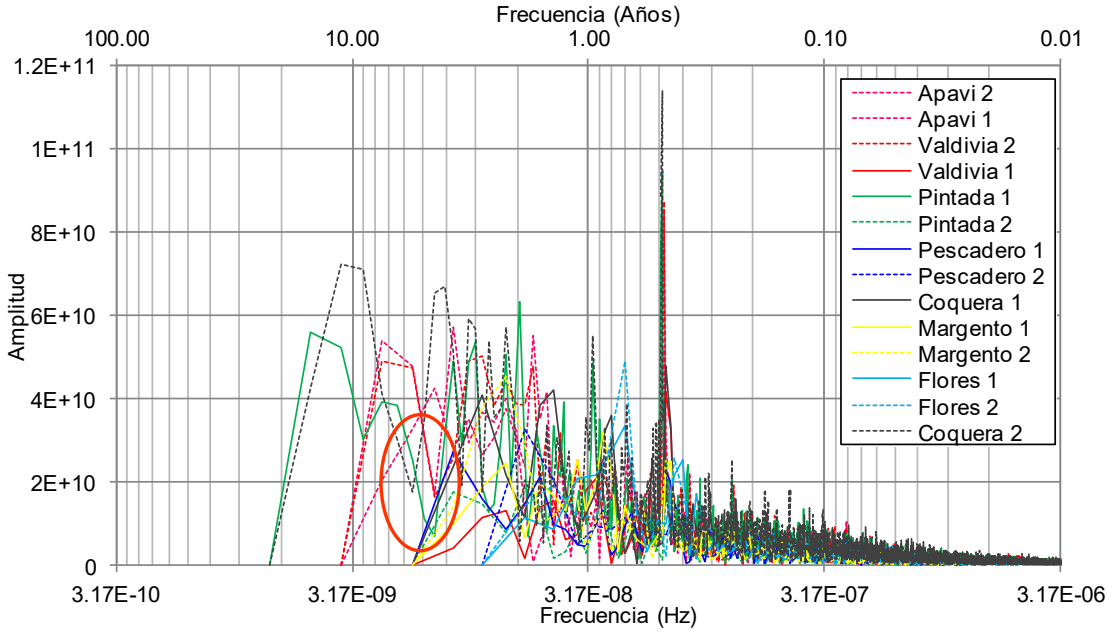


Figura 11—6 Frecuencia característica entre 5 y 6 años identificada en las estaciones

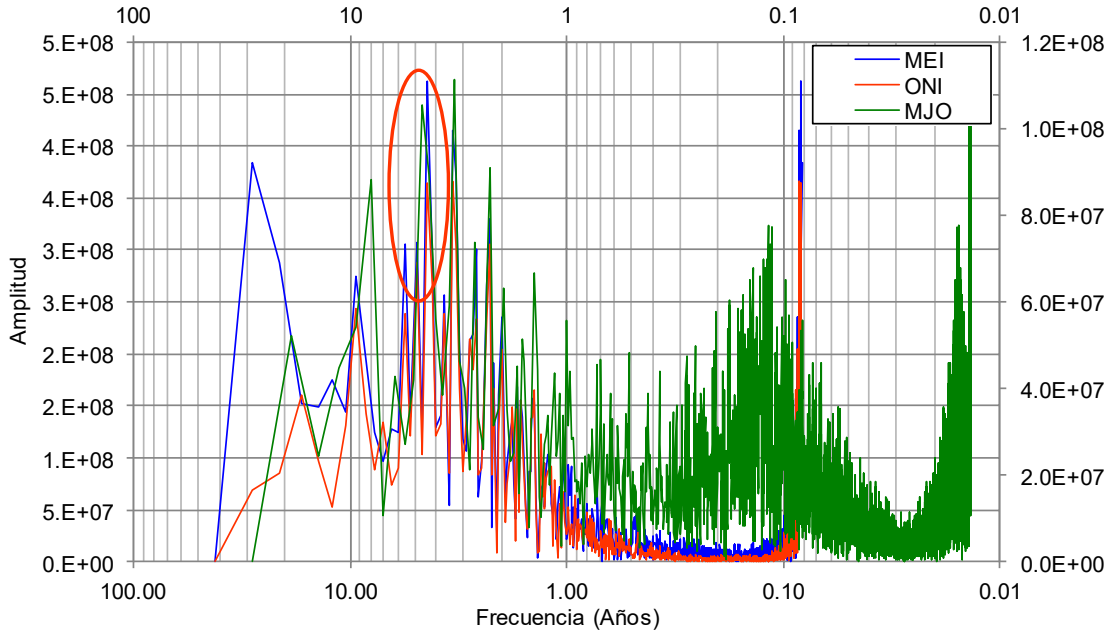


Figura 11—7 Frecuencia característica entre 5 y 6 años identificada en los índices climáticos

Finalmente, si hizo evidente que ninguno de los índices climáticos analizados, presenta frecuencias dominantes de aproximadamente 0,5 años (6 meses), esta frecuencia es supremamente recurrente en toda las estaciones, como se muestra en la Figura 11—8.

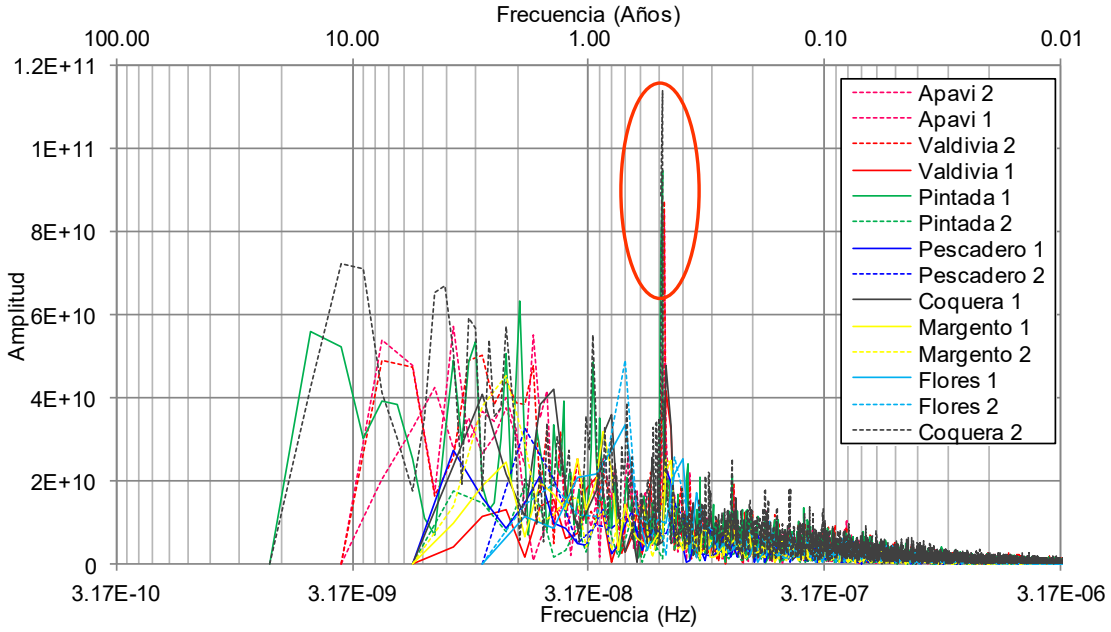


Figura 11—8 Frecuencia característica de 0,5 años identificada en las estaciones

Esta frecuencia dominante de 0,5 años está asociada a la oscilación meridional de la Zona de Convergencia Intertropical (ZCIT), La cual tiene un recurrencia de 6 meses, es unos de los fenómenos que más influencia tiene en nuestro país, generando un clima de características bimodales, es decir, generando dos inviernos y dos veranos al año.

Por otro lado, se identificaron frecuencias características entre 14,9 y 11,2 años, de la cuales no se encontró una asociación directa entre los fenómenos climáticos y los registros de las estaciones, el origen de estas frecuencias deben ser estudiadas con más detalle.

12 METODOLOGÍA PARA COMPLETAR SERIES DE CAUDALES

En el anterior capitulo se encontró que la Trasformada Rápida de Fourier representa de manera confiable las influencia de los principales fenómenos macroclimáticos en las series de caudales del río Cauca. En el presente capitulo, con base en esta herramienta matemática se formuló una metodología para completar registros faltantes en las series de caudales.

Para el planteamiento de la metodología de completar registros de caudales medios, a modo de ejemplo, se utilizó la serie de caudales medios diarios registrada en la estación limnigráfica La Coquera; esta estación cuenta con registros históricos de unos 46 años, desde Septiembre de 1966 hasta Diciembre de 2010, con algunos períodos incompletos, los cuales son los que se necesitan completar, entre los que se destacan: Abril 1 de 1969 a Marzo de 1971, y 1 de Enero al 31 de Diciembre de 2003.

En la Figura 12—1 se presenta la serie de caudales del río Cauca en la estación La Coquera.

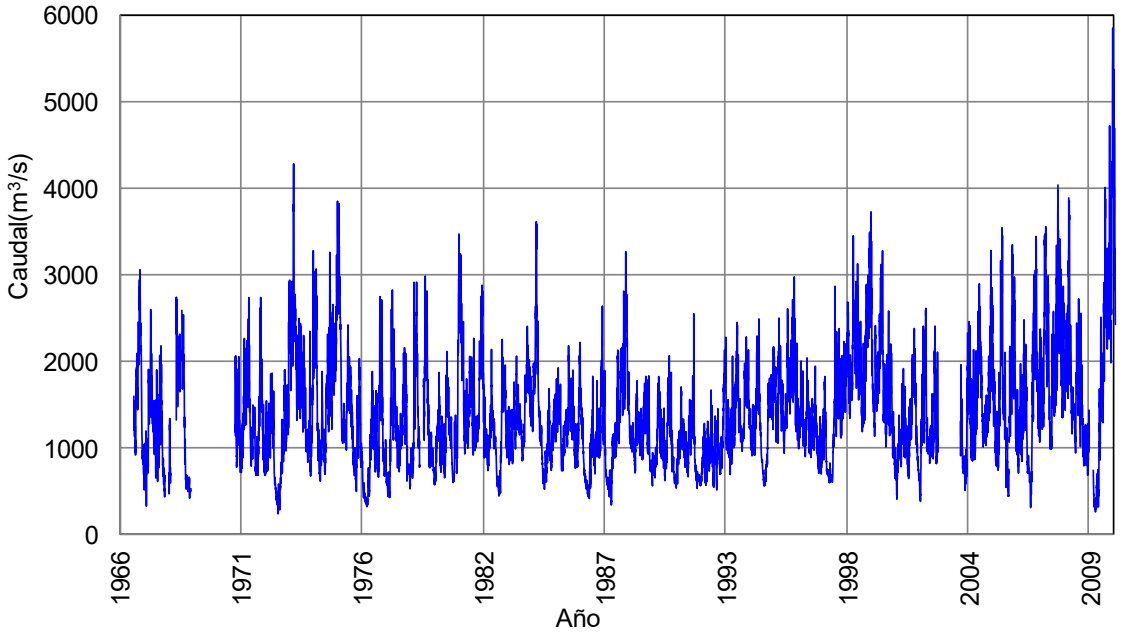


Figura 12—1 Serie original de caudales medios diarios estación La Coquera

Para tener una serie de caudales homogénea en el tiempo, se consideró que la metodología debería tener en cuenta los siguientes componentes para completar la información faltante:

- Análisis con año pivote: Este análisis consistió en seleccionar un año pivote, el cual es un periodo de tiempo que se caracteriza por tener registrado el 100% de los datos, y que además la media durante este periodo, es similar al caudal medio multianual de la serie original.

Para la estación La Coquera se seleccionó a 1986 como año pivote, ya que este año cuenta con el 100% de los registros y su media es de 1265 m^3/s cercana a la media de toda la serie, la cual equivale a 1397 m^3/s como se muestra en la Figura 12—2.

Figura 12—2 Año pivote de 1986

- Análisis con las frecuencias dominantes: Este análisis consistió en tomar las frecuencias características del modelo de la FFT validados con los fenómenos macroclimáticos, y con base en estas frecuencias reconstruir una serie de armónicos que represente el comportamiento de estos fenómenos a través del tiempo.

  En el presente estudio se parte de una señal muestreada, que es una serie hidrológica discreta y periódica, a la cual se le aplica la transformada discreta de Fourier mediante el algoritmo de la FFT, obteniendo una sucesión finita de coeficientes que indican que tan ponderadas están las frecuencias con respecto a los fenómenos macroclimáticos.

  La serie de la FFT está centrada en el ancho de banda de la señal, la cual es la frecuencia de Nyquist, esta serie es una sucesión finita de coeficientes, y pensar en devolverse al dominio temporal con una función continua que aproxime la serie temporal no es matemáticamente viable, porque al involucrar una operación de muestreo, el mapeo ya no sería uno a uno. Sin embargo cualitativamente se puede llegar a una aproximación.

  Asumiendo que las componentes reales e imaginarias obtenidas a partir de la FFT aproximan a los coeficientes a_n y b_n de la serie de Fourier respectivamente, se plantea hacer una analogía con una serie hidrológica continua periódica, la cual genera una función continua con la forma de la serie hidrológica sin ruido; teniendo en cuenta que las amplitudes de esta función no tienen una relación aparente con los valores de caudales, con sus frecuencias se normaliza tanto la serie como la función armónica encontrada, como se presenta en la Figura 12—3.

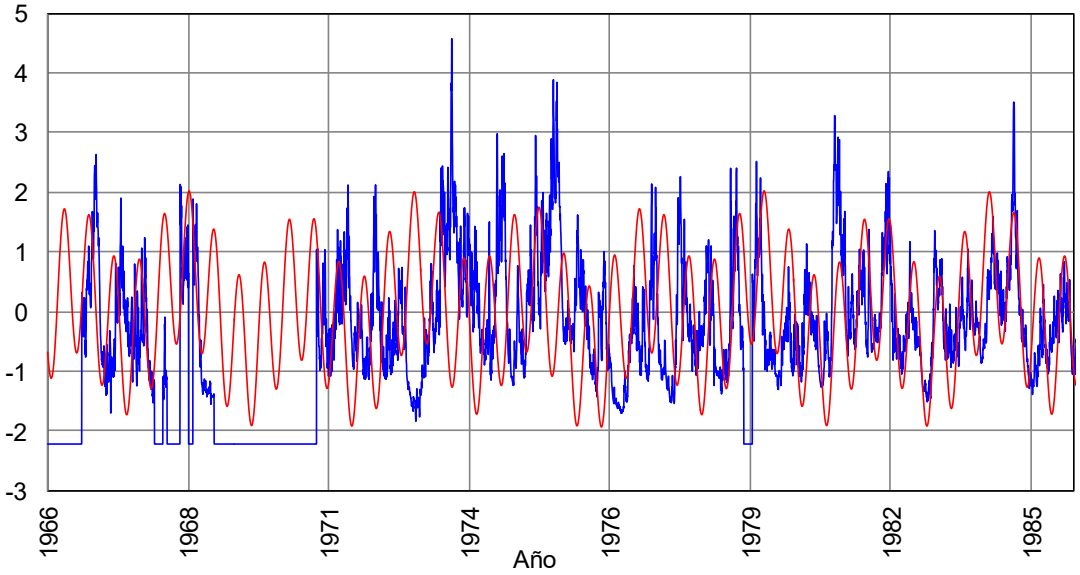


Figura 12—3 Normalización de serie y función de la estación La Coquera

$$Z = \frac{x - \mu}{\sigma}$$

Ecuación 12-1 Ecuación de normalización

Con base en los análisis anteriores, la reconstrucción definitiva de la serie de la estación La Coquera, consintió en rellenar los vacios de la serie normalizada con los valores de la función armónica normalizada como se muestra en la Figura 12—4 .

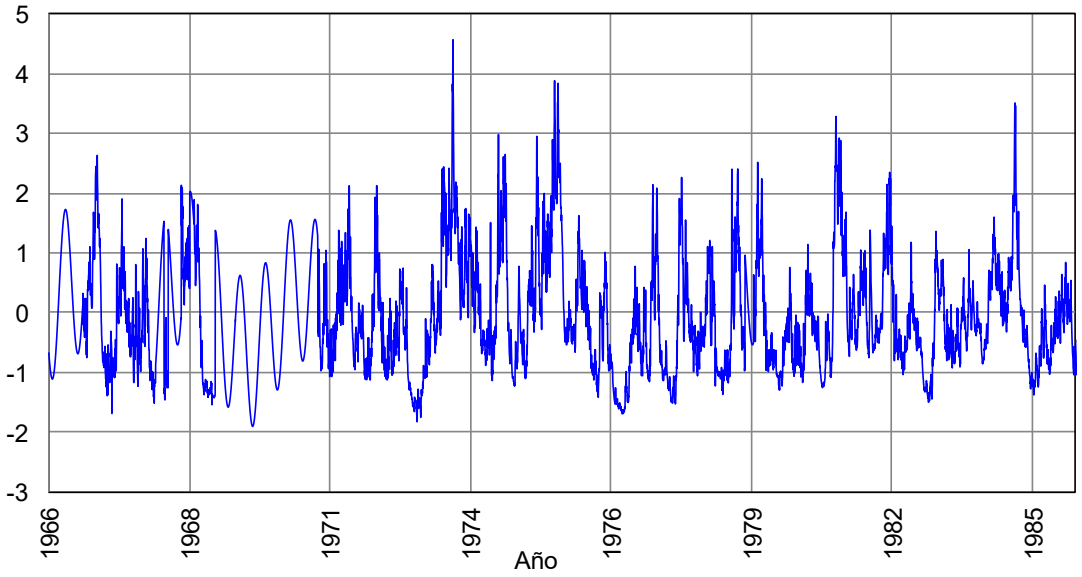


Figura 12—4 Serie normalizada rellenada

Con este procedimiento se garantiza una serie que tiene en cuenta la variabilidad macroclimática de ese periodo, posteriormente se hace el procedimiento contrario a la normalización de la serie completa, con respecto a la media y desviación estándar de la serie original, para de esta manera garantizar que no se genere mucha dispersión al rellenarla, o se sobrestime el caudal medio como se aprecia en la Figura 12—5.

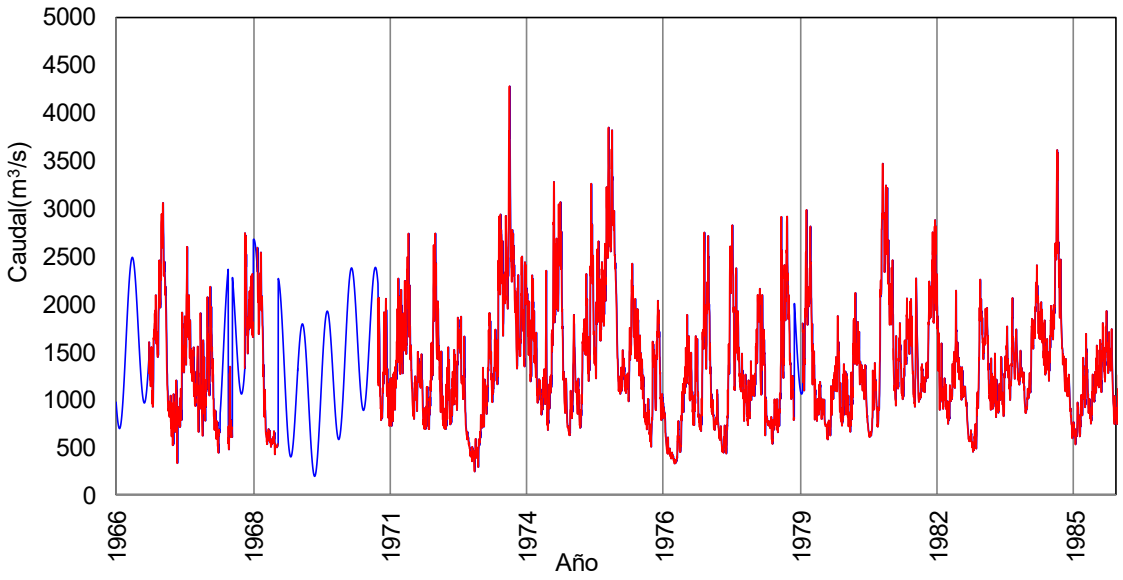


Figura 12—5 Serie de caudales rellenada con armónicos

Finalmente, se tomó el año pivote seleccionado y se estimó la diferencia del caudal con la media del año pivote, esta diferencia se ubican sobre la tendencia del armónico rellenado, en los registros faltantes para el día y mes indicado, encontrando así los caudales medios diarios faltantes para cada año; sin embargo, se deben revisar que si esta diferencia sobre el armónico genera caudales menores al caudal mínimo de toda la serie original, se rellena con este ultimo valor.

Realizando el análisis anteriormente descrito, y tomando a 1986 como el año pivote, la serie reconstruida de caudales medios diarios para la estación La Coquera se presenta gráficamente en la Figura 12—6.

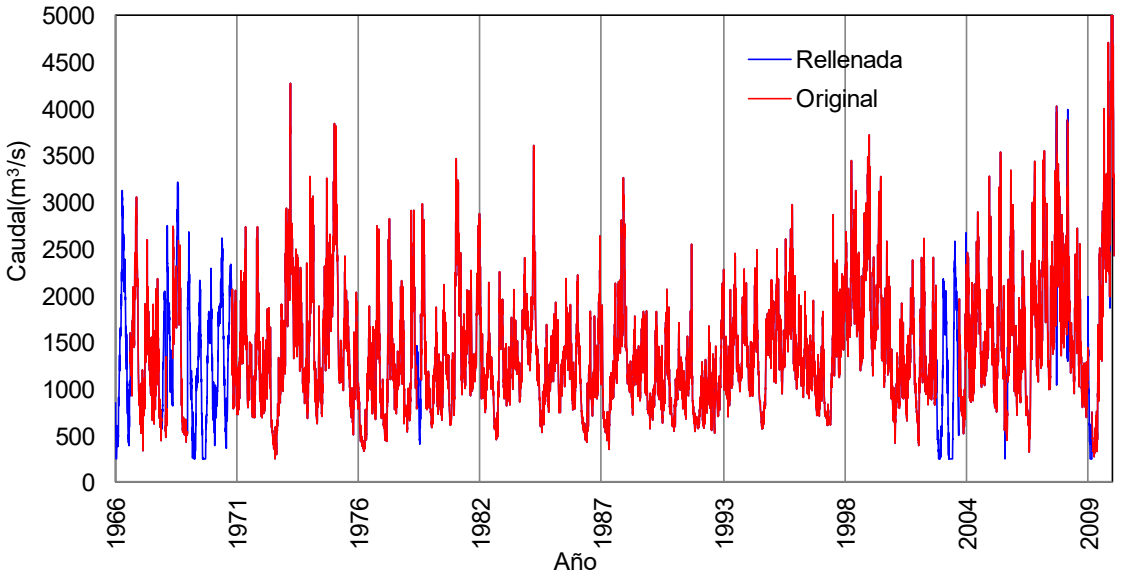


Figura 12—6 Serie definitiva de caudales medios diarios

Con el fin de validar esta metodología se elimino un tramo del registro existente de esta estación y se rellenó este tramo con esta metodología, para de esta manera poder comparar y estimar el error (Mean Absolute Error**)** entre el registro real y el estimado, en la Figura 12—7 se muestra el registro original (línea rojo) comparado con los valores estimados (línea azul).

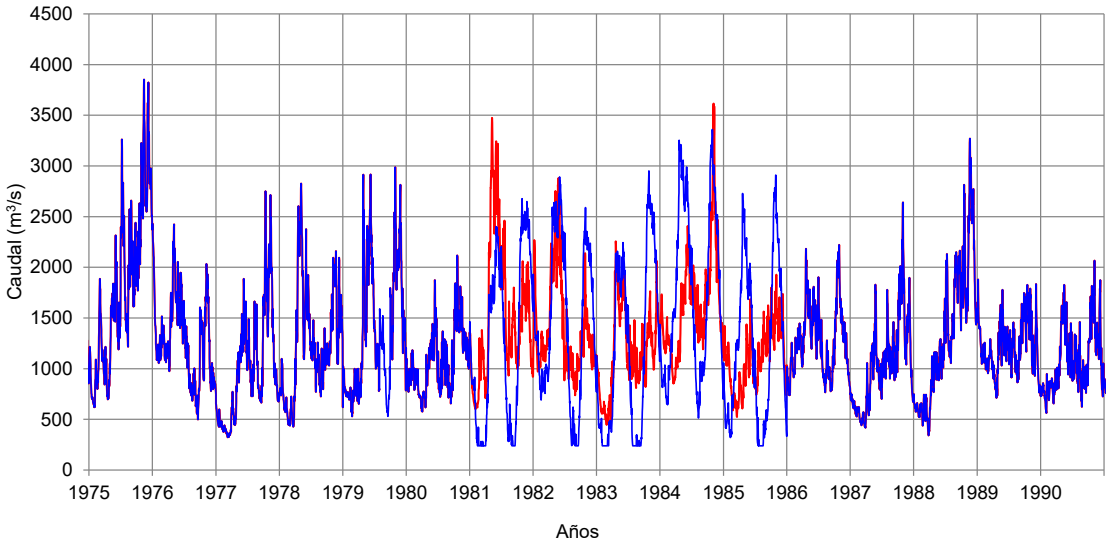


Figura 12—8 Validación metodología para complementar información

Se obtuvo un error de 559 m^3/s lo cual equivale a un 30% de la media, sin embargo con el fin de comparar cada una de las metodologías por aparte, se replico la metodología para el rellenado contemplando solo el armónico y solo el año característico (año pivote). Los errores para cada una de estas metodologías se presenta en la Tabla 12-1.

Tabla 12-1 Errores en m^3/s de la metodología para completar caudales

Método	Mean Absolute Error (m^3/s)
Armónico + año característico	559
Año característico	383
Armónico	426

13 METODOLOGÍA PARA TRASLADAR SERIES DE CAUDALES

Como parte de los objetivos planteados en el presente proyecto, se propuso una metodología para trasladar información de caudales a un punto cualquiera de la cuenca de rio cauca con base en la en las Transformada Rápida de Fourier.

Para lograr este objetivo se propone una función normalizada, la cual se obtiene a partir de los promedios de las funciones armónicas de las estaciones analizadas y se presenta gráficamente en la Figura 13—1 y los valores se encuentran en el Anexo 4.

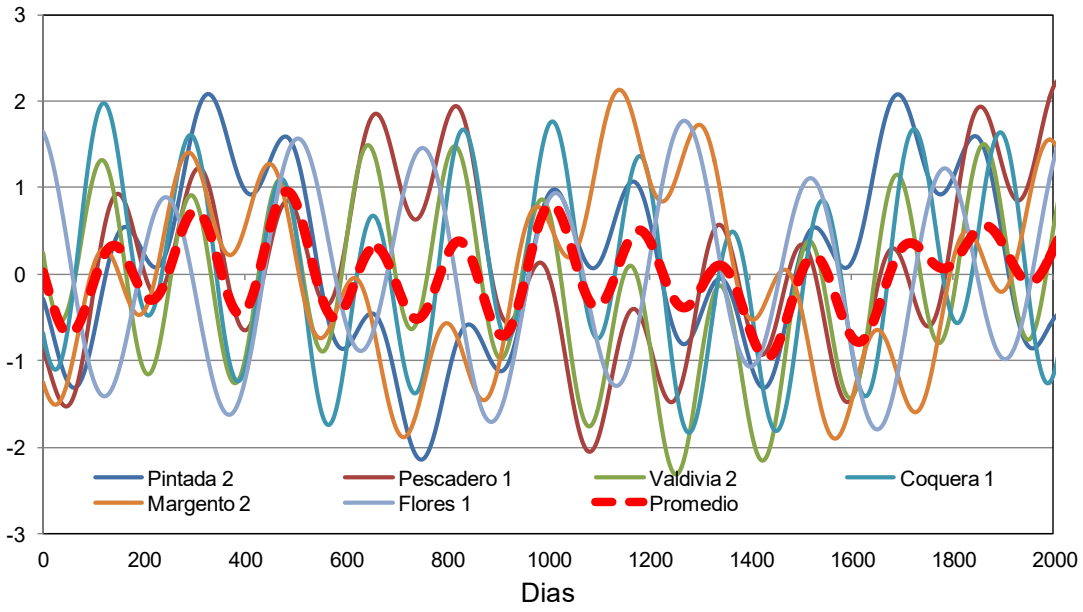


Figura 13—1 Función armónica normalizada

Para obtener la serie de caudales en cualquier lugar dentro de la cuenca de estudio, se necesita el caudal medio de dicho punto, este caudal puede ser obtenido mediante aforos o un estudio hidrológico de la cuenca.

En las estaciones analizadas se encontró una relación entre la desviación estándar y la media de cada una de estas, como se muestra en la Tabla 13-1. Con base en estos resultados se consideró que la desviación estándar tiene una relación promedio que corresponde al 46% de la media de la serie, es decir, la desviación estándar es aproximadamente el 46% de la media de cualquier punto de la cuenca.

Tabla 13-1 Desviación estándar y media de las estaciones de estudio

Estación	La Pintada	Pescadero	Valdivia	Apaví	La Coquera	Margento	Las Flores
Media (m^3/s)	707	1049	1177	1188	1397	1347	2228
Desviación	407	461	532	549	631	600	913
Desv./Media	58%	44%	45%	46%	45%	45%	41%

Finalmente, y tomando los parámetros anteriormente encontrados junto con la función armónica normalizada, y el caudal medio del punto de análisis, se puede reconstruir una serie de caudales teórica en el punto de estudio.

Inicialmente se toma el caudal medio estimado en el punto de análisis, el cual es multiplicado por 0,46 para así obtener la desviación estándar de la serie en el punto. Con estos dos valores se procede a regresar la normalización de la función armónica con respecto al caudal medio, y la desviación estándar encontradas.

Con el fin de validar esta metodología, se realizó una comparación entre el registro original de la estación Pescadero y el armónico generado con la metodología anteriormente planteada basa en la media de esta estación (1049 m^3/s).

Como se muestra en la Figura 13—2, el armónico (línea roja) se asemeja al comportamiento que tiene el registro original de la estación (línea azul), en este caso se encontró un Mean Absolute Error (MAE) equivale a 446 m^3/s, lo cual en porcentaje del caudal medio es equivalente a un 40%, este error elevado es consecuencia de la comparación de armónicos que representan fenómenos macroclimáticos con la serie original que contempla variaciones diarias.

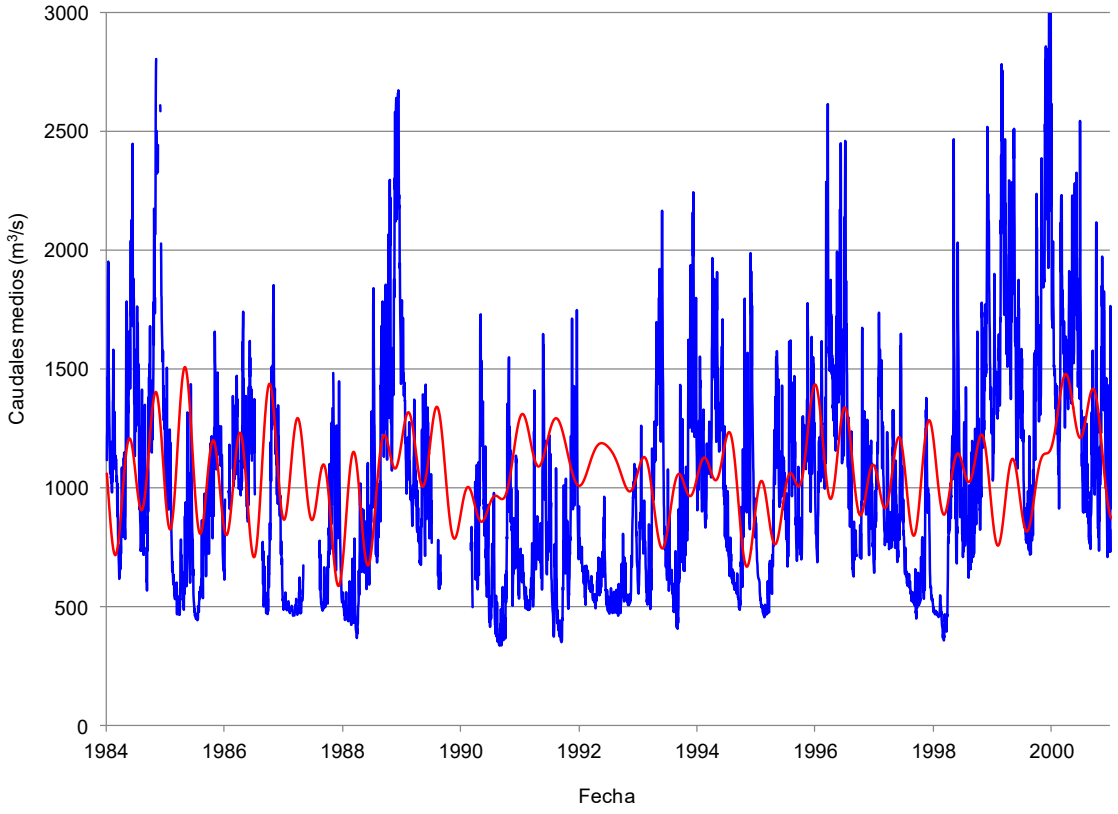


Figura 13—2 Validación metodología traslado de información

14 CONCLUSIONES

En la elaboración de un estudio hidrológico las metodologías de análisis están fundamentadas principalmente en conceptos estadísticos, por lo que se requiere de una importante longitud de registros hidroclimatológicos, los cuales en muchos casos son insuficientes, tienen periodos en los que no se recopiló la información, o incluso no se dispone de los datos en el sitio de interés. Es por esta razón, que la información hidroclimatológica, antes de ser analizada, debe cumplir un rigusoso proceso para analizar la calidad de los registros, y definir los procedimientos más idóneos para obtener un estudio hidrológico confiable.

Uno de los procesos de validación de los registros, consiste en la búsqueda de datos anómalos o Outliers. En este trabajo se determinó que la prueba de Studentized Deviates es la más conservadora, dado que marca como Outliers, valores que no están tan alejados de la muestra. De otro lado, se identificó que la prueba que menos valores identifica como anómalos es la Lm, y por este motivo se considera la metodología menos conservadora.

La identificación de una dato extremo puede afectar la caracterización de un determinado fenómeno climatológico dentro de la cuenca, por esta razón es claro que la eliminación de outliers depende de la importancia y el comportamiento que rija la variable que se está analizando. Por tal motivo, se le debe dar importancia a la experiencia en los análisis de series de caudales, dado que esta variable en las cuencas hidrográficas esta directamente correlacionada con otras variables climatológicas, geológicas, geomorfológicas e incluso antrópicas.

Dado la variabilidad del caudal en una cuenca hidrográfica, la selección del área de estudio fue fundamental para tener la posibilidad de comparar la variación de esta variable en diferentes zonas. Es así, y mediante el análisis de la media de caudal como se identificó que la tendencia de los mismos en el río Cauca es ascendente con respecto al tiempo, dado que las todas las estaciones analizadas presentan este comportamiento excepto las estaciones Pintada y Valdivia. Este comportamiento ascendente se debe probablemente al cambio de usos del suelo de la cuenca del río Cauca en la región andina, pero esta hipótesis deberá der analizada con mayor detalle en estudios posteriores.

Adicionalmente, es importante mencionar que las estaciones que dan tendencia descendente, también fueron analizadas mediante la prueba de carreras, obteniéndose que las mismas no presentaban dependencia respecto al tiempo. Estos resultados parecen indicar que en estas dos estaciones hubo un cambio de equipo de medición, o del sitio en el que se realizaron los aforos. Esta conclusión se deberá validar al considerar información secundaria como las curvas de calibración de los sitios, las cuales no estaban disponibles para la realización de este estudio.

La aplicación de la Transformada Rápida de Fourier (ó FFT por sus siglas en inglés) a las series de caudal de las estaciones permitió identificar ciertos eventos que tienen una frecuencia característica. Estas señales identificadas fueron comunes a lo largo de la cuenca de estudio.

Luego de realizar la FFT a índices macroclimáticos como el MEI, el ONI y Maiden Julian, se identificaron señales que estaban en la misma frecuencia de las estaciones

de caudal analizadas. Por tal razón, se concluye que las oscilaciones hidro-climatoligicas generadas por los fenómenos macroclimáticos que tienen influencia directa sobre la zona de estudio, ya que se identificaron claramente la influencia de los fenómenos como la oscilación meridional de la Zona de Convergencia Intertropical (ZCIT) la cual tiene una recurrencia de 6 meses, así mismo se evidenció la influencia de la oscilación cuasi-bienal (QBO), fenómeno que tiene un periodicidad entre 2 y 2,5 años.

También se identificaron anomalías en los caudales, con una recurrencia entre 5 y 6 años, asociadas a las anomalías hidroclimáticas durante las dos fases del fenómeno El Niño/Oscilación del Sur (ENSO) que se presentan en la sur América tropical incluyendo Colombia.

Respecto al procedimiento de la Transformada de Fourier empleado en este trabajo, en lo que respecta al completar la longitud de los registros para lograr una exponencial de 2, se encontró que este artificio matemático no es muy recomendable, y en su lugar se recomienda eliminar algunos registros para acortar la longitud deseada, o en su defecto, rellenarla con la media de la serie tratada, ya que completar la serie de caudales con valores de ceros no tiene sustento hidrológico.

Para estudios posteriores se recomienda realizar un análisis similar con la transformada de Wavelets, dado que mediante la investigación realizada en el presente estudio, se encontró que teóricamente esta transformada permite trabajar con series no continuas y adicionalmente detecta singularidades, de esta manera podría servir para identificar registros anómalos, por lo cual permitiría hacer un análisis más completo y versátil.

Con respecto a la metodología planteada para rellenar las series de caudales, se encontró que la propuesta planteada tiene una mayor validez, con respecto a la que consiste en completar los registros con un año pivote; este último procedimiento es ampliamente aplicado en el medio por su simplicidad y agilidad. Es importante mencionar que la metodología planteada en este estudio, tiene en cuenta la variabilidad hidro-climatológica a gran escala de los caudales del río Cauca, por lo que se considera más conveniente completar registros de caudal siguiendo los procedimientos planteados en este estudio.

Adicionalmente se encontró que las señales entre las estaciones del río Cauca, son similares, y que con la variabilidad media de las mismas es posible reconstruir la información en cualquier punto comprendido entre las estaciones de monitoreo con un buen grado de precisión.

BIBLIOGRAFÍA

BOTERO, Juan Carlos. Dinámica de estructuras. Sistemas de un grado de libertad. Medellín: Universidad EAFIT, 2011. 244p. ISBN 978-958-720-085-0

CLOUGH Ray y PENZIEN Joseph, Dynamics of structures, 2nd edition, New York, McGraw-Hill,1993. p.738. ISBN 0-07-011394-7

COOLEY J W y TUKEY J W. An algorithm for the machine calculation of complex Fourier series. Math. Comput., 1965. p.301

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

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

HARNETT, Donald L. y MURPHY, James L. Introducción al análisis estadístico. 2ed. México: Addison-Wesley Iberoamericana, 1980. 712 p. ISBN-968-6630-34-1

HELSEL, D.R. y HIRSCH, R.M., Statistical Methods in Water Resources. Amsterdam, Elsevier, 1992. p. 522.

HOEL, Pauk. Estadistica Elemental. 2ed. México: Continental S.A., 1964. 388 p.

INTEGRAL, Proyecto hidroeléctrico Pescadero-Ituango: evaluación de impacto ambiental. Informe Final 1999.

INTEGRAL, Proyecto hidroeléctrico San Andrés: informe de diseño, 2012

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

KOTTEGODA, Nathabandu T. y ROSSO, Renzo. APPLIED STATISTICS FOR CIVIL AND ENVIRONMENTAL ENGINEERS. 2ed. Oxford: Blackwell,2008. 717p. ISBN 978-1-4051-7917-1

PUERTAS,Olga L.,CARVAJAL,Yesid y QUINTERO Mauricio. Estudio de Tendencias de la Precipitación Mensual en la Cuenca Alta-media del Río Cauca, Colombia.

SHUMWAY, Robert H. y STOFFER, David S. Time Series Analysis and Its Applications. With R Examples. 2ed. New York: Springer, 2006. 575 p. ISBN-13: 978-0387-29317-2.

SMITH, Ricardo A. y CAMPUZANO, Claudia. Detección de puntos anormalmente extremos (Outliers) en series hidrológicas. En: SEMINARIO NACIONAL DE HIDRÁULICA E HIDROLOGÍA (14: 2000: Villa de Leyva, Boyacá)

SMITH, Ricardo A. y VELEZ, Maria V. Hidrología de Antioquia. Medellín: Secretaria de Obras Publicas, 1997. 165p.

SPIEGEL, Murray R. Estadistica. 2ed. Mexico: McGRAW-HILL, 1991. 555p. ISBN 970-10-0047-1

TIETJEN, G.L y MOORE, R.H,. Some Grubbs-Type statistics for the detection of several outliers. Technometrics, 1972. p. 597.