

INSTITUTO TECNOLÓGICO DE NUEVO LEÓN
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL



**AJUSTE DE FUNCIONES
PROBABILÍSTICAS A DATOS
ESPACIALES DE ESTACIONES
PLUVIOMÉTRICAS DE LA CUENCA
DEL RÍO SAN JUAN.**

POR

SUSANA FABIOLA MARTÍNEZ FUENTES
PROYECTO DE TESIS

Cd. Guadalupe, Nuevo León, 6 de mayo de 2019

INSTITUTO TECNOLÓGICO DE NUEVO LEÓN

DEPARTAMENTO DE INGENIERÍA INDUSTRIAL



**AJUSTE DE FUNCIONES
PROBABILÍSTICAS A DATOS
ESPACIALES DE ESTACIONES
PLUVIOMÉTRICAS DE LA CUENCA
DEL RÍO SAN JUAN.**

POR
SUSANA FABIOLA MARTÍNEZ FUENTES
PROYECTO DE TESIS

Cd. Guadalupe, Nuevo León, 6 de mayo de 2019

INSTITUTO TECNOLÓGICO DE NUEVO LEÓN

DEPARTAMENTO DE INGENIERÍA INDUSTRIAL



**AJUSTE DE FUNCIONES
PROBABILÍSTICAS A DATOS
ESPACIALES DE ESTACIONES
PLUVIOMÉTRICAS DE LA CUENCA
DEL RÍO SAN JUAN.**

Por

SUSANA FABIOLA MARTÍNEZ FUENTES

PROYECTO DE TESIS

Cd. Guadalupe, Nuevo León, 6 de mayo de 2019

Instituto Tecnológico de Nuevo León

Departamento de Ingeniería Industrial.

Los miembros del Comité de evaluación de Proyecto de Tesis recomendamos que la Tesis “AJUSTE DE FUNCIONES PROBABILÍSTICAS A DATOS ESPACIALES DE ESTACIONES PLUVIOMÉTRICAS DE LA CUENCA DEL RÍO SAN JUAN.”, realizada por el alumno Susana Fabiola Martínez Fuentes, matrícula 14480248, sea aceptada para su revisión como opción de Titulación por Tesis.

Integrantes del equipo de Revisión

Dr. Francisco Gerardo
BENAVIDES BRAVO

M.A. Marta Gabriela
Ríos NAVA

M.E.B.D Angela Gabriela
BENAVIDES RÍOS

Lic. Miguel Ángel
SALAZAR SALINAS

Dr. Juan Ramón
GARCÍA JIMÉNEZ

Cd. Guadalupe 6 de mayo de 2019

Declaración del Autor

Yo, Susana Faviola Martínez Fuentes, declaro que el presente trabajo titulado «AJUSTE DE FUNCIONES PROBABILÍSTICAS A DATOS ESPACIALES DE ESTACIONES PLUVIOMÉTRICAS DE LA CUENCA DEL RÍO SAN JUAN.» es resultado de un proyecto de investigación dentro del Instituto Tecnológico de Nuevo León a cargo del Dr. Francisco Gerardo Benavides Bravo. Confirmando que:

- Este trabajo fue realizado principalmente, para obtener el título de Licenciatura.
- Cualquier parte de esta proyecto no ha sido previamente sometido a revisión en esta Institución o cualquier otra institución.
- Donde he consultado la obra publicada de los otros investigadores, se ha atribuido con claridad.
- He citado los trabajos de los otros autores, dando siempre la fuente.
- He reconocido todas las principales fuentes de ayuda.
- Este trabajo es realizado por mí persona y en forma conjunta con mis asesores, y he dejado claro, exactamente, lo que se hizo por ellos y lo que he contribuido yo.

Firma:

Fecha:

Agradecimientos

Deseo expresar mi agradecimiento al Dr. Francisco Gerardo BENAVIDES BRAVO, revisor de este trabajo de Residencia Profesional, maestro del Instituto Tecnológico de Nuevo León, quien con su apoyo influyó en mi formación académica. También agradezco a la M.A.C. Marta Gabriela Ríos Nava y al Doc. Juan Ramón García Jiménez, Lic. Miguel Angel Salazar Salinas, M.M.C. Ángela Gabriela Benavides Ríos, a todos ellos gracias ya que sin sus observaciones y consejos no hubiese sido posible el desarrollo de esta trabajo.

Deseo agradecer a todos mis compañeros de Licenciatura, quienes constantemente me apoyaron y alentaron en esta proyecto que emprendí.

A mis compañeros de trabajo, mi reconocimiento por sus aportes y palabras de apoyo.

Al igual que al Instituto Tecnológico de NUEVO LEÓN, por el apoyo brindado para la realización de este proyecto.

Agradezco la ayuda prestada por la Comisión Nacional del Agua (CONAGUA) ya que proporcionó la información para la realización de esta investigación, sin este apoyo, nunca hubiese sido posible el desarrollo de la misma.

A los generadores del conocimiento relacionado a los temas que fortalecieron este trabajo, gracias por compartir su ciencia.

También la dedico a, mis padres: JUAN ROBERTO Y MARIA ISABEL, mis hermanos: FATIMA Y JUAN ROBERTO.

Mensaje

«Una serie de temas, hacen que se pueda abrir ideas nuevas, cuando prácticamente no eran claras, el orden académico que se adquiere, permite ajustar estos y obtener efectos satisfactorios.»

SUSANA FAVIOLA.

Resumen

por

Un tema importante en el estudio del comportamiento de las series temporales y, en particular, series de tiempo meteorológicas, es la dependencia a largo plazo. En esta tesis se analiza el comportamiento de las variaciones de precipitación en diferentes periodos, utilizando el análisis de las correlaciones de largo alcance. Variogramas y exponente de Hurst se aplicaron a los datos históricos de diferentes estaciones pluviométricas de la cuenca del Río Bravo-San Juan, en la región hidrográfica RH-24 México. La base de datos fue proporcionada por la Comisión Nacional del Agua (CONAGUA). A los semivariogramas, se les obtuvo el exponente de Hurst y se utilizó como una entrada para llevar a cabo un análisis de agrupamiento de estaciones de lluvia. Grupos de muestras homogéneas que pueden ser útiles en un análisis de frecuencia regional se obtuvieron a través del proceso.

Palabras clave: Series de Tiempo, correlaciones de Largo-alcance, lluvia, semivariogramas, exponente de Hurst, análisis de clústers.

Software: Matlab, GNU-R

An important topic in the study of the time series behavior and, in particular, meteorological time series, is the long-range dependence. This thesis explores the behavior of rainfall variations in different periods, using long-range correlations analysis. Variograms and Hurst exponent were applied to historical data in different pluviometric stations of the Río Bravo-San Juan watershed, at the hydrographic RH-24 Mexico region. The database was provided by the Water National Commission (CONAGUA). Using the semivariograms, the Hurst exponent was obtained and used as an input to perform a cluster analysis of rainfall stations. Groups of homogeneous samples that might be useful in a regional frequency analysis were obtained through the process.

Keywords: Time series, Long-range correlations, rainfall, semivariogram, Hurst exponent, cluster analysis

Software: Matlab, GNU-R

Índice

| | |
|--|-----------|
| Índice | 2 |
| 1. Introducción | 3 |
| 1.1. Antecedentes y descripción del problema. | 8 |
| 1.2. Justificación del problema. | 12 |
| 1.3. Planteamiento del problema. | 14 |
| 1.4. Hipótesis. | 14 |
| 1.5. Objetivo General. | 14 |
| 1.6. Objetivo Específico. | 15 |
| 2. Metodología | 16 |
| 2.1. Análisis | 17 |
| 2.2. Depuración de datos | 18 |
| 2.3. Descripción de series de tiempo | 18 |
| 2.4. Estimación de la dimensión fractal en la práctica | 20 |
| 2.5. Semivariograma | 21 |
| 2.6. Medición del exponente de Hurst (H). | 22 |
| 2.6.1. Mediciones del exponente de Hurst | 25 |
| 2.7. Identificación de regiones homogéneas | 25 |
| 3. Presentación de resultados | 27 |
| 3.1. Discusión del problema | 28 |
| 3.2. Descripción del problema | 32 |
| 3.3. Resultados | 35 |
| 4. Conclusiones y trabajo futuro | 43 |
| 5. Bibliografía | 46 |
| 6. Anexos | 50 |
| 6.1. Estimación de la dimensión fractal en la práctica | 67 |

Capítulo 1

Introducción

Introducción

El comportamiento de las diversas variables climatológicas en el tiempo y en el espacio, han llamado la atención del ser humano desde la antigüedad, ya que condicionan parte del medio ambiente en el cual este desarrolla sus actividades.

El presente trabajo tiene la finalidad de analizar el comportamiento de 33 estaciones pluviométricas instaladas en la región hidrológica RH-24 o cuenca del Río Bravo y efectuar una clasificación de acuerdo a sus características de comportamiento de estas.

La partición espacial de un conjunto de estaciones en grupos sustancialmente homogéneos con respecto a parámetros similares, facilita estudios posteriores de variabilidades en el tiempo o correlaciones con otras variables externas. Particularmente, la clasificación de estas, según su ciclo anual, conduce a la obtención de una regionalización según el régimen de lluvias. Esta regionalización puede ser útil en la búsqueda de relaciones con agentes físicos externos a la hora de ajustar pronósticos y quizá en múltiples aplicaciones hidrológicas.

La mayor parte de los métodos desarrollados se efectúa mediante el análisis estadístico de frecuencias de máximos de una variable hidrológica. La estimación de parámetros, a partir de una muestra pequeña, presenta dificultades, debido a la incertidumbre existente respecto a su representatividad. Es incierto efectuar análisis de dispersiones o de coeficientes de variación debido a que las bases de datos son irregulares y establecer relaciones entre estaciones es casi fortuito. Lo anterior conduce a métodos que asumen una región homogénea respecto a ciertas características estadísticas, lo que permite aprovechar el conjunto de información disponible en la región. La fase más importante en la utilización de información regional, es la de definir las estaciones de precipitación que se consideran similares entre sí, y que puedan ser agrupadas según el grado de heterogeneidad que se quiera asumir para tener un beneficio en el tratamiento conjunto de la información. Aunque no existe un procedimiento que asegure correctamente la definición de una región para el análisis de precipitación, Lettenmaier y Potter (1985), reportan las ventajas de agrupar los datos de precipitación máxima de distintas estaciones con coeficientes de variación (C_v) bajos y homogéneos. En los análisis de precipitación máxima, la mayoría de los métodos toma como base la regionalización.

Una parte de la teoría de eventos extremos se basa en el supuesto de que dichos valores se extraen de poblaciones idénticas e independientes e igualmente distribuidas (iid) ya que esto simplifica las operaciones de muchos métodos estadísticos. Estas condiciones no se cumplen en las series de tiempo meteorológicas porque comúnmente muestran estacionalidades y dependencias.

En el caso de una serie estacionaria se remedia el problema de la dependencia analizando la serie en bloques. Esto se justifica con el hecho que los valores extremos de cada bloque tienden a volverse independientes entre sí conforme aumenta el tamaño de los bloques.

Una serie temporal, que no es otra cosa más que una sucesión de valores de una variable observada en intervalos de tiempo igualmente espaciados y que generalmente es aleatoria, puede verse influenciada por como fueron almacenados sus registros históricos; mensual, bimestral, trimestral, semestral, días, horas, etc. A este tipo de series le llamaremos “*estacionales*” ya que esa manera de registrarlos las afecta.

Se dice que estas poseen tendencia, cuando observamos cierto comportamiento en ellas o cuando sus valores oscilan alrededor de curvas que podemos modelar mediante una ecuación y la cual nos permite establecer criterios como, el de estacionalidad, creciente ó decreciente, tendencia, etc..

En nuestro trabajo, las series de tiempo son registros históricos que fueron analizadas con el interés de efectuar una clusterización, esto es, regiones que poseen un comportamiento similar, para, encontrar la función de probabilidad que permita describirlos.

El caso de las series estacionales no es tan simple, ya que no existe una metodología general. Un método para analizar una serie estacional es asignarle una dependencia en tiempo a los parámetros de la distribución de probabilidad con la que se modelan los datos. Por ejemplo, una serie de varianza constante y con un valor medio que varía en tiempo, podría modelarse con una distribución normal $N(a_0 + a_1t + a_2t^2; \sigma^2)$ y calcular los coeficientes a_i , usando técnicas de máxima verosimilitud.

En otros casos es posible analizar la serie, dividiéndola en segmentos ó estaciones más pequeñas, ya que la función de máxima verosimilitud puede subdividirse para tomarlas en cuenta. Por ejemplo, una serie que muestra variaciones fuertes durante el año podría subdividirse en los 12 meses del año y plantear la función de verosimilitud como un producto de doce funciones (una para cada mes).

La complejidad de los modelos medioambientales es que virtualmente cualquier proceso físico lleva consigo una variabilidad en el espacio y tiempo ya que la interacción supone que los casos cercanos están en el espacio. Se disponen de datos que se revisaron buscando cierto patrón espacial mediante observaciones gráficas.

Esto puede permitir apreciar cierto comportamiento y analizar más a fondo períodos particulares en donde se pueda hacer un mejor análisis ó bien, una clasificación de zonas homogéneas, que formen conglomerados ó clústeres espaciales, además si observamos los

acumulados de algunos meses, nos permite decir que si, en algún período particular, después de ciertos acumulados mensuales, se le agrega la presencia de un evento extremo, puede ocurrir una avenida de agua como las ocurridas con los huracanes, Beulah, Gilberto y Alex.

En particular, nuestro trabajo se centrará en efectuar una clusterización de las estaciones pluviométricas con similares comportamientos utilizando para ello el exponente de Hurst, aplicándolo tanto a los datos directos así como al variograma de estos y ajustar distribuciones de probabilidad a estas, posteriormente se efectuaran simulaciones con estas distribuciones analizándolas con el coeficiente de Jaccard para observar si existe semejanza entre los conglomerados generados por el exponente de Hurst y los generados en la simulación.

El **variograma**, utilizado en la Geoestadística impulsada por George Matheron[2] (1930-2000), es una técnica que se aplicó a nuestros datos la cual permitió observar un suavizamiento en las series de tiempo, se optó por medir esa variabilidad, aplicando un análisis de fractalidad ó análisis de Reescalado, encontrando así el denominado exponente de Hurst[3], que permite medir dicha variabilidad y capturar su magnitud, con la cual generaremos intervalos de valores numéricos de manera natural y por análisis de jerarquización de k-medias, conjeturando que, aquellos que estén en ciertos intervalos, poseen características semejantes, y con ello poder efectuar una clusterización de las estaciones que fueron analizadas.

La Comisión Nacional del Agua (CONAGUA) organismo administrativo desconcentrado de la Secretaría de Medio Ambiente y Recursos Naturales, creado en 1989, con la responsabilidad de administrar, regular, controlar y proteger las aguas nacionales en el país, proporcionó las bases de datos de 41 estaciones pluviométricas distribuidas en la cuenca del Río San Juan que fué la zona de estudio elegida para el presente trabajo. Las estaciones constan de datos desde 1980 hasta el 2012, distribuidos en meses.

Aquí se propone establecer una metodología para identificar comportamiento similar en estaciones pluviométricas, analizando los datos de los registros históricos de estas, aplicando índices de fractalidad, el exponente de Hurst, tanto a los datos directos como a los variogramas de estos y que nos permitirá llevar a cabo una clusterización de estaciones. Una vez hecho esto, se estimará los parámetros de comportamiento como: promedio, desviación, curtosis y asimetría, mediante L-Momentos, para proponer una función de probabilidad que los caracterice.

Organización de la tesis

A lo largo de los próximos capítulos se emplea una terminología que es importante aclarar para una correcta comprensión de los mismos. El elemento fundamental de análisis considerado en esta investigación son las series temporales representadas por un conjunto de datos con estructura compleja y cuando se habla de estos se hace referencia al conjunto de todos aquellos atributos propios de cada una de las 33 estaciones que analizamos. La organización de esta tesis comprenderá lo siguiente:

Capítulo 1 Introducción, antecedentes, descripción del problema, justificación.

Capítulo 2 Metodología aplicada en el desarrollo de la Tesis.

Capítulo 3 Presentación de Resultados. Aquí se discuten las contribuciones del presente trabajo.

Capítulo 4 Conclusiones, trabajo futuro y recomendaciones.

* Se presenta una explicación breve de la revisión de los datos y su filtración para llevar a cabo el análisis.

Anexos * Introducción a las series de tiempo.

* Explicación de Fractales.

* Explicación de Variogramas, técnica aplicada a la presente investigación.

* Rango reescalado, técnica aplicada posteriormente a los variogramas para la obtención del exponente de fractalidad.

* Se esboza el porque establecer una clasificación de regiones homogéneas utilizando el exponente de Hurst[3].

1.1. Antecedentes y descripción del problema.

A nivel nacional se cuenta con numerosas estaciones pluviométricas como se muestra en la Fig.1.1.

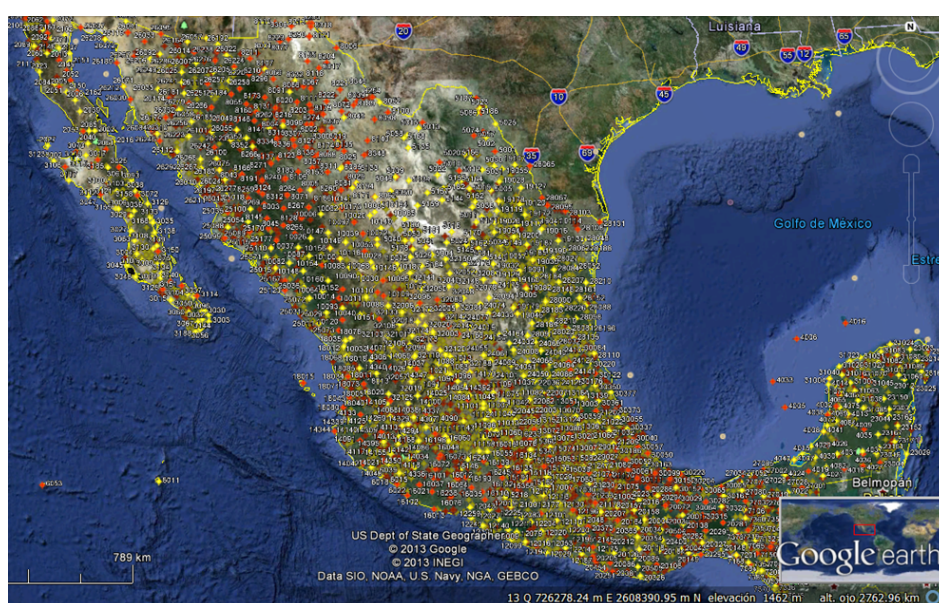


FIGURA 1.1: Estaciones pluviométricas de la Nación. Imagen obtenida de: "Google Earth"

La cuenca del Río San Juan integra diversos centros urbanos entre los cuales se encuentra la Ciudad de Monterrey con su área metropolitana y Saltillo (capital de Coahuila), que en conjunto a las zonas citricolas y agrícolas conforman un sistema económico que es crítico tanto para la región como para el país. En el Anexo 1 damos una descripción de la ubicación de la cuenca del Río San Juan.

Eventos atípicos, han mostrado lo susceptible de la región ante fenómenos extremos como el caso de los huracanes, Beulah en 1967, Gilberto en 1988, Emily en 2005 y el ocurrido más recientemente, Alex, en el 2010.

Este tipo de riesgos hidrológicos ocasionales de grandes avenidas, junto a otro riesgo de grandes sequías, conforman parte de los comportamientos extremos de la zona de análisis.

Si se desea lograr una mejor planificación urbana, elaboración de planes de contingencia y paliación de riesgos, se requiere contar con controles regionales del perfil de riesgo de estos comportamientos extremos.

Entre las estrategias de análisis de riesgo se encuentra la modelación espacial de niveles de precipitación para diferentes períodos en las zonas bajo estudio.

En esta investigación, se realizó una revisión de la información histórica proveniente de 41 bases de datos que dan seguimiento a las estaciones meteorológicas de la cuenca del Río San Juan. Cada estación está caracterizada por su ciclo anual en un período de 1984-2011. De estas 41 estaciones se seleccionaron 33 debido a que algunas poseen menos de esta información y otras que dentro de ese período carecen de algunos años de registro. Los datos fueron recopilados por la CONAGUA. Se observó el comportamiento de las diferentes estaciones y se efectuó un análisis de fractalidad y se agrupó en conglomerados mediante el exponente de Hurst, también se efectuó análisis de variograma y se efectuó nuevamente clusterización mediante el exponente de Hurst, para ajustar posteriormente modelos probabilísticos que permitirán hacer simulación y pronósticos de precipitación de lluvias para distintos períodos de interés.

Los fenómenos extremos causados por precipitaciones en la región generalmente ocurren en verano y otoño.

Estas lluvias están asociadas a la influencia de sistemas atmosféricos de gran escala, como la zona de convergencia intertropical y los vientos alisios, así como a sistemas transitorios y ciclones tropicales. La zona intertropical de convergencia es una región de bajas presiones alrededor del globo y cerca al Ecuador, donde tiene lugar el encuentro de los vientos del este (alisios) provenientes de cada hemisferio, eso da lugar a movimiento de ascenso de la masa de aire y a una abundante precipitación.

El decir fenómeno extremo, es como sinónimo de acontecimiento poco común bajo la perspectiva humana. Sin embargo, la formación de una gota de lluvia es un fenómeno natural de la misma manera que un huracán, esta expresión también se refiere, en general, a los peligrosos fenómenos naturales también llamados "*desastres naturales*", la lluvia, por ejemplo, no es en sí un "*desastre*", pero, "*puede serlo*", dependiendo de la perspectiva humana, si ciertas condiciones se reúnen. La mala planificación urbana, con la construcción de estructuras en lugares vulnerables e inundaciones, puede causar efectos desastrosos para los seres humanos. Los huracanes, como un factor importante de las lluvias de verano en México, se les asocia con este tipo de situaciones de desastre y pérdida de vidas humanas, constituyen los principales sistemas productores de lluvia en verano para regiones como el noreste del país. Un caso de este tipo ocurre cuando el flujo que acompaña a un ciclón tropical choca con una cadena de montañas, lo que produce un ascenso de la masa de aire cargada de humedad, causando una magnificación del proceso convectivo. Ejemplos de esta situación se presentaron en Nuevo León y Tamaulipas durante el huracán Gilberto en

septiembre de 1988, el huracán Paulina (afectando a Guerrero y Oaxaca) en 1997 y Alex en 2010. Las inundaciones extraordinarias, así como la persistencia de la sequía en grandes extensiones de la república, es lo que motiva a efectuar algunos planteamientos que pueden servir para entender mejor estos fenómenos y, con esa base, disminuir los daños que se derivan de los escurrimientos extremos.

En el estudio de los escurrimientos máximos por precipitaciones extremas, las inundaciones constituyen la principal preocupación. Dependiendo de su duración, la magnitud de las áreas afectadas y el tipo de afectación, las inundaciones provocadas por lluvias intensas en cuencas con respuesta rápida provocan los denominados “*flash floods*” (inundaciones repentinas), los cuales se acompañan casi siempre de una gran cantidad de lodo. Algunas de estas se han presentado en el Valle de México, en el arroyo Topo Chico y Santa Catarina en Monterrey y en las serranías de Puebla y Veracruz.

Inundaciones de larga duración, por la lentitud con que se producen, causan lamentablemente pérdidas humanas e importantes pérdidas económicas tanto en zonas urbanas como en zonas rurales. Encharcamientos se presentan casi siempre en sectores urbanos, cuya principal consecuencia es el retraso en el desarrollo de las actividades productivas de la población.

Dada la diversidad de características de aquellas avenidas que pueden producir inundaciones de distinto tipo, el problema de la estimación del riesgo es complejo y es necesario caracterizarlo ya que esto permitiría diseñar una política de contingencia o mitigación de los daños con medidas preventivas o resolutivas.

Debido a los riesgos presentados anteriormente es necesario realizar estudios regionales de lluvias extremas. Al llevar a cabo el estudio regional, es necesario identificar los sitios similares y agruparlos en conglomerados y, de acuerdo con el tipo de afectación, analizar con probabilidades los registros históricos con objeto de definir, en cada caso, la probabilidad de que en un año cualquiera se presente una inundación asociada a una lluvia extrema o cuál será el nivel de retorno correspondiente.

Para la modelación de estos fenómenos se puede utilizar la metodología de análisis regional de frecuencias, modelos jerárquicos basados en estadística bayesiana y procesos **max-estables**. En el Anexo 2 damos una breve explicación del porque es requerido diferentes análisis al respecto. El primero de ellos ha sido aplicado en estudios en EE.UU, el Noroeste de México, Chile, Turquía y algunas regiones de Europa. Este análisis permite realizar predicciones decrecientes, es decir, estimaciones asociadas a una determinada probabilidad de excedencia, con base en todos los datos observados en varias estaciones hidrométricas de una región.

Para la aplicación de esta metodología se siguen cinco etapas: (1) revisión y preparación de los datos, (2) identificación de regiones homogéneas, (3) selección de la distribución de frecuencia, (4) cálculo de parámetros y estimación de la función de cuantiles, y (5) mapeo del periodo de retorno.

Algunas de las distribuciones que más se utilizan para el ajuste de dicha información hidrológica son: distribuciones de: Gumbel, Pearson, Log Pearson, Gamma, Normal, Log Normal, General de Valores Extremos, entre otras.

La base de esta investigación fueron datos provenientes de 41 estaciones pluviométricas, a las que se les realizó un análisis estadístico de reescalado o índice fractal de Hurst[3] y Variogramas, elaborando una clusterización considerando un comportamiento similar en un intervalo, se tomó la mejor distribución de probabilidad, estimando los parámetros: media, desviación, asimetría y curtosis mediante L-momentos.



FIGURA 1.2: Ubicación de las estaciones pluviométricas de estudio: "Google Earth"

1.2. Justificación del problema.

Los modelos determinísticos o los modelos probabilísticos son la base para analizar un “*fenómeno natural*”.

a) En los modelos determinísticos, dadas ciertas condiciones, “*siempre*” ocurre el mismo resultado. Lo único que podría variar el resultado, es un “*error*” en la medida. Muchos fenómenos físicos y químicos pertenecen a esta categoría.

b) En los modelos probabilísticos, el fenómeno ocurre con una *determinada frecuencia* que no es constante y es aquí donde la mayoría de los *fenómenos* caen en esta categoría, ya que están influenciados por muchos factores variables y a su vez por factores no identificados.

Cuando el modelo es *predominantemente probabilístico*, se necesitan *técnicas estadísticas* para alcanzar el conocimiento científico – *la presencia de variabilidad requiere análisis estadístico*.

Las *técnicas estadísticas* son: *herramientas objetivas* con las cuales se pueden definir las *tendencias de los grupos* a partir de cierto número de *observaciones individuales* (también llamadas *muestras*).

Los promedios, porcentajes, desviaciones, coeficientes de variación y asimetrías de nuestros datos que se calculan para ajustar la posible distribución de probabilidad, representan una pequeña parte del método, importante, pero solamente restringido a describir las características de la zona de estudio. En realidad, en el razonamiento estadístico, se elabora *inferencias inductivas*, de lo particular (muestras), a lo general (población), dado que no es posible medir un parámetro específico en la *totalidad* de la población investigada.

En esta investigación se combinarán, el *Análisis Regional de Frecuencias, varigramas y exponente de fractalidad de Hurst*, los cuales constituyen una herramienta secuencial extremadamente potente para efectuar una clusterización de estaciones con comportamiento similar y así resumir la forma de distribuciones de probabilidad, en especial cuando:

- Se espera que la distribución de frecuencia sea altamente asimétrica (precipitaciones mensuales, anuales en ciertas regiones, lluvias máximas, períodos de retorno, etc.),
- Se dispone de registros cortos (cuando hay asimetría, un registro corto incluso es aquel menor a 100 datos),
- Ante la presencia de valores atípicos,

-
- Cuando se requiere de una estimación robusta, insesgada, y con un abundante fundamento teórico, altamente estimable y prácticamente axiomático.
 - Cuando se desea utilizar distribuciones de tres o más parámetros, muy flexibles, pero que con el procedimiento convencional requieren momentos de mayor orden que causan sesgos enormes en las estimaciones.

1.3. Planteamiento del problema.

El estudio de una regionalización parte del principio que se requiere subdividir o fraccionar cierta zona en un conjunto y sistema de regiones menores dentro de los límites de manera que facilite el ejercicio del control ya sea administrativo, asignación de los recursos, políticas de dirección. Regionalización es un procedimiento para modificar el orden territorial en unidades más pequeñas con características comunes y representa una herramienta metodológica básica en la planeación ambiental, pues permite un manejo adecuado donde se puede analizar la dependencia entre objetos de análisis. Con el objetivo de la identificación de regiones homogéneas, es necesario encontrar correlaciones entre las diferentes bases de datos y es necesario recortar, en algunos casos, información para encontrarlas.

El presente trabajo pretende responder y aportar información en relación a las siguientes preguntas: ¿ Existe la posibilidad de capturar la magnitud en la variación de los datos mediante el análisis fractal ?, ¿ Es el exponente de Hurst otra herramienta estadística para hacer esto ?, ¿ Podrá servirnos como relación ?, y de ser así, ¿ Puede esta medida, permitirnos efectuar cúmulos que tengan características similares ?.

1.4. Hipótesis.

Procedimientos como el análisis regional de frecuencias son utilizados para ajustar una muestra de datos a un tipo de distribución, asociando la forma de la distribución a un número finito de parámetros. La asimetría y la curtosis se utilizan comúnmente para establecer la proximidad de los valores observados (muestra) a diferentes tipos de distribuciones. Sin embargo, el cálculo de estos estadígrafos es sensible al tamaño de la muestra. ¿ Puede el exponente de Hurst, formar parte de una estrategia para identificar regiones con comportamiento similar ?

1.5. Objetivo General.

Establecer una modificación a la metodología para la identificación de Regiones Homogéneas dentro del análisis regional de frecuencias utilizando ahora el exponente de fractalidad de Hurst.

1.6. Objetivo Específico.

Proponer una regionalización de las estaciones de la región RH-24, mediante el análisis de comportamiento fractal y así identificar comportamientos similares en las diferentes bases de datos, de nuestras estaciones pluviométricas, para nuestro caso de estudio. Posteriormente aplicar el método de L-Momentos y estimar de parámetros para el ajuste de las funciones de probabilidad que se utilizan tradicionalmente adecuándose a la información hidrológica (Generalizada de valor extremo, lognormal, Pearson tipo III, Gamma o Gumbel).

Capítulo 2

Metodología

2.1. Análisis

Aunque en la cuenca del Río San Juan hay muchas estaciones meteorológicas, gestionadas por diferentes instituciones públicas y privadas, un buen número de estas se encuentran inactivas, otras han empezado a operar en las últimas décadas o tienen series de menos de 10 años continuos.

Después de un análisis en función de su altitud, longitud y latitud se complementaron las series de datos. De las 41 bases de datos de las estaciones gestionadas se seleccionaron 33 y estas se detallan en el capítulo 3. Se analizaron las series de tiempo y se les aplicó un método orientado a corregir los valores anómalos, correcciones de datos, relleno de celdas, análisis de posibles(outliers), para tener las bases homogéneas agrupadas por meses, con datos comprendidos desde 1984-2011.

En el presente estudio, se integra el análisis fractal a los datos directos así como también a los variogramas, el cual se utiliza para cuantificar las correlaciones de largo alcance de datos de las diferentes estaciones pluviométricas con registros mensuales.

Teniendo en cuenta el análisis de la muestra de la serie histórica, se realizó un análisis de rango reescalado R/S para obtener una medida particular del exponente de Hurst[3]. El proceso se repite para cada estación pluviométrica en la región bajo análisis.

Este exponente de Hurst, se utilizó como referencia para identificar las estaciones que presentan patrones similares. Como consecuencia, se aplicó un análisis de clústers de conglomerado o frecuencia para identificar nuestras estaciones homogéneas.

Una ventaja del exponente de Hurst es la simplicidad de su algoritmo que puede ser utilizado para medir la condición de persistencia o antipersistencia de un proceso y proporciona, una métrica que puede utilizarse para clasificar series de tiempo diferentes.

2.2. Depuración de datos

La depuración fue una actividad inevitable y difícil dentro del análisis de datos estadísticos. Esta primera actividad, nos permitió detectar citas erróneas en la codificación de las variables cuantitativas. Se entiende por errores de digitación o codificación, a los valores o datos que son inconsistentes, valores ausentes, fuera del rango, duplicados de las variables cuantitativas que nos interese analizar y/o a códigos numéricos o no numéricos no definidos. Los diferentes tipos de error que se presentaron en los datos, se analizaron con técnicas desarrolladas por diferentes investigadores para detectarlos y corregirlos.

Dado que existe una multiplicidad de técnicas para la depuración de los datos, resulta que no es trivial decidir cuál o cuáles deben ser utilizadas en nuestro caso particular. En esta tesis, se consignan algunos pasos como guía metodológica que puede apoyar para el análisis de datos de acuerdo con la naturaleza y la distribución de los mismos detallado en el Anexo 3.

2.3. Descripción de series de tiempo

Una serie temporal cronológica es un conjunto de observaciones ordenadas en el tiempo, que pueden representar la evolución de una variable a lo largo de él. También podemos considerarla como la realización de un proceso estocástico en tiempo discreto, donde los elementos están ordenados y corresponden a instantes equidistantes del tiempo. El conjunto de observaciones se simboliza: $\{X(t_i), i = 1, 2, \dots, n\}$ donde t_i es independiente e indica sucesivos instantes o tiempos determinados (quinquenos, años, trimestres, meses, . . . , etc).

“ X ” es la variable cuyo comportamiento a través del tiempo se desea estudiar o sea que la serie de tiempo es una serie estadística (información cuantitativa) cuyos valores han sido observados en el tiempo.

La serie puede simbolizarse como $\{X_n, n = 1, 2, \dots, n\}$, la cual representara el paso del tiempo, configurando un proceso estocástico que tendrá su propia función de distribución con sus respectivos momentos. Normalmente, para reconocer y caracterizar las distribuciones resulta complejo, basta con especificar la media y la varianza para cada $X(t)$ y la covarianza para variables referidas a distintos valores de t :

$$\mu_t = E[X(t)] \quad (2.1)$$

$$\sigma_t^2 = Var(X(t)) = E[X(t) - \mu_t]^2 \quad (2.2)$$

$$\gamma_{t_1, t_2} = Cov(X(t_1), Y(t_2)) = E[(X(t_1) - \mu_{t_1})(X(t_2) - \mu_{t_2})] \quad (2.3)$$

donde $E[X(t)]$ es el valor esperado de $X(t)$, μ_t es la media, σ_t^2 es la varianza y γ_{t_1, t_2} la covarianza. $X(t)$ puede ser:

- Fuertemente estacionaria si todas las funciones de distribución conjuntas son constantes, o dicho con más propiedad, son “invariantes con respecto a un desplazamiento en el tiempo” (variación de t). Es decir, considerando que $t, t+1, t+2, \dots, t+k$ reflejan períodos sucesivos: $F(X_i, X_{i+1}, \dots, X_{i+k}) = F(X_{i+m}, X_{i+1+m}, \dots, X_{i+k+m})$ para todo t, k y m .
- Débilmente estacionario si:
 - Las esperanzas matemáticas de las variables aleatorias no dependen del tiempo, son constantes: $E[X_i] = E[X_{i+m}]$ para todo m .
 - Las varianzas no dependen del tiempo (y son finitas): $Var[X_i] = Var[X_{i+m}] \neq \infty$
 - Las covarianzas entre dos variables aleatorias del proceso correspondientes a períodos distintos de tiempo (distintos valores de t) sólo dependen del lapso de tiempo transcurrido entre ellas: $Cov(X_{t_1}, Y_{t_2}) = Cov(X_{t_1+m}, Y_{t_2+m})$

De esta última condición se desprende que, si un fenómeno es estacionario, sus variables pueden estar relacionadas linealmente entre sí, pero de forma que la relación entre dos variables sólo depende de la distancia temporal k transcurrida entre ellas.

Sea ρ la Función de Autocorrelación (FAC) de $X(t)$. Si la serie de tiempo es débilmente estacionaria, FAC estaría dada por:

$$\rho(k) = \frac{E[X(t) - \mu]E[X(t+k) - \mu]}{\sigma^2} \quad (2.4)$$

La serie de tiempo, $X(t)$, tiene la propiedad de Dependencia de Largo Alcance (DLA) si $\sum_{k=-\infty}^{k=\infty} \rho(k)$ diverge, esta puede determinarse de dos maneras:

- En el dominio del tiempo, donde se manifiesta con un alto grado de correlación entre sitios separados a cierta distancia.
- En el dominio de la frecuencia, donde se manifiesta como un nivel significativo de las frecuencias próximas a cero.

Existen diferentes técnicas para su estimación. Por ejemplo: el rango reescalado R/S , el método de momentos absolutos, el método modificado R/S o método de $Lo[12]$, periodograma, Wavelets, entre otros. En esta investigación empleamos el método R/S , el cual es uno de los más aplicados, debido a la efectividad y sencillez de su algoritmo. Éste fue propuesto por Hurst(1951), afinado, posteriormente por Mandelbrot y Wallis[13](1969) y Mandelbrot[14] (1972).

2.4. Estimación de la dimensión fractal en la práctica

Algunos fenómenos u objetos de la vida real pueden mostrar propiedades fractales como lo mencionamos en el Anexo 5 y podemos calcular su dimensión fractal. Es aquí donde puede ser útil obtener la dimensión fractal de un conjunto de datos de una muestra. Este cálculo no se puede obtener de forma exacta sino que debe estimarse.

Esto se usa en una variedad de áreas de investigación tales como la física, análisis de imagen, acústica, ceros de la función zeta de Riemann é incluso procesos electroquímicos.

Tales relaciones de escala familiares pueden definirse matemáticamente por la regla de escala general, donde la variable N , es el número de particiones, ϵ es el factor de escala, y D es la dimensión fractal:

$$N \propto \epsilon^{-D} \quad (2.5)$$

Para nuestro caso de estudio, al revisar las series de tiempo de las precipitaciones fenómeno climático, con amplia variabilidad y comportamiento aleatorio, de diferentes estaciones pluviométricas, mostraron ser irregulares, y nos llevó a enfrentamos a una pregunta, ¿Qué tan irregulares están?. Lo que podíamos decir, depende de su resolución. Si utilizamos técnicas para el análisis de series de tiempo, las conclusiones pueden ser diferentes dependiendo del tipo de herramienta. De hecho, cada vez que utilizemos una nueva técnica obtenemos un resultado con más y más detalle. Pero entonces, ¿cuál será la mejor manera de efectuar un análisis fractal de series espacio temporales ?.

La pregunta así formulada es incorrecta, pues como dijimos, depende de nuestra resolución de medida.

Al aplicar fractalidad o Análisis Reescalado (R/S) pudimos capturar su dimensión fractal, mediante el exponente de Hurst, mostrando persistencia significativa a largo plazo, el cual es uno de los factores más importantes que caracterizan a las precipitaciones, debido a los errores aleatorios sistemáticos en ellas (Mandelbrot y Wallis, 1969; McGregor y Nieuwolt, 1998) donde podemos considerar la propiedad de memoria a largo plazo donde la dependencia temporal persiste, aún entre observaciones, mostrando que pueden ser caracterizadas a través de la dimensión fractal. De esto, los valores obtenidos pueden ser utilizadas como instrumento de clasificación o closterización.

En un determinado rango, obtenemos una línea recta como función al representar en una gráfica *log-log*. La pendiente de esa recta es el exponente de fractalidad que puede caracterizar las diferentes cambios variacionales, y la distingue de otras.

2.5. Semivariograma

El semivariograma o variograma, $\gamma(h)$, es un ajuste o modelado espacial considerado como un estimador de la varianza poblacional y de análisis estructural, donde la población debe tener una tendencia de estacionalidad, se utiliza para describir la relación de observaciones pareadas separadas por una distancia h y en otros casos con una dirección.

Es una técnica geoestadística, la cual permite una medida cuantitativa de la persistencia a largo plazo en series de tiempo no estacionarias Witt[16], Haslett[17], Dmowska[18] et al. Establece correlaciones a través del tiempo a un fenómeno regionalizado en el espacio, generando patrones que pueden ser utilizados para describir el comportamiento de un conjunto de observaciones.

Matemáticamente, el variograma estima la diferencia cuadrada prevista entre variables aleatorias vecinas, dando un soporte fundamental y permitiendo representar cuantitativamente esta relación. Este proceso continúa para cada punto de medición.

Teniendo en cuenta una serie de tiempo o procesos estocásticos $\{X_t, t \geq 0\}$, la función de autocovarianza en el punto $(t, t+h)$ se define como $C_X(t, t+h) = E[X_t X_{t+h}] - E[X_t]E[X_{t+h}]$ con $E[X_t]$ la media del proceso en tiempo t .

Sin embargo, para una sola serie de tiempo, $\{X_n, n = 1, 2, \dots, n\}$, se espera que el valor puede ser estimado suponiendo una hipótesis de ergodicidad, i.e., es decir, un principio estadístico de equivalencia según la cual “*el promedio a través del tiempo y el promedio a través del ensamble son los mismos*” Lefevbre[19].

Así las diferencias $X_{t+h} - X_t$, que se obtendría con un proceso infinitamente reproducible, son “simulados” ó “clonados” de la “serie madre”.

Por lo tanto, el valor medio de las diferencias $X_{t+h} - X_t$ es estimado por:

$$\gamma(h) = \frac{1}{2n(h)} \sum_{t=1}^{n(h)} (x_{t+h} - x_t)^2 \quad (2.6)$$

Una explicación mas a detalle se da en el Anexo 7. Este estimador de momentos, es un promedio de diferencias al cuadrado, que puede ser influenciado por un número pequeño de valores que ocasionan discrepancias al final de los cálculos debido a las particiones realizadas. Pero, se considera un estimador robusto, ya que disminuye la importancia de las diferencias grandes y al cuadrado.

También se considera robusto en el sentido de que es resistente a distribuciones normales contaminadas y afloramientos posiblemente generados por distribuciones de colas pesadas. Esto puede verse por el uso de la raíz cuadrada de las diferencias, en lugar de diferencias de cuadrados, en el estimador insesgado.

La aplicación de variogramas a nuestros datos mostró periodos largos de comportamiento similar, con diferente duración, un comportamiento o patrón cíclico, pero no periódico.

2.6. Medición del exponente de Hurst (H).

El famoso hidrólogo británico Harold Edwin Hurst (1880-1978), trabajó las fluctuaciones de los niveles del Río Nilo, durante largos períodos de tiempo, su interés era proyectar las capacidades de las reservas y tomar medidas de precaución en épocas de sequía. Para esto, ideó una nueva metodología estadística, la cual consiste en saber si las tendencias de la serie de tiempo tienen persistencia o no, luego de medir la duración de ciclos de las series de tiempo y posteriormente determinar si la serie de tiempo es fractal, ó comprobar si esta posee memoria. Esto, con el fin poder proyectar los resultados a futuro.

El método contiene una serie de pasos básicos, que son necesarios para calcular un valor H denominado exponente de Hurst, indispensable para determinación de la persistencia o antipersistencia de una serie de tiempo, además proporciona información sobre la dimensión fractal, dato importante para el desarrollo de nuestra tesis, ya que con el configuraremos regiones con un mismo comportamiento.

Los pasos para estimar el exponente de Hurst, a partir de una serie temporal de longitud N son, dividir la serie en un conjunto de d -subseries de tiempo más cortas, cada una de longitud m , efectuando lo siguiente pasos para cada subserie desde $n=1, \dots, d$:

- Calcular la media E_n , y la desviación estándar S_n ;
- Normalizar los datos (X_{in}), sustrayendo a cada uno, la media de la subserie;

$$Z_{in} = X_{in} - E, \quad i = 1, 2, \dots, m; \quad (2.7)$$

- Obtener las sumas parciales para cada serie de tiempo.

$$Y_{in} = \sum_{j=1}^i Z_{in}, i = 1, 2, \dots, m; \quad (2.8)$$

- Calcular el rango de cada subserie;

$$R_n = \max_{i=1:m}(Y_{in}) - \min_{i=1:m}(Y_{in}) \quad (2.9)$$

- Se reescala o normaliza el rango calculando;

$$\frac{R_n}{S_n} \quad (2.10)$$

- Una vez calculados los reescalados para cada subserie de longitud m se promedian:

$$\left\langle \frac{R}{S} \right\rangle_m = \frac{1}{d} \sum_{n=1}^d \frac{R_n}{S_n} \quad (2.11)$$

Hurst[12] encuentra que la relación del estadístico (R/S) está dado por la siguiente ley de potencia:

$$\left\langle \frac{R}{S} \right\rangle_m \approx c * m^H \quad (2.12)$$

donde H , es el exponente de Hurst y c es una constante positiva.

Dos factores que intervienen en la determinación del coeficiente de Hurst son: la forma en que la serie temporal es dividida en un conjunto de subseries, donde el rango de los valores de t sobre el cual la pendiente de $\log(\langle R/S \rangle_t)$ y $\log(t)$ es calculado y el segundo factor que interviene en la determinación de H , es el resultado del comportamiento asintótico del rango reescalado, es decir, cuando el valor de t tiende a infinito.

El análisis reescalado $\langle R/S \rangle_t$ sobre varios valores de t , es estimado aplicando log/log a la expresión dada, esto es:

$$\log\left(\left\langle \frac{R}{S} \right\rangle_m\right) = \log(c) + H * \log(m) \quad (2.13)$$

Para obtener el coeficiente H , se lleva a cabo un ajuste lineal de los puntos de la relación $\langle R/S \rangle_t$ vs. $\log(t)$ por el método de mínimos cuadrados.

La pendiente de dicha línea es entonces el **coeficiente de Hurst, H** .

Este exponente es un índice fractal que proporciona información sobre una medida de memoria a las correlaciones a largo plazo que se presentan en las series de tiempo.

En la práctica los valores del exponente de Hurst se encuentran entre 0 y 1. Apoyado en dicho valor, ver Mendelbort y Wallis[13], se pueden catalogar las series de tiempo y se puede establecer que las series son:

- Anti-persistentes, lo que significa que para un incremento es más probable que sea seguido por un decremento, y viceversa.
- Aleatoria, corresponde a falta de correlación en la serie (denominado, ruido blanco Gaussiano).
- Persistente, es decir, a un incremento es muy probable que le siga un incremento, y a un decremento es muy probable un decremento

Aunque el parámetro de Hurst esta bien definido matemáticamente, medirlo es problemático. Los datos se medirán a frecuencias de intervalos, grandes/pequeños, donde están contenidas las lecturas menores ó mayores (no lluvia ó huracanes, tormentas), los cuales afectan el cálculo de H .

Otros estimadores son parciales y convergen lentamente de acuerdo a la cantidad de datos disponibles, además son vulnerables a las tendencias y periodicidad en los datos, y quizá, a otras fuentes de ruido. Muchos estimadores asumen formas funcionales específicas con modelos subyacentes y se aplican erróneamente al querer simplificar un trabajo.

Esta método, en nuestra investigación, se eligió porque el análisis R/S es una técnica bien conocida, que se ha utilizado durante algún tiempo al efectuar mediciones del parámetro de Hurst.

Estimar el valor de H , puede ser un paso importante en el análisis de series de tiempo real. Nos permite clasificar (por lo menos aproximadamente) la serie según su correlación a largo plazo dependiendo de el número de particiones en la serie.

Mendelbrot[14] da una justificación formal para el uso de esta prueba.

La estimación del exponente de Hurst, puede obtenerse como se muestra en la siguiente subsección.

2.6.1. Mediciones del exponente de Hurst

Hurst simple (Hs): Una serie de tiempo de longitud completa N no se divide en un número de series más cortas, aquí $n = N$. Solo calculamos el rango redimensionado medio. Esto es, tomamos la serie completa y calculamos el coeficiente como una sola partición.

Hurst simple corregido (Hsc): Este se calcula efectuando el análisis $(R/S) - E[R/S]$.

Hurst empírico (He): Este es el algoritmo de rango reescalado que se explicó al inicio de esta sección.

Hurst ANNIS-LLOYD (Hal): Este cálculo se denomina así ya que es adjudicado a A.A. ANNIS y E.H. LLOYD. Dado que el análisis anterior produce estimaciones sesgadas del exponente que rige esa ley de potencia y que para pequeñas particiones existe una desviación significativa, encuentran una mejor estimación para calcular esa ley de potencias y es efectuando el cálculo siguiente:

$$[E(R/S)] = \begin{cases} \frac{\Gamma(\frac{n-1}{2})}{\sqrt{\pi}\Gamma(\frac{n}{2})} \sum_{i=1}^{n-1} \sqrt{\frac{n-i}{i}}, & \text{para } n \leq 340 \\ \frac{1}{\sqrt{n\frac{\pi}{2}}} \sum_{i=1}^{n-1} \sqrt{\frac{n-i}{i}}, & \text{para } n > 340 \end{cases} \quad (2.14)$$

dónde Γ es la función gamma de Euler. En nuestro caso de estudio encontraremos estos cuatro coeficientes y utilizaremos el promedio de ellos.

Hurst teórico (Ht):

2.7. Identificación de regiones homogéneas

La identificación de regiones homogéneas es a menudo la etapa más complicada, al requerir de la toma de decisiones subjetivas.

En zonas montañosas, las características fisiográficas influyen en la distribución espacial irregular de las variables climáticas. Por lo que en estos sitios la división en regiones se debe realizar teniendo en cuenta tanto aspectos hidroclimáticos como fisiográficos, sin tomar en cuenta la continuidad geográfica de la cuenca. Nathan et al. (1990), mencionaron

que regiones homogéneas definidas por la similitud hidrológica de las cuencas o las características de éstas, pueden no tener significancia geográfica.

La Cuenca del río San Juan, es una región de relieve montañoso donde actualmente existen más de 80 estaciones hidrométricas en operación. El objetivo de este trabajo fue identificar zonas homogéneas dentro de la Cuenca, mediante la aplicación de técnicas estadísticas. Además, se obtuvo una clasificación de estaciones con comportamiento similar.

Esta clasificación se pueden llevar a cabo utilizando distintos tipos de datos y empleando métodos univariantes o multivariantes, lo cual varía en función del tipo de zonificación que se quiere encontrar en relación a una o más variables de interés.

La evaluación comparativa basada en la similitud estructural permite identificar regiones a través de comparaciones sistemáticas con otros y así marcar una diferencia. Esta metodología de identificación de regiones homogéneas, permite la evaluación comparativa y puede ser de gran ayuda en la toma de decisiones estratégicas.

En este trabajo se presenta un método para la detección de zonas homogéneas a través de un análisis fractal sobre los datos que caracterizan las series de tiempo de las 31 estaciones pluviométricas de la Cuenca del Río San Juan.

El objetivo en esta investigación es formar conglomerados de estaciones que satisfagan aproximadamente la condición de homogeneidad, esto es, que podamos decir de alguna manera que poseen la misma distribución de frecuencias, excepto por un factor de escala.

Capítulo 3

Presentación de resultados

3.1. Discusión del problema

Cuando se dispone de observaciones limitadas de eventos hidrológicos, se ve comprometida la capacidad de proveer una adecuada caracterización, análisis y predicciones de un fenómeno. Sin embargo, el análisis puede mejorarse mediante la identificación de muestras homogéneas que pueden usarse en combinación para hacer mejores estimaciones de un modelo de probabilidad.

Esta es una de las principales preocupaciones dentro de la práctica del análisis regional de frecuencia (ARF), donde el resultado final es obtener una estimación de eventos extremos dentro de una zona geográfica, la cual se puede utilizar como entrada en un análisis de riesgos, administración del agua, zonificación y aplicación del agua a usos de la tierra, Hosking y Wallis[5].

Sin embargo, la estimación de eventos extremos se considera un problema complejo, sobre todo porque la información es generalmente limitada, existe correlación serial, muchos puntos de cambio que pudieran estar presentes y observaciones a seguir sobre tendencias y patrones estacionales.

Para enfrentar estos problemas, muchos estudios de series de tiempo hidrológicas han sido aplicado con éxito en el pasado Machiwal y JAI[24], sin embargo otros esfuerzos de investigación habían centrado su atención en pruebas de detección de tendencia, dejando a un lado otras propiedades importantes tales como de estacionalidad, la homogeneidad, la periodicidad y la persistencia. Al abordar estas propiedades, sería posible una mejor selección de muestras homogéneas, y como consecuencia, los profesionales podrían lograr mejores predicciones.

Trabajos previos sobre el análisis de series de tiempo en climatología con aplicaciones en la precipitación se remontan a Bhuiya[25], con el desarrollo de una prueba de estacionariedad después que componentes periódicos y de tendencia se restaron de la serie hidrológica.

Otras pruebas de tendencia de Buishand[26] las utilizó para evaluar la diferencia de precipitación entre zonas rurales y urbanas de Amsterdam y Rotterdam. Buishand[27,28] construyó varias pruebas de homogeneidad y la media de la serie usando las sumas acumulativas, pruebas probabilísticas e inferencia bayesiana. Kothyari[29], et al. evaluó tres estaciones en la India, Agra, Dehradun y Delhi para probar cambios en la precipitación y temperatura, proporcionando la evidencia de un cambio en el número de días de lluvia durante la temporada de monzones y un incremento en la temperatura.

Giakoumakis y Baloutsos[30] realizó un análisis de tendencia en la serie histórica de precipitaciones anuales de la cuenca del Eveno Riven en Grecia. Aplicando diferentes pruebas de aleatoriedad, disminuyendo las tendencias encontradas en los registros de precipitación. Otros autores relacionados con análisis de tendencias, puntos de cambio encontradas en la literatura y homogeneidad son Ángel y Huff [31], Mirza[32], et al., Tarhule y Woo[33], De Lués[34] et al., Kripalani, Adamowski y Kulkarni[35], Bougadis [36], Yu[37] et al., Kumar[38], et al.

Una revisión exhaustiva de estos trabajos puede encontrarse en Machiwal y JAI[24] con descripciones de acontecimientos relacionados con en análisis de series de tiempo hidrológicas.

El exponente de Hurst, como parte de un análisis fractal, se utilizó para evaluar la dependencia de largo alcance y la posibilidad de tendencias en los datos. Otras obras incluyen el exponente de Hurst como los desarrollos de Golder[41], et al., donde utiliza el exponente de Hurst para explorar las correlaciones a largo plazo, y observaciones de precipitación acumulada las modelan usando la ley de probabilidad *alpha-estable* para hacer frente a distribuciones de colas pesadas.

Chang[42] extendió la aplicación del exponente de Hurst mediante el desarrollo de un enfoque de cálculo estimandolo sobre series de tiempo comparadas con un movimiento browniano fraccional de tiempo discreto o ruido fraccional Gaussiano. Yu[43], et al., también estudió correlaciones a largo plazo usando al exponente de Hurst y realizó un análisis fractal múltiple o multifractal de las series de precipitación (véase Kantelhardt[44]) basado en un modelo de cascada multiplicativa y un análisis multifractal de fluctuación de poblamientos.

En este trabajo, llevamos a cabo un enfoque de agrupamiento que utilizamos para concentrar las estaciones en muestras homogéneas, después de resumir los resultados del análisis fractal utilizando el exponente de Hurst tanto a datos directos como a semivariogramas.

Como caso de estudio, tomamos una muestra de 33 estaciones pluviométricas, de la cuenca del Río Bravo-San Juan, en la región RH-24 de México (Fig.3.1).

Un mapa de la cuenca del Río Bravo-San Juan se muestra en la figura 3.2. Esta región se encuentra en México entre los Estados de Nuevo León Coahuila y Tamaulipas con una superficie aproximada de 29420 km^2 . Algunas de las estaciones de lluvias se muestran en la figura 3.3. Los datos utilizados ha sido proporcionados por la CONAGUA, la institución local responsable de la gestión del agua en el país.



FIGURA 3.1: La cuenca del Río San Juan obtenida de Google Earth. Se muestra en negro, la cuenca del Río Bravo-San Juan.

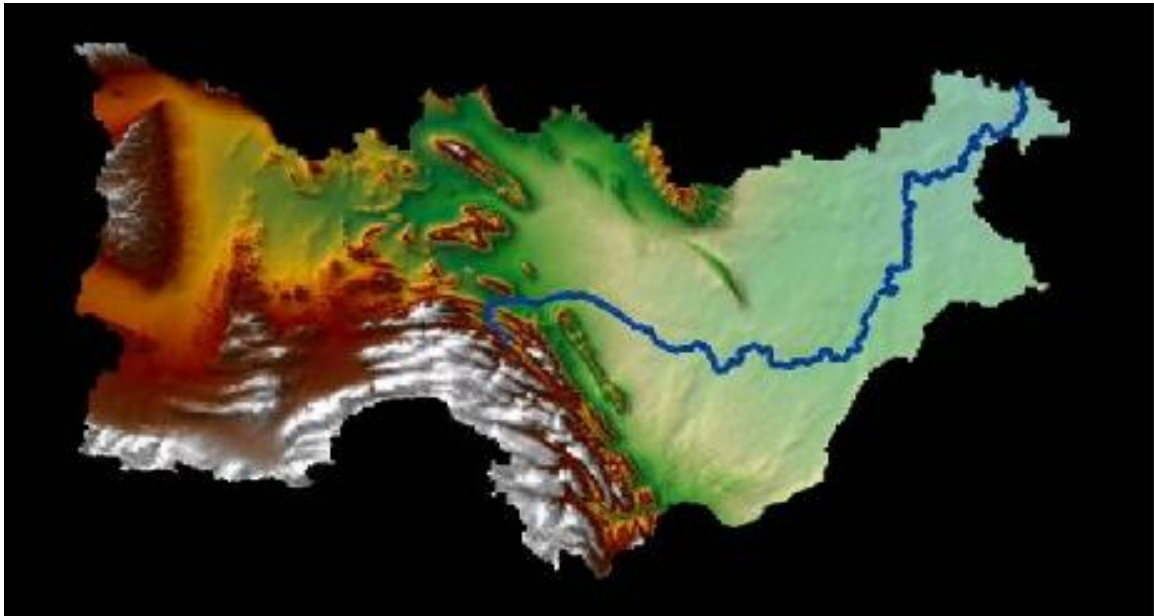


FIGURA 3.2: Cuenca del Río Bravo-San Juan (Río San Juan en azul). Imagen obtenida de: "Administración del agua en la cuenca del Río San Juan, en el sur del Río Bravo Región Hidrológica de México" de <http://earthzine.org/2012/08/13/>

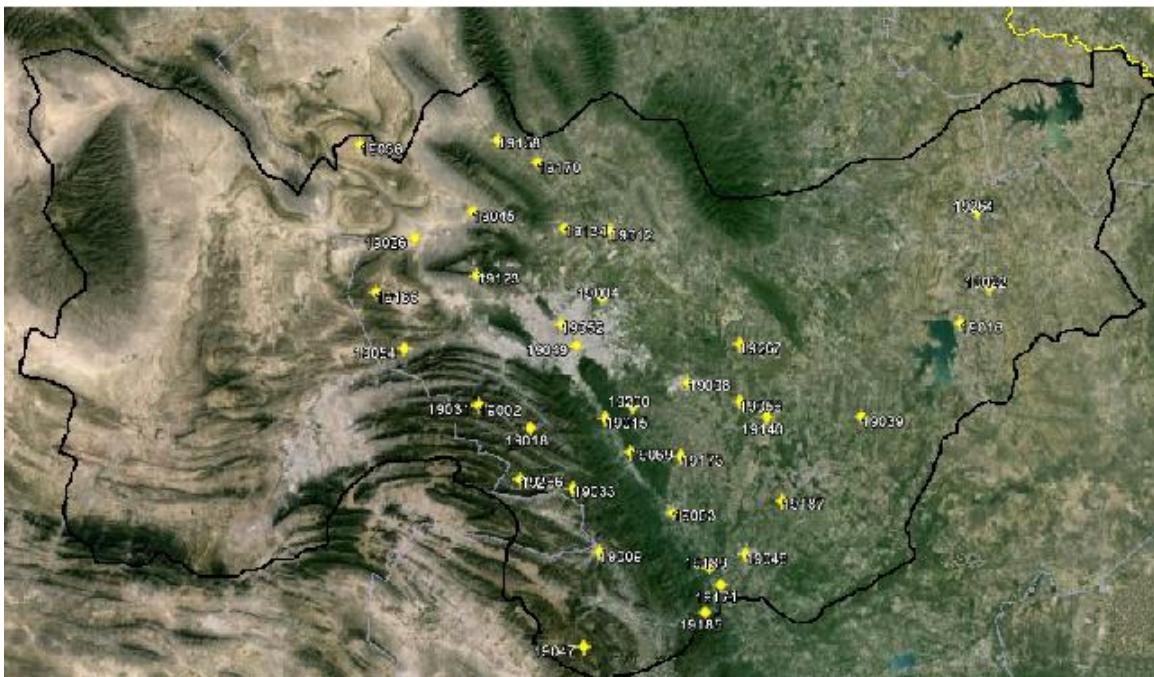


FIGURA 3.3: Localización Geográfica (de Google Earth) de las estaciones pluviométricas de la cuenca del Río Bravo-San Juan. Obtenida de la base de datos: www.conagua.gob.mx

3.2. Descripción del problema

En la práctica, para realizar un análisis regional de frecuencia, se requiere identificar regiones homogéneas donde los datos siguen patrones similares que pueden analizarse conjuntamente para mejorar la identificación de modelos de probabilidad que a su vez puede utilizarse para estimar eventos extremos y su frecuencia en términos de periodos de retorno. Este análisis generalmente se ejecuta cuando se trata con sequías, contaminación, movimiento del viento, temperatura, presión atmosférica y observaciones de precipitación, para nombrar algunos.

En esta investigación nos ocupamos del problema de encontrar grupos de estaciones de lluvias que crean regiones homogéneas, considerando estructuras fractales analizadas a través de los semivariogramas y exponentes de Hurst. Los datos de la precipitación de una muestra 33 estaciones de la región hidrográfica México RH-24, cuenca del Río Bravo-San Juan, fueron utilizadas como estudio de caso para evaluar la propuesta.

Los semivariogramas, se utilizaron como cuantificador de las correlaciones de largo alcance para los datos de las diferentes estaciones pluviométricas con registros mensuales. Teniendo en cuenta el análisis de la muestra de la serie histórica, se realizó un análisis de rango reescalado ó análisis R/S , para obtener una medida del exponente de comportamiento fractal en nuestro caso el exponente de Hurst[48].

Este exponente se utilizó como una medida para cada una de las estaciones pluviométricas en particular. El proceso se repitió para cada estación pluviométrica en la región bajo análisis. El exponentes de Hurst se utilizó como referencia para identificar las estaciones que presentan patrones similares.

Como consecuencia, se aplicó un análisis de clúster para identificar muestras homogéneas.

Una ventaja del exponente de Hurst es la simplicidad de su algoritmo que puede ser utilizado para medir la condición de persistencia o antipersistencia de un proceso, y proporciona una métrica que puede utilizarse para clasificar series de tiempo diferentes.

El semivariograma o variograma $\gamma(h)$ se utilizó para describir la relación de las observaciones pareadas y separadas por una distancia h . Es una técnica geoestadística que permite una medida cuantitativa de la persistencia a largo plazo en series de tiempo no estacionarias Witt[16], Haslett[17], Dmowska[18], et al.

En las correlaciones en el tiempo y el espacio, se crean patrones que pueden utilizarse para describir el comportamiento de un conjunto de observaciones, y el variograma estima la diferencia al cuadrado prevista entre variables aleatorias vecinas. Este cálculo se realiza sobre los valores de h diferentes.

El coeficiente de Hurst calculado con el método del rango reescalado ó también llamado crecimiento del rango, mide el crecimiento de las fluctuaciones de las estaciones, al aumentar el intervalo de tiempo Δt . Esto significa que cuanto mayor es el exponente de Hurst de una población más rápido aumenta el rango de las fluctuaciones. Si no tenemos en cuenta el valor de la constante c de proporcionalidad en la fórmula fractal, para un mismo tamaño poblacional, los valores más grandes del coeficiente de Hurst se podrían asociar a un mayor aumento en los valores o disminución de los mismos (Sugihara y May).

Para estimar el exponente o coeficiente de Hurst de una serie temporal $\{X_k\}$, con $k \in 1, 2, \dots, N$ la serie es dividida en un grupo de d -subseries de longitud m . Realmente, la medida de m es un valor promedio.

Una forma estándar, aunque no la única, de obtener el tamaño m de la subserie, es particionar la serie original en potencias de base 2. De esta manera, en cada una de las particiones sucesivas, el valor aproximado de m es: $N, N/2, N/2^2, N/2^3, \dots$, y así sucesivamente. Para cada subserie $n = 1, 2, \dots, d$, desarrollando los pasos de la sección 8.1.

Dos factores que intervienen en la determinación del coeficiente de Hurst son: la manera en que es dividida la serie de tiempo en grupos de subseries y el comportamiento asintótico del análisis de rango reescalado.

El primero, el rango de valores m se utilizan para calcular la pendiente de $\log(\langle R/S \rangle_m)$, dada la relación

$$\log\left(\left\langle \frac{R}{S} \right\rangle_m\right) = \log(c) + H \log(m) \quad (3.1)$$

El segundo, la determinación de H es el resultado del comportamiento asintótico del rango reescalado, i.e., cuando el valor de m tiende a infinito.

El análisis de reescaldado $\langle R/S \rangle_m$ sobre algunos valores de m es estimado usando log/log expresión dada en (3.1). Para obtener el coeficiente H , se utiliza el método de mínimos cuadrados. La pendiente es el coeficiente de Hurst, H .

Este exponente es considerado un índice fractal, Mandelbrot and Wallis[13], y proporciona información acerca de correlaciones a largo plazo de una serie de observaciones;

para una revisión teórica del exponente de Hurst véase Mandelbrot[14].

En la práctica, el exponente de Hurst puede tomar valores entre 0 y 1, donde:

- $0 < H < 0,5$ indica no persistencia en una serie, i.e., a un incremento es más probable que sea seguida por un decremento y viceversa.
- $H = 0.5$ indica ausencia de correlación serial (ruido blanco Gaussiano).
- $0,5 < H < 1$ indica persistencia, es decir, un incremento corre el riesgo de ser seguido por un incremento y un decremento por otro decremento.

3.3. Resultados

Para ilustrar el procedimiento, se mostrará el análisis de una estación de lluvia que seleccionamos de entre las 33 estaciones para presentarla en esta sección el resto de ellas se mostrara en el anexo 4.

Los valores medidos de precipitación mensual en milímetros de las estaciones de Apodaca, El Cuchillo y La Boca se muestran en la Fig. 3.4, y se presentan los resultados obtenidos teniendo en cuenta todas las estaciones (siguiendo el mismo proceso) en el anexo 2 de esta tesis.

Como puede verse, los patrones y las relaciones entre diferentes estaciones son difíciles de evaluar sólo a través un análisis “visual”. Sin embargo, cuando se obtiene el semivariogram de cada una de estas estaciones, como se muestra en la Fig. 3, una huella de los datos, se hace más evidente. Una inspección más cercana en cada estación muestra un patrón estacional que se repite cada 12 observaciones en el semivariograma. Esto puede explicarse debido al hecho de que se utilizaron las observaciones mensuales en el análisis.

Una vez que se obtuvieron los semivariogramas, se calculó el exponente de Hurst para la serie de los valores de γ . Se puede observar que los coeficientes de Hurst están cercanos a 1, lo que indica una dependencia larga positiva de los datos en los semivariogramas.

El exponentes de Hurst de todas las estaciones bajo análisis se presentan en la tabla ??, en el que el tiempo de dependencia en todos variogramas se hace evidente. Algunos de los coeficientes que aparecen en la tabla exceden el intervalo establecido de los posibles valores del exponente Hurst, $0 < H < 1$, este es un error conocido debido al sesgo de la estimación o un posible error de cálculo en la regresión lineal.

Para abordar el problema de encontrar muestras homogéneas, un análisis de conglomerados se llevó a cabo, utilizando las estimaciones del exponente de Hurst. Como se muestra en la Fig. 3.7, se utilizó un histograma de frecuencias de las estaciones separadas en 6 grupos. Estos clústers y el resultado de las distribuciones de probabilidad de ajuste a través de los datos de cada estación de lluvias se muestran en la tabla 3.2.

Se seleccionaron las distribuciones basándonos en un análisis de bondad de ajuste. Después de identificar un conjunto de distribuciones posibles con p -valores más grandes que un nivel significativo de 0,05, en todos los casos, se seleccionó la distribución con el más alto promedio p -valor para cada estación. Gamma y General de Valores Extremos eran las distribuciones que dieron el mejor ajuste en los datos analizados. Los parámetros de

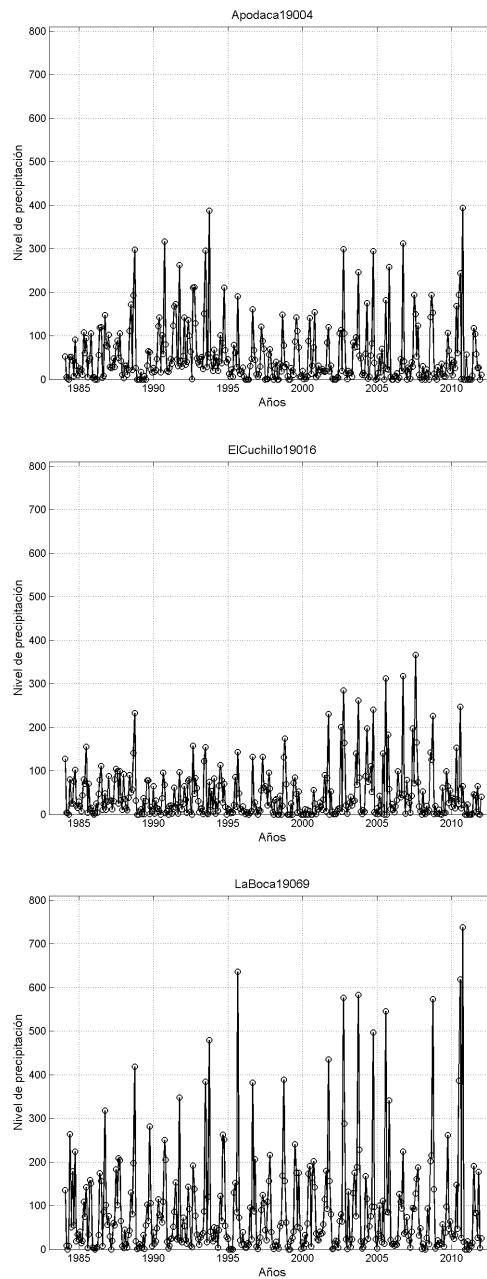


FIGURA 3.4: Mediciones de precipitación de tres estaciones de La cuenca del Río Bravo-San Juan, México. Desde la parte superior a la parte inferior, respectivamente: Apodaca, Estación número 19004, Enero 1940-Diciembre 2012; El Cuchillo, Estación número 19016, Enero 1939-Diciembre 2012 ; La Boca, Estación número 19069, Enero 1923-Diciembre 2012.

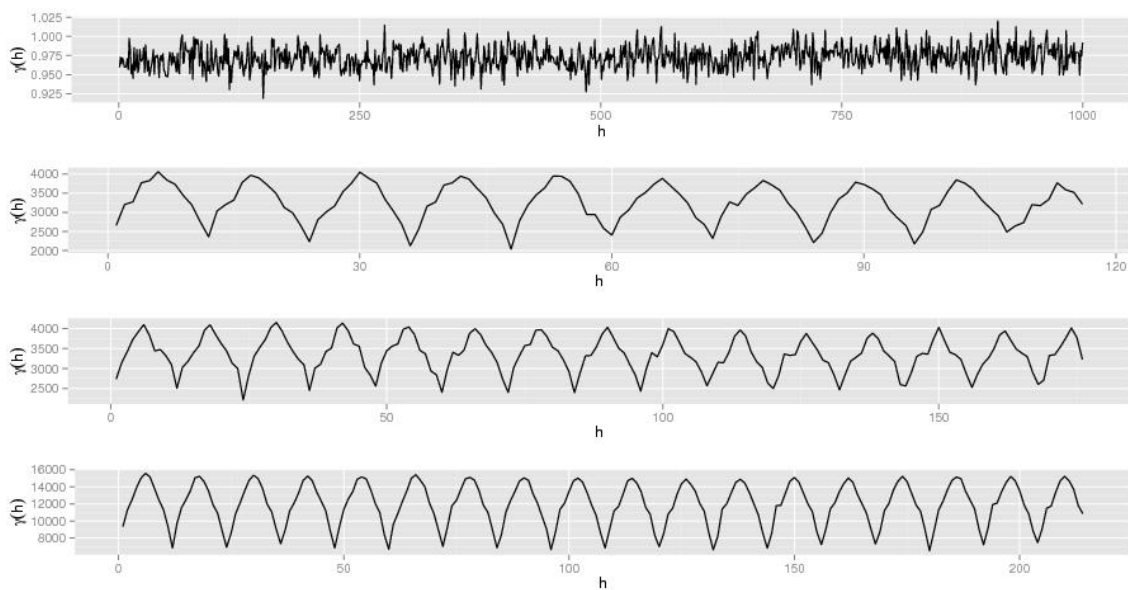


FIGURA 3.5: Semivariogramas correspondientes a la medición de las series temporales de lluvia de las tres estaciones de la cuenca del Río Bravo-San Juan se muestran en 3.4. En cada caso el máximo lag $h_{max} = 0,20*N$ siendo N el tamaño de las series temporales. Para la comparación, se incluye el semivariogram correspondiente al ruido blanco gaussiano. Desde la parte superior a la inferior, respectivamente: ruido blanco Gaussiano, Apodaca(Estación número 19004), El Cuchillo(Estación número 19016), La Boca(Estación número 19069).

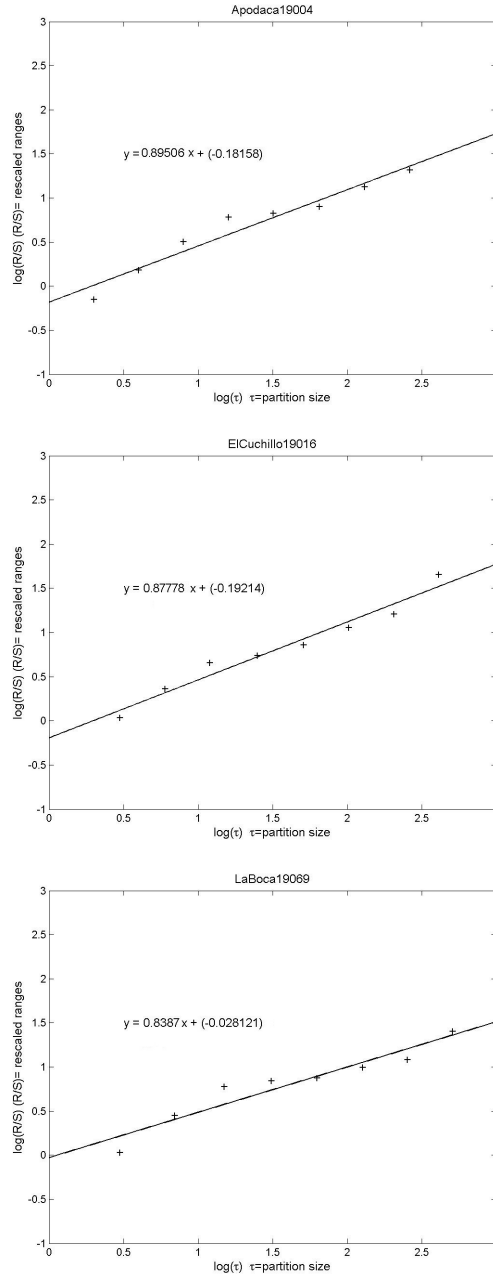


FIGURA 3.6: Mediciones del exponente de Hurst de tres estaciones de La cuenca del Río Bravo-San Juan, México. El valor de la pendiente es el exponente de Hurst. Calculado por mínimos cuadrados. Desde la parte superior a la parte inferior, respectivamente: Apodaca, Estación número 19004, Enero 1940-Diciembre 2012, $m = 0,6375$; El Cuchillo, Estación número 19016, Enero 1939-Diciembre 2012, $m = 0,6551$; La Boca, Estación número 19069, Enero 1923-Diciembre 2012, $m = 0,5137$.

CUADRO 3.1: Información general de las estaciones pluviométricas de la cuenca Río Bravo-San Juan.

| Estación | Nombre | Latitud | Longitud | Datos |
|----------|----------------------|----------|-----------|-----------|
| 19015 | El Cerrito | 25 30 36 | 100 11 36 | 1984-2012 |
| 19039 | Las Enramadas | 25 30 05 | 099 31 17 | 1984-2012 |
| 19018 | El Pajonal | 25 29 23 | 100 23 20 | 1984-2012 |
| 19012 | Ciénega de Flores | 25 57 08 | 100 10 20 | 1984-2012 |
| 19003 | Allende | 25 17 01 | 100 01 13 | 1984-2012 |
| 19069 | La Boca | 25 25 46 | 100 07 44 | 1984-2012 |
| 19009 | Casillas | 25 11 47 | 100 12 51 | 1984-2012 |
| 19187 | California | 25 18 23 | 099 44 02 | 1984-2012 |
| 19173 | Palmitos | 25 25 02 | 099 59 50 | 1984-2012 |
| 19002 | Agua Blanca | 25 32 39 | 100 31 23 | 1984-2012 |
| 19134 | Salinas Victoria | 25 57 33 | 100 17 34 | 1984-2012 |
| 19031 | La Cruz | 25 32 47 | 100 31 23 | 1984-2012 |
| 19036 | La Popa | 26 09 50 | 100 49 40 | 1984-2012 |
| 19185 | El Canadá | 25 02 48 | 099 56 29 | 1984-2012 |
| 19056 | San Juan | 25 32 36 | 099 50 25 | 1944-2012 |
| 19016 | El Cuchillo | 25 43 05 | 099 15 21 | 1960-2012 |
| 19052 | Monterrey(Obs) | 25 44 01 | 100 16 01 | 1984-2012 |
| 19026 | Icamole | 25 56 28 | 100 41 13 | 1984-2012 |
| 19200 | La Cienega | 25 32 10 | 100 07 15 | 1984-2012 |
| 19004 | Apodaca | 25 47 37 | 100 11 50 | 1984-2012 |
| 19047 | Mimbres | 24 58 26 | 100 15 31 | 1984-2012 |
| 19140 | Tepehuaje | 25 30 19 | 099 46 15 | 1984-2012 |
| 19048 | Montemorelos | 25 10 55 | 099 49 56 | 1984-2012 |
| 19022 | General Bravo | 25 48 05 | 099 10 32 | 1984-2012 |
| 19158 | Rancho de Gomas | 26 10 11 | 100 27 52 | 1984-2012 |
| 19045 | Mina | 26 00 08 | 100 32 00 | 1984-2012 |
| 19054 | Rinconada | 25 40 52 | 100 43 03 | 1984-2012 |
| 19165 | Chupaderos del Indio | 25 48 49 | 100 47 24 | 1984-2012 |
| 19171 | Lampacitos | 25 06 38 | 099 53 57 | 1984-2012 |
| 19264 | Dr. Coss | 25 51 16 | 099 56 36 | 1984-2012 |
| 19170 | El Hojase | 26 06 55 | 100 21 38 | 1984-2012 |
| 19033 | Laguna de Sanchez | 26 06 55 | 100 21 38 | 1984-2012 |
| 19123 | Grutas de García | 26 06 55 | 100 21 38 | 1984-2012 |

CUADRO 3.2: Coeficientes de Hurst Simple(He), Hurst simple corregido (Hsc), Hurst empírico (He), Hurstal(Hal), Hurst Teórico (Ht) y Hurst Promedio (HP). Mediante (6.4) de estaciones pluviométricas de la cuenca Río Bravo-San Juan.

| Estación | Hs | Hsc | He | Hal | Ht | HP |
|----------|--------|--------|--------|--------|--------|--------|
| 19002 | 0.6277 | 0.6389 | 0.4915 | 0.4531 | 0.5429 | 0.5508 |
| 19003 | 0.5721 | 0.5902 | 0.6291 | 0.5871 | 0.5379 | 0.5833 |
| 19004 | 0.5507 | 0.5457 | 0.5333 | 0.4920 | 0.5419 | 0.5327 |
| 19009 | 0.4993 | 0.4811 | 0.5102 | 0.4710 | 0.5407 | 0.5407 |
| 19012 | 0.6517 | 0.6755 | 0.7254 | 0.6784 | 0.5397 | 0.6541 |
| 19015 | 0.4397 | 0.3595 | 0.3616 | 0.3281 | 0.5410 | 0.4060 |
| 19016 | 0.5640 | 0.5859 | 0.6120 | 0.5719 | 0.5392 | 0.5746 |
| 19018 | 0.5389 | 0.5422 | 0.5780 | 0.5388 | 0.5417 | 0.5479 |
| 19022 | 0.5218 | 0.5470 | 0.6097 | 0.5721 | 0.5382 | 0.5578 |
| 19026 | 0.5599 | 0.5859 | 0.6217 | 0.5829 | 0.5389 | 0.5779 |
| 19031 | 0.6630 | 0.7468 | 0.6127 | 0.5705 | 0.5416 | 0.6269 |
| 19033 | 0.5783 | 0.5734 | 0.5800 | 0.5365 | 0.5413 | 0.5619 |
| 19036 | 0.6282 | 0.6948 | 0.6199 | 0.5831 | 0.5426 | 0.6137 |
| 19039 | 0.5663 | 0.5874 | 0.6885 | 0.6410 | 0.5397 | 0.6046 |
| 19047 | 0.5386 | 0.5543 | 0.6493 | 0.6093 | 0.5407 | 0.5785 |
| 19048 | 0.5276 | 0.5416 | 0.6101 | 0.5677 | 0.5397 | 0.5573 |
| 19052 | 0.5561 | 0.6045 | 0.7051 | 0.6605 | 0.5360 | 0.6125 |
| 19054 | 0.5908 | 0.6309 | 0.6252 | 0.5836 | 0.5412 | 0.5944 |
| 19056 | 0.6011 | 0.6334 | 0.6791 | 0.6338 | 0.5392 | 0.6173 |
| 19069 | 0.5342 | 0.5144 | 0.5319 | 0.4505 | 0.5384 | 0.5219 |
| 19123 | 0.6725 | 0.7653 | 0.8554 | 0.7965 | 0.5479 | 0.7275 |
| 19140 | 0.5842 | 0.6183 | 0.6538 | 0.6100 | 0.5466 | 0.6026 |
| 19158 | 0.5787 | 0.5818 | 0.4499 | 0.4106 | 0.5461 | 0.5134 |
| 19165-1 | 0.6011 | 0.6334 | 0.6791 | 0.6338 | 0.5392 | 0.6173 |
| 19165-2 | 0.5877 | 0.6015 | 0.5377 | 0.4915 | 0.5479 | 0.5533 |
| 19170 | 0.6033 | 0.6449 | 0.6400 | 0.5905 | 0.5479 | 0.6054 |
| 19171 | 0.5310 | 0.5412 | 0.5626 | 0.5117 | 0.5483 | 0.5390 |
| 19173-1 | 0.5998 | 0.6376 | 0.6389 | 0.5908 | 0.5467 | 0.6027 |
| 19173-2 | 0.5998 | 0.6376 | 0.6389 | 0.5908 | 0.5467 | 0.6027 |
| 19185 | 0.5418 | 0.5553 | 0.6131 | 0.5605 | 0.5469 | 0.5635 |
| 19187 | 0.5435 | 0.5155 | 0.4812 | 0.4385 | 0.5479 | 0.5053 |
| 19261 | 0.5823 | 0.6100 | 0.6204 | 0.5707 | 0.5461 | 0.5859 |
| 19264-1 | 0.6018 | 0.6399 | 0.6356 | 0.5863 | 0.5479 | 0.6023 |
| 19264-2 | 0.5946 | 0.6360 | 0.5342 | 0.4912 | 0.5460 | 0.5604 |

CUADRO 3.3: Clusterización de las estaciones pluviométricas mediante la aplicación de variogramas y el exponente de Hurst.

| Clúster | Estación | Hurst | Distribución | Parámetros |
|---------|----------|-------------------------|---------------------------|---------------------------|
| 3 | 19069 | 0.463980 | Gamma(α, β) | (0.53024, 161.53) |
| | 19015 | 0.425844 | Gamma(α, β) | (0.56713, 135.9793) |
| | 19045 | 0.550170 | Gamma(α, β) | (0.43528, 64.87) |
| 19 | 19002 | 0.51633 | GEV(k, σ, μ) | (0.375, 31.409, 22.699) |
| | 19018 | 0.51834 | GEV(k, σ, μ) | (0.40419, 23.903, 15.742) |
| | 19056 | 0.57269 | Gamma(α, β) | (0.66796, 89.45) |
| | 19171 | 0.53847 | GEV(k, σ, μ) | (0.45234, 31.257, 20.775) |
| | 19009 | 0.48223 | GEV(k, σ, μ) | (0.41444, 22.618, 12.778) |
| | 19165 | 0.55299 | GEV(k, σ, μ) | (0.48155, 9.733, 5.2045) |
| | 19185 | 0.543203 | Gamma(α, β) | (0.59620, 70.85) |
| | 19047 | 0.530818 | GEV(k, σ, μ) | (0.22971, 31.208, 22.582) |
| | 19187 | 0.54467 | Gamma(α, β) | (0.63920, 95.31) |
| | 19134 | 0.55995 | Gamma(α, β) | (0.70652, 55.42) |
| | 19048 | 0.54338 | Gamma(α, β) | (0.54095, 134.70) |
| | 19052 | 0.54884 | GEV(k, σ, μ) | (0.4252, 24.947, 18.532) |
| | 19158 | 0.55010 | GEV(k, σ, μ) | (0.41386, 17.099, 11.222) |
| | 19003 | 0.560854 | GEV(k, σ, μ) | (0.39327, 38.646, 31.372) |
| | 19031 | 0.55591 | GEV(k, σ, μ) | (0.31561, 36.006, 24.569) |
| | 19054 | 0.51812 | GEV(k, σ, μ) | (0.46328, 9.9797, 5.5734) |
| | 19012 | 0.52662 | Gamma(α, β) | (0.69427, 65.16) |
| 19039 | 0.56713 | GEV(k, σ, μ) | (0.4426, 26.541, 15.9) | |
| 19033 | 0.56747 | GEV(k, σ, μ) | (0.38319, 32.435, 21.751) | |
| 6 | 19200 | 0.57142 | Gamma(α, β) | (0.54323, 119.49) |
| | 19140 | 0.59096 | Gamma(α, β) | (0.58043, 98.07) |
| | 19004 | 0.58956 | GEV(k, σ, μ) | (0.36706, 28.543, 20.751) |
| | 19170 | 0.58586 | GEV(k, σ, μ) | (0.49082, 18.836, 11.01) |
| | 19022 | 0.59508 | Gamma(α, β) | (0.615691, 72.00) |
| | 19173 | 0.59711 | Gamma(α, β) | (0.53839, 118.69) |
| 2 | 19026 | 0.66907 | GEV(k, σ, μ) | (0.51724, 6.8559, 3.2313) |
| | 19123 | 0.65470 | GEV(k, σ, μ) | (0.56217, 16.124, 8.2436) |
| 2 | 19036 | 0.56404 | GEV(k, σ, μ) | (0.60326, 6.0381, 2.4608) |
| | 19016 | 0.61413 | Gamma(α, β) | (0.55195, 79.17) |
| 1 | 19264 | 0.55037 | GEV(k, σ, μ) | (0.47709, 16.441, 8.179) |

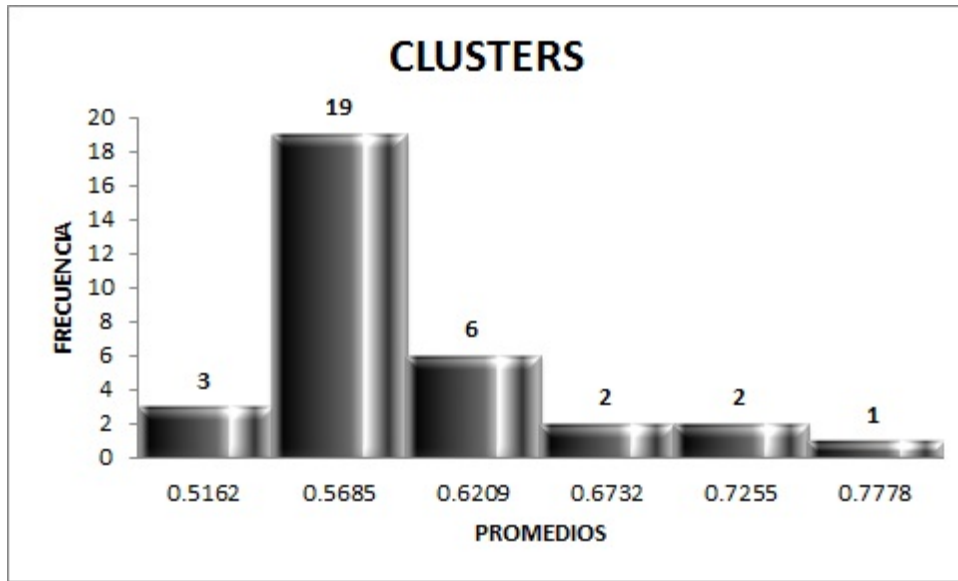


FIGURA 3.7: Histograma de las estaciones pluviométricas utilizando el exponente de Hurst.

ubicación, dispersión, asimetría y curtosis, fueron calculados con los L-momentos. Estas funciones son:

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}}{\beta^{\alpha} \Gamma(\alpha)} e^{-\frac{x}{\beta}}, \quad (3.2)$$

y

$$f(x; k, \sigma, \mu) = \frac{1}{\sigma} e^{-(1+kz)^{-1/k}} (1+kz)^{-1-1/k}, \quad z = \frac{x - \mu}{\sigma} \quad (3.3)$$

respectivamente.

Como resultado del análisis, las distribuciones estimadas no se mezclan dentro de clúster. Esta entrada puede ser utilizada en un análisis posterior para obtener estimadores de máxima verosimilitud de los parámetros de distribución utilizando todos los datos al mismo tiempo.

Capítulo 4

Conclusiones y trabajo futuro

Conclusiones

Como caso de estudio para desarrollar esta investigación, utilizamos una muestra de estaciones de lluvia, las cuales están contenidas en la cuenca del Río Bravo-San Juan de la región hidrográfica RH-24 México. Fueron utilizadas 33 de las 41 estaciones en el análisis. En esta muestra se revisaron las series de tiempo y se le aplicó la técnica de Variogramas, posterior a esto, se analizó su fractalidad o exponente de Hurst.

El mapa de Variogramas brinda una forma muy eficiente y automática de determinar si un conjunto de datos presenta un comportamiento que puede ser claramente observado y además seguido del análisis de R/S o estimación del exponente de Hurst, revela mayor información utilizable como herramienta de un análisis de conglomerados.

El exponente de Hurst, provee una medida para determinar si una serie de tiempo es ruido blanco gaussiano o tiene una tendencia subyacente y, puede ser utilizada para agrupar las estaciones pluviométricas de acuerdo con a los valores obtenidos al aplicarse el análisis de reescalado a los semivariogramas.

La dependencia de largo alcance se encuentra en cada uno de los variogramas evaluados con el exponente de Hurst, sin embargo, se encontró todavía más útil como una herramienta para un análisis de conglomerados.

Una bondad del proceso de ajuste se ejecuta con cada serie, y los resultados mostraron que existen distribuciones dominantes dentro de un conjunto factible (que se encuentra de forma independiente en cada estación) no se superponen sobre las agrupaciones.

Se encontró que las distribuciones de probabilidad anidan dentro de cada grupo. Esto es indicativo de que los patrones homogéneos fueron identificados dentro de los grupos y, los grupos fueron heterogéneos entre sí.

El estudio de las estaciones de lluvia con semivariogramas y el análisis R/S proporciona una poderosa herramienta que permite a los profesionales encontrar correlaciones a largo plazo y además analizar la agrupación de las series hidrológicas.

El análisis de agrupamiento se utiliza, intentando identificar las estaciones dentro de la cuenca con similar régimen hidrológico.

Los parámetros de ubicación, dispersión, asimetría y curtosis de las distribuciones ajustadas a cada conglomerado, se obtuvieron utilizando L-momentos, los cuales dan una mejor caracterización de la distribución de probabilidad de nuestras muestras.

Se identificaron seis regiones de la cuenca, que son hidrológicamente más parecidas entre sí que a captaciones en otros lugares pero, no del mismo modo distintivas entre sí ya que sus distribuciones son, generalmente, distribuciones; Gamma y GEV.

En trabajos futuros, un análisis de L -momentos, podrá ser utilizado en el estudio de ondas espectrales y para mejorar la comprensión de las complejas series de tiempo de los niveles de precipitaciones pluviométricas.

Si se considerasen más factores de igual importancia relacionados a la hidrología del área, igualmente importantes, sería como un complejo mosaico de zonas hidrológicamente homogéneas, y más difícil relacionarlas a una gran escala realista, pero mediante este tipo de análisis, se podría hacer una mejor custerización.

Capítulo 5

Bibliografía

Bibliografía

- [1] Soley, F. Javier. Descripción de dos métodos de rellenado de datos ausentes en series de tiempo meteorológicas. *Revista de Matemática: Teoría y Aplicaciones*, vol. 16, núm. 1, enero-junio, 2009, pp. 60-75
- [2] Georges Matheron. *Curso de Geostatística*.
- [3] Hurst, H. E. (1951). Long-Term Storage Capacity of Reservoirs. *Transactions of the American Society of Civil Engineers*, 116(1), 770-799.
- [4] Tate Dalrymple. Flood-frequency analyses, *Manual of Hydrology: Part 3 Water Supply Paper 1543-A*
- [5] Hosking, J. R. M., & Wallis, J. R. (2005). *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press.
- [6] Flood frequency analysis. A Ramachandra Rao, Khaled H. Hamed
- [7] Lena M. Tallaksen, Henny A.J. Van Lanen. *Hydrological Drought, processes and estimation methods for streamflow and groundwater*.
- [8] Durrans and Tomic. *Advances in Data-based Approaches for Hydrologic Modeling*.
- [9] Murugesu Sivapalan. *Effects of spatial variability and scale with implications to hydrologic modeling*
- [10] Dominic Mazvimavi. *A review of aspects of hydrological sciences research in Africa over the last decade*
- [11] T. M. L. WIGLEY. *Global temperature variations between 1861 and 1984*
- [12] Andrew W. Lo. *Behaviors as the product of evolution in stochastic environments*.
- [13] Mandelbrot, B. B., & Wallis, J. R. (1969). Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. *Water Resources Research*, 5(5), 967-988.
- [14] Mandelbrot, B. (1972). Statistical methodology for nonperiodic cycles: from the covariance to R/S analysis. In *Annals of Economic and Social Measurement*, 1(3), 259-290.
- [15] Aristid Lindenmayer. *Modelled the behaviour of cells of plants. L-systems nowadays are also used to model whole plants*.
- [16] Witt, A., Malamud, B. D. (2013). Quantification of Long-Range Persistence in Geophysical Time Series: Conventional and Benchmark-Based Improvement Techniques. *Surv Geophys*, 34, 541–651.

- [17] Haslett, J. (2002). On the sample variogram and the sample autocovariance for non-stationary time series. *Journal of the Royal Statistical Society: Series D*, 46(4), 475-484.
- [18] Dmowska, R., Saltzman, B. (Editors) (1999). *Advances in geophysics, Vol 40. Long-Range Persistence in Geophysical Time Serie*. Academic Press, USA.
- [19] Lefevbre, M. (2007). *Applied stochastic processes*. Springer, USA.
- [20] A. I. McKerchar. Application of seasonal parametric linear stochastic models to monthly flow data.
- [21] M. G. Schaefer. Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington State
- [22] M.C. Acreman, C.D. Sinclair. Classification of drainage basins according to their physical characteristics - An application for flood frequency analysis in Scotland. *Journal of Hydrology*, 84, 365- 380
- [23] Nathaniel B. Guttman. 1992-1993 Winter precipitation in southwest Arizona.
- [24] Machiwal, D., & Jha, M. K. (2012). Current Status of Time Series Analysis in Hydrological Sciences. In *Hydrologic Time Series Analysis: Theory and Practice* (pp. 96-136). Springer Netherlands.
- [25] Bhuiya, R. K. (1971). Stochastic analysis of periodic hydrologic process. *Journal of the Hydraulics Division*, 97(7), 949-962.
- [26] Buishand, T. A. (1979). Urbanization and changes in precipitation, a statistical approach. *Journal of Hydrology*, 40(3), 365-375.
- [27] Buishand, T. A. (1982). Some methods for testing the homogeneity of rainfall records. *Journal of hydrology*, 58(1), 11-27.
- [28] Buishand, T. A. (1984). Tests for detecting a shift in the mean of hydrological time series. *Journal of Hydrology*, 73(1), 51-69.
- [29] Kothyari, U. C., Singh, V. P., & Aravamuthan, V. (1997). An investigation of changes in rainfall and temperature regimes of the Ganga Basin in India. *Water resources management*, 11(1), 17-34.
- [30] Giakoumakis, S. G., & Baloutsos, G. (1997). Investigation of trend in hydrological time series of the Evinos River basin. *Hydrological sciences journal*, 42(1), 81-88.
- [31] Ángel, J. R., & Huff, F. A. (1997). Changes in heavy rainfall in Midwestern United States. *Journal of Water Resources Planning and Management*, 123(4), 246-249.
- [32] Mirza, M. Q., Warrick, R. A., Ericksen, N. J., & Kenny, G. J. (1998). Trends and persistence in precipitation in the Ganges, Brahmaputra and Meghna river basins. *Hydrological Sciences Journal*, 43(6), 845-858.
- [33] Tarhule, A., & Woo, M. K. (1998). Changes in rainfall characteristics in northern Nigeria. *International Journal of Climatology*, 18(11), 1261-1271.

- [34] Luis, M. D., Raventós, J., González-Hidalgo, J. C., Sánchez, J. R., & Cortina, J. (2000). Spatial analysis of rainfall trends in the region of Valencia (East Spain). *Int. J. Climatol*, 20(12), 1451-1469.
- [35] Kripalani, R. H., & Kulkarni, A. (2001). Monsoon rainfall variations and teleconnections over South and East Asia. *International Journal of Climatology*, 21(5), 603-616.
- [36] Adamowski, K., & Bougadis, J. (2003). Detection of trends in annual extreme rainfall. *Hydrological Processes*, 17(18), 3547-3560.
- [37] Yu, P. S., Yang, T. C., & Kuo, C. C. (2006). Evaluating long-term trends in annual and seasonal precipitation in Taiwan. *Water Resources Management*, 20(6), 1007-1023.
- [38] Kumar, V., Jain, S. K., & Singh, Y. (2010). Analysis of long-term rainfall trends in India. *Hydrological Sciences Journal–Journal des Sciences Hydrologiques*, 55(4), 484-496.
- [39] Golian, S., Saghafian, B., Sheshangosht, S., & Ghalkhani, H. (2010). Comparison of classification and clustering methods in spatial rainfall pattern recognition at Northern Iran. *Theoretical and applied climatology*, 102(3-4), 319-329.
- [40] Shi, P., Ma, X., Chen, X., Qu, S., & Zhang, Z. (2013). Analysis of variation trends in precipitation in an upstream catchment of Huai River. *Mathematical Problems in Engineering*, 2013.
- [41] Golder, J., Joelson, M., Neel, M. C., & Di Pietro, L. (2014). A time fractional model to represent rainfall process. *Water Science and Engineering*, 7(1), 32-40.
- [42] Chang, Y. C. (2014). Efficiently Implementing the Maximum Likelihood Estimator for Hurst Exponent. *Mathematical Problems in Engineering*, 2014.
- [43] Yu, Z. G., Leung, Y., Chen, Y. D., Zhang, Q., Anh, V., & Zhou, Y. (2014). Multifractal analyses of daily rainfall time series in Pearl River basin of China. *Physica A: Statistical Mechanics and its Applications*, 405, 193-202.
- [44] Kantelhardt, J. W. (2009). Fractal and multifractal time series. In *Encyclopedia of Complexity and Systems Science* (pp. 3754-3779). Springer, New York.
- [45] Carbone, M., Turco, M., Brunetti, G., & Piro, P. (2015). A Cumulative Rainfall Function for Subhourly Design Storm in Mediterranean Urban Areas. *Advances in Meteorology*, 2015.
- [46] Chou, C. M. (2014). Complexity analysis of rainfall and runoff time series based on sample entropy in different temporal scales. *Stochastic Environmental Research and Risk Assessment*, 28(6), 1401-1408.
- [47] García-Marín, A. P., Estévez, J., Medina-Cobo, M. T., & Ayuso-Muñoz, J. L. (2015). Delimiting homogeneous regions using the multifractal properties of validated rainfall data series. *Journal of Hydrology*, 529, 106-119.

Capítulo 6

Anexos

Anexo 1. Ubicación de la cuenca

El marco natural de la cuenca del río San Juan como lo muestra la Fig. 6.1 comprende parte de los estados de Coahuila, Nuevo León y Tamaulipas; así como los diferentes aprovechamientos hidráulicos que en esta existen, tales como el Distrito de Riego 026 Bajo Río San Juan, el Distrito de Riego 031 Las Lajas, el acueducto China-Monterrey, el acueducto regional China-Aldamas y las unidades de riego que se ubican en el Río Pesquería, arroyo Ayancual y Río San Juan.

Esta cuenca tiene una superficie de 32,972 km² y es el segundo afluente en importancia de los aportadores mexicanos al Río Bravo cuya confluencia ocurre a 58 km aguas abajo de la cortina de la presa Internacional Falcón y 383 km aguas arriba de la desembocadura en el Golfo de México.

La cuenca se localiza entre los paralelos 24°50' y 26°25' de latitud Norte y los meridianos 98°45' y 101°59' de longitud Oeste.



FIGURA 6.1: Ubicación de la cuenca: “Google Earth”

Este Río es uno de los más importantes de la región Noreste del país, abarcando territorio de tres estados que son Coahuila con 13,123 km², con 18,860 km y Tamaulipas con

989 km² (Secretaría de Agricultura y Recursos Hidráulicos (SARH), 1973).

La presa Rodrigo Gómez “La Boca”, se construyó en la cabecera del Río San Juan en el municipio de Santiago Nuevo León, es ubicada entre las coordenadas en UTM Nuevo León, X Mínima 383871.168, X Máxima 386780.921, con Y Mínima 2810047.796 y Y Máxima 2815299.763. Cuenta con una superficie aproximada de 455Ha. y un volumen de almacenamiento de 40 millones de metros cúbicos (Mm³). Es alimentada principalmente por el arroyo La Chueca con dirección sureste y capta las aguas de arroyos perennes que descargan de la Sierra Madre Oriental a altitudes entre 2,000 a 2,300 msnm, estos arroyos son: Cavazos, Cristalinas, Dolores, Escamilla, Puerco y San Antonio; continua como efluente después de la compuerta con el nombre de Río San Juan.

Anexo 2. Técnicas de regionalización

En la década de los años 80, da inició formalmente el denominado análisis regional de los datos para predicción en sitios donde no se tiene tal información y obteniendo estimaciones más confiables para aplicarlos en cuencas con datos insuficientes.

Aunque este método da el índice de avenida, data de 20 años antes (Dalrymple, 1960[4]), es hasta finales de los 80 que se hace una exploración de este enfoque, en cuanto a métodos o procesos y sus ventajas (A. Ramachandra, Khaled H. Hamed 1988[6]).

El análisis regional de frecuencias de crecientes (ARFC), permite realizar predicciones, es decir, estimaciones asociadas a una determinada probabilidad de excedencia, con base en todos los datos observados en varias estaciones hidrométricas de una región, y que incluso se utilizan en las pruebas que emplean características climáticas y/o fisiográficas de cuencas (Lena M. and Henny[7]).

El análisis regional de frecuencia es la estimación de cuan a menudo un evento específico puede ocurrir. La estimación de la frecuencia de eventos extremos es de particular importancia, ya que existen numerosas fuentes de incertidumbre sobre los procesos físicos que amplifican a los eventos observados, es así como una aproximación estadística para el análisis de datos es necesaria.

Los métodos estadísticos son nobles con la existencia de incertidumbre, permiten cuantificar sus efectos y los procedimientos para el análisis de frecuencias estadísticas son un agregado singular, que también nos es de utilidad.

Por ejemplo, las observaciones meteorológicas del medio ambiente, de una misma variable, en diferentes lugares de medición con una frecuencia de eventos, tiene similares o diferentes cantidades observadas, entonces la conclusión exacta podría alcanzarse analizando todos los datos de las muestras, usando únicamente una simple muestra.

En aplicaciones ambientales este método es conocido como **ANÁLISIS REGIONAL DE FRECUENCIAS**, donde los datos de las muestras analizadas son típicamente observaciones de la misma variable en un número de lugares de medición dentro de lo que se define como **“REGIÓN”**. Los principios de análisis de frecuencia regional, sin embargo, son aplicables, siempre que las muestras múltiples de los datos similares están disponibles.

Diversas técnicas para llevar a cabo una regionalización hidrológica se han desarrollado para facilitar y hacer más rápido los análisis. Durrans & Tomic[8](1996) indicaron

que ciertas técnicas se pueden clasificar en dos tipos.

La primera, dedicada a la predicción de caídas de agua en cuencas (Sivapalan[9], 2003), en la que la relación de ciertas características hidrológicas (por ejemplo, el pico de descarga o de bajo flujo) con características climáticas y fisiográficas se establecen para medir las cuencas hidrográficas. Esa relación se puede aplicar sin datos en cuencas y predecir características hidrológicas mediante observaciones fisiográficas y características climáticas.

La segunda, la de análisis regional, conocida como análisis regional de frecuencia o análisis de frecuencia regional. Su objetivo es mejorar la estimación en algunos sitios medidos por medio del uso de la información de otros sitios calibrados con datos de períodos más largos en una región homogénea. Hosking y Wallis[5], consideran este método como una forma de “Negociación en espacio-temporal”.

Métodos de regresión múltiple también han sido utilizado para este propósito durante muchos años (p. ej. Mazvimavi[10] et al., 2004).

Con el desarrollo de geotecnología de la información, como los sistemas de información geográfica (SIG) y teledetección, más y más información está disponible actualmente.

Anexo 3. Depuración de Datos

Se importaron los datos desde un origen de datos externo.

Se creó una copia de seguridad de los datos originales en otra carpeta diferente.

Se complementaron los datos en el mismo formato, tabla con filas y columnas:

- Mismo tipo de datos en cada columna
- Todas las columnas y filas visibles
- Se buscó que no hubiese filas en blanco

Se ejecutaron, primero, las tareas que no necesiten la manipulación de columnas. Y posteriormente, las que requiera la manipulación de columnas. Se manipularon las columnas: Se insertaron columnas nuevas junto a las columnas originales, las cuales requerían depuración. Se rellenó la nueva columna, y se seleccionó para copiarla, pegarla.

Al revisar la información de las bases de datos se encontraron:

- Logotipos
- Comas en lugar de puntos
- Espacios entre números
- Letras en la captura
- Falta de información en ellas como S/D, s.d. (sin dato), NaN (Not a Number), INAP (no aplica), etc.
- Información que no corresponde a los datos como notas del capturista o registros del analista.

Esto fue un problema común en todas de las bases de datos. La tabla 1. muestra lo anteriormente expuesto.

Se utilizó librerías de MATLAB, como: xlsread (nombre del archivo, número de hoja). Instrucciones que permiten leer archivos desde Excel, representando valores que no son números reales con un valor especial llamado NaN, que significa “no es un número”. De esta forma, se logra detectar celdas con texto, comas, espacios, etc.

El cuadro 6.1 y 6.2 muestra lo anterior.

Ya depurada la base de datos, mediante interpolación se completó las faltantes y/o se tomaron en cuenta las observaciones de estaciones cercanas para el caso de celdas datos al inicio de la base o al final de esta.

CUADRO 6.1: COMISIÓN NACIONAL DEL AGUA
ORGANISMO DE CUENCA RÍO BRAVO
DIRECCIÓN TÉCNICA
DATOS DE PRECIPITACIÓN MENSUAL EN mm.

| AÑOS | ENE | FEB | MAR | ABR | MAY | JUN | JUL | AGO | SEP | OCT | NOV | DIC |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| 1987 | 50.0 | 63.0 | 35.5 | 68.0 | 79.5 | 356.1 | 146.1 | 241.5 | 356.0 | 69.0 | 13.0 | 9.0 |
| 1988 | 36.3 | 11.8 | 12.5 | 41.0 | 69.5 | 209.0 | 130.5 | 236.0 | 388.0 | 41.0 | 20,8 | 9.0 |
| 1989 | 30.0 | 41.5 | 8.5 | 54.0 | 15.5 | 51.5 | 116,0 | 44.5 | 251.0 | 192.0 | 28.5 | 74.0 |
| 1990 | 17.5 | 21.5 | 47.5 | 122.0 | 141.5 | 16.0 | 101.5 | 58.0 | 316.5 | 80.0 | 19.5 | 17.0 |
| 1991 | 26.5 | 47.5 | 35.5 | 123.5 | 167.5 | 172.5 | 46.0 | 30.5 | 262.0 | 42.0 | 29.5 | 54.0 |
| 1992 | 142.5 | 33.5 | 40.0 | 136.5 | 100.5 | 63.5 | 1.0 | 210.5 | 212.0 | 129.0 | 48.0 | 39.0 |
| 1993 | 33.5 | 46.5 | 51.5 | 23.8 | 151.0 | 298.5 | 28.0 | 58.5 | 367.5 | 50.5 | 31.0 | 19.5 |
| 1994 | 67.5 | 31.0 | 50.5 | 19.5 | 41.5 | 101.0 | 36.5 | 83.5 | 210.3 | 67.2 | 39.0 | 46.0 |
| 1995 | 28.5 | 27.0 | 50.0 | S/D | S/D | S/D | s/d | S/D | S/D | S/D | S/D | S/D |
| 1996 | S/D | S/D | S/D | S/D | S/D | S/D | S/D | S/D | 30.1 | 179.1 | 7.2 | 10.2 |
| 1997 | 18.7 | 61.8 | 72.2 | 121.2 | 128.8 | 68.1 | 21.4 | 69.3 | 42.8 | 226.4 | 36.7 | 14.8 |
| 1998 | 7.7 | 17.9 | 13.7 | 9.0 | 0.0 | 12.5 | 28.3 | 262.2 | 232.0 | 38.7 | 63.3 | 5.2 |
| 1999 | 0,0 | 0.0 | 25.8 | 25.7 | 34.1 | 103.9 | 145.3 | 0 | 72.4 | 25.7 | 0.0 | 14,3 |
| 2000 | 0.0 | 33.9 | 12.6 | 13.7 | 100.1 | 104.7 | 20.5 | 93.3 | 72.1 | 279.1 | 44.5 | 2 |
| 2001 | 47.1 | 5.3 | 68.7 | 78.0 | 15.4 | 99.6 | 25.9 | 25.8 | 564.9 | 42.0 | 131.2 | 24.4 |
| 2002 | 4.3 | S/D | INAP | INap | INAP | S/D | 118.6 | 93.2 | 407.2 | 308.2 | 8.6 | 5.3 |
| 2003 | 179.2 | 20.2 | 40.4 | 0.0 | 45.2 | 80.7 | 33.3 | 249.3 | 529.2 | 70.6 | 54.6 | 0.0 |
| 2004 | 0.0 | 38.7 | 221.3 | 167.0 | 19.4 | 23.3 | 28.8 | 253.8 | 415.3 | 70.1 | 10.3 | 6.0 |
| 2005 | 40.6 | 135.1 | 55.0 | 29.6 | 127.7 | 17.8 | 397.4 | 33.3 | 65.7 | 361.2 | 12.5 | 14.3 |
| 2006 | 0.0 | 28.7 | 19.6 | 10.4 | 49.7 | 56.5 | 74.0 | 87.0 | 244.6 | 19.3 | 36.0 | INAP |
| 2007 | 61.8 | 43.5 | 4.5 | 65.0 | 97.0 | 124.0 | 89.3 | 163.0 | 99.9 | 42.0 | 46.0 | 2.5 |
| 2008 | 25.5 | 5.0 | 34.5 | 91.5 | 58.0 | 1.0 | 169.0 | 169.0 | 789.5 | 130.0 | 2.0 | 14.5 |
| 2009 | 32.5 | 16.0 | 22.5 | 9.0 | 62.5 | 58.0 | 17.0 | 31.0 | 106.6 | 65.6 | 19.5 | 44.5 |
| 2010 | S.D. | S.D. | 20.0 | 299.2 | S.D. | 93.5 | 444.8 | 17.5 | 169.5 | 35.3 | 0.0 | S.D. |
| 2011 | 21.7 | 0.0 | 23.0 | 0.0 | 152.2 | 103.1 | 82.8 | 83.8 | 87.5 | 89.0 | 2.5 | 27.7 |

Cuadro 6.1 con asignaciones S/D, s/d, S.D., INAP, comas, espacios etc..

CUADRO 6.2: DATOS CON ASIGNACIÓN NAN

| AÑOS | ENE | FEB | MAR | ABR | MAY | JUN | JUL | AGO | SEP | OCT | NOV | DIC |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| 1987 | 50.0 | 63.0 | 35.5 | 68.0 | 79.5 | 356.1 | 146.1 | 241.5 | 356.0 | 69.0 | 13.0 | 9.0 |
| 1988 | 36.3 | 11.8 | 12.5 | 41.0 | 69.5 | 209.0 | 130.5 | 236.0 | 388.0 | 41.0 | NaN | 9.0 |
| 1989 | 30.0 | 41.5 | 8.5 | 54.0 | 15.5 | 51.5 | NaN | 44.5 | 251.0 | 192.0 | 28.5 | 74.0 |
| 1990 | 17.5 | 21.5 | 47.5 | 122.0 | 141.5 | 16.0 | 101.5 | 58.0 | 316.5 | 80.0 | 19.5 | 17.0 |
| 1991 | 26.5 | 47.5 | 35.5 | 123.5 | 167.5 | 172.5 | 46.0 | 30.5 | 262.0 | 42.0 | 29.5 | 54.0 |
| 1992 | 142.5 | 33.5 | 40.0 | 136.5 | 100.5 | 63.5 | 1.0 | 210.5 | 212.0 | 129.0 | 48.0 | 39.0 |
| 1993 | 33.5 | 46.5 | 51.5 | 23.8 | 151.0 | 298.5 | 28.0 | 58.5 | 367.5 | 50.5 | 31.0 | 19.5 |
| 1994 | 67.5 | 31.0 | 50.5 | 19.5 | 41.5 | 101.0 | 36.5 | 83.5 | 210.3 | 67.2 | 39.0 | 46.0 |
| 1995 | 28.5 | 27.0 | 50.0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 1996 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 30.1 | 179.1 | 7.2 | 10.2 |
| 1997 | 18.7 | 61.8 | 72.2 | 121.2 | 128.8 | 68.1 | 21.4 | 69.3 | 42.8 | 226.4 | 36.7 | 14.8 |
| 1998 | 7.7 | 17.9 | 13.7 | 9.0 | 0.0 | 12.5 | 28.3 | 262.2 | 232.0 | 38.7 | 63.3 | 5.2 |
| 1999 | NaN | 0.0 | 25.8 | 25.7 | 34.1 | 103.9 | 145.3 | 0 | 72.4 | 25.7 | 0.0 | NaN |
| 2000 | 0.0 | 33.9 | 12.6 | 13.7 | 100.1 | 104.7 | 20.5 | 93.3 | 72.1 | 279.1 | 44.5 | 2 |
| 2001 | 47.1 | 5.3 | 68.7 | 78.0 | 15.4 | 99.6 | 25.9 | 25.8 | 564.9 | 42.0 | 131.2 | 24.4 |
| 2002 | 4.3 | NaN | NaN | NaN | NaN | NaN | 118.6 | 93.2 | 407.2 | 308.2 | 8.6 | 5.3 |
| 2003 | 179.2 | 20.2 | 40.4 | 0.0 | 45.2 | 80.7 | 33.3 | 249.3 | 529.2 | 70.6 | 54.6 | 0.0 |
| 2004 | 0.0 | 38.7 | 221.3 | 167.0 | 19.4 | 23.3 | 28.8 | 253.8 | 415.3 | 70.1 | 10.3 | 6.0 |
| 2005 | 40.6 | 135.1 | 55.0 | 29.6 | 127.7 | 17.8 | 397.4 | 33.3 | 65.7 | 361.2 | 12.5 | 14.3 |
| 2006 | 0.0 | 28.7 | 19.6 | 10.4 | 49.7 | 56.5 | 74.0 | 87.0 | 244.6 | 19.3 | 36.0 | NaN |
| 2007 | 61.8 | 43.5 | 4.5 | 65.0 | 97.0 | 124.0 | 89.3 | 163.0 | 99.9 | 42.0 | 46.0 | 2.5 |
| 2008 | 25.5 | 5.0 | 34.5 | 91.5 | 58.0 | 1.0 | 169.0 | 169.0 | 789.5 | 130.0 | 2.0 | 14.5 |
| 2009 | 32.5 | 16.0 | 22.5 | 9.0 | 62.5 | 58.0 | 17.0 | 31.0 | 106.6 | 65.6 | 19.5 | 44.5 |
| 2010 | NaN | NaN | 20.0 | 299.2 | NaN | 93.5 | 444.8 | 17.5 | 169.5 | 35.3 | 0.0 | NaN |
| 2011 | 21.7 | 0.0 | 23.0 | 0.0 | 152.2 | 103.1 | 82.8 | 83.8 | 87.5 | 89.0 | 2.5 | 27.7 |

Cuadro 6.2 con asignaciones NaN utilizando matlab.

CUADRO 6.3: DATOS SIN ASIGNACIÓN NAN

| AÑOS | ENE | FEB | MAR | ABR | MAY | JUN | JUL | AGO | SEP | OCT | NOV | DIC |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| 1987 | 50.0 | 63.0 | 35.5 | 68.0 | 79.5 | 356.1 | 146.1 | 241.5 | 356.0 | 69.0 | 13.0 | 9.0 |
| 1988 | 36.3 | 11.8 | 12.5 | 41.0 | 69.5 | 209.0 | 130.5 | 236.0 | 388.0 | 41.0 | 20.8 | 9.0 |
| 1989 | 30.0 | 41.5 | 8.5 | 54.0 | 15.5 | 51.5 | 116.0 | 44.5 | 251.0 | 192.0 | 28.5 | 74.0 |
| 1990 | 17.5 | 21.5 | 47.5 | 122.0 | 141.5 | 16.0 | 101.5 | 58.0 | 316.5 | 80.0 | 19.5 | 17.0 |
| 1991 | 26.5 | 47.5 | 35.5 | 123.5 | 167.5 | 172.5 | 46.0 | 30.5 | 262.0 | 42.0 | 29.5 | 54.0 |
| 1992 | 142.5 | 33.5 | 40.0 | 136.5 | 100.5 | 63.5 | 1.0 | 210.5 | 212.0 | 129.0 | 48.0 | 39.0 |
| 1993 | 33.5 | 46.5 | 51.5 | 23.8 | 151.0 | 298.5 | 28.0 | 58.5 | 367.5 | 50.5 | 31.0 | 19.5 |
| 1994 | 67.5 | 31.0 | 50.5 | 19.5 | 41.5 | 101.0 | 36.5 | 83.5 | 210.3 | 67.2 | 39.0 | 46.0 |
| 1995 | 28.5 | 27.0 | 50.0 | 53.4 | 70.6 | 90.03 | 31.5 | 78.8 | 120.2 | 123.2 | 23.1 | 28.1 |
| 1996 | 23.6 | 44.4 | 61.1 | 87.3 | 99.7 | 79.1 | 26.4 | 74.1 | 30.1 | 179.1 | 7.2 | 10.2 |
| 1997 | 18.7 | 61.8 | 72.2 | 121.2 | 128.8 | 68.1 | 21.4 | 69.3 | 42.8 | 226.4 | 36.7 | 14.8 |
| 1998 | 7.7 | 17.9 | 13.7 | 9.0 | 0.0 | 12.5 | 28.3 | 262.2 | 232.0 | 38.7 | 63.3 | 5.2 |
| 1999 | 3.9 | 0.0 | 25.8 | 25.7 | 34.1 | 103.9 | 145.3 | 0 | 72.4 | 25.7 | 0.0 | 14.3 |
| 2000 | 0.0 | 33.9 | 12.6 | 13.7 | 100.1 | 104.7 | 20.5 | 93.3 | 72.1 | 279.1 | 44.5 | 2 |
| 2001 | 47.1 | 5.3 | 68.7 | 78.0 | 15.4 | 99.6 | 25.9 | 25.8 | 564.9 | 42.0 | 131.2 | 24.4 |
| 2002 | 4.3 | 12.8 | 554.5 | 39.0 | 30.3 | 91.2 | 118.6 | 93.2 | 407.2 | 308.2 | 8.6 | 5.3 |
| 2003 | 179.2 | 20.2 | 40.4 | 0.0 | 45.2 | 80.7 | 33.3 | 249.3 | 529.2 | 70.6 | 54.6 | 0.0 |
| 2004 | 0.0 | 38.7 | 221.3 | 167.0 | 19.4 | 23.3 | 28.8 | 253.8 | 415.3 | 70.1 | 10.3 | 6.0 |
| 2005 | 40.6 | 135.1 | 55.0 | 29.6 | 127.7 | 17.8 | 397.4 | 33.3 | 65.7 | 361.2 | 12.5 | 14.3 |
| 2006 | 0.0 | 28.7 | 19.6 | 10.4 | 49.7 | 56.5 | 74.0 | 87.0 | 244.6 | 19.3 | 36.0 | 8.4 |
| 2007 | 61.8 | 43.5 | 4.5 | 65.0 | 97.0 | 124.0 | 89.3 | 163.0 | 99.9 | 42.0 | 46.0 | 2.5 |
| 2008 | 25.5 | 5.0 | 34.5 | 91.5 | 58.0 | 1.0 | 169.0 | 169.0 | 789.5 | 130.0 | 2.0 | 14.5 |
| 2009 | 32.5 | 16.0 | 22.5 | 9.0 | 62.5 | 58.0 | 17.0 | 31.0 | 106.6 | 65.6 | 19.5 | 44.5 |
| 2010 | 27.1 | 8.0 | 20.0 | 299.2 | 107.4 | 93.5 | 444.8 | 17.5 | 169.5 | 35.3 | 0.0 | 36.1 |
| 2011 | 21.7 | 0.0 | 23.0 | 0.0 | 152.2 | 103.1 | 82.8 | 83.8 | 87.5 | 89.0 | 2.5 | 27.7 |

Cuadro 6.3 con asignaciones numéricas mediante interpolación.

El cuadro 6.3 muestra los datos ya efectuada la interpolación.

Dado que los posibles problemas presentados por los datos son muchos y la cantidad de técnicas existentes también es alta, elaborar una guía metodológica completa, la cual contemple los diferentes problemas y que detecte la mayor cantidad posible para corregirlos, es una tarea ardua que vale la pena para garantizar la buena toma de decisiones.

Anexo 4. Series de Tiempo y sus gráficos

El marco del trabajo propuesto ha sido probado sobre datos procedentes del CONAGUA. El objetivo del análisis de una serie temporal es el conocimiento de su patrón de comportamiento, para así prever su evolución futura, suponiendo que las condiciones no variarán.

Dado que no se trata de fenómenos deterministas, sino sujetos a una aleatoriedad, el estudio del comportamiento pasado ayuda a inferir la estructura que permita predecir su comportamiento futuro. La particular forma de la información disponible de una serie cronológica (se dispone de datos en periodos regulares de tiempo) hace que las técnicas habituales de inferencia estadística no sean válidas para estos casos, ya que nos encontramos ante elementos procedentes de poblaciones de características y distribución desconocidas.

Actualmente existen muchas razones por las cuales es importante poder pronosticar utilizando datos de cualquier tipo. Los diferentes métodos para realizar un pronóstico dependen básicamente de la experiencia, la cantidad de información disponible, del nivel de dificultad que presenta la situación y del grado de exactitud o confianza necesaria en el pronóstico.

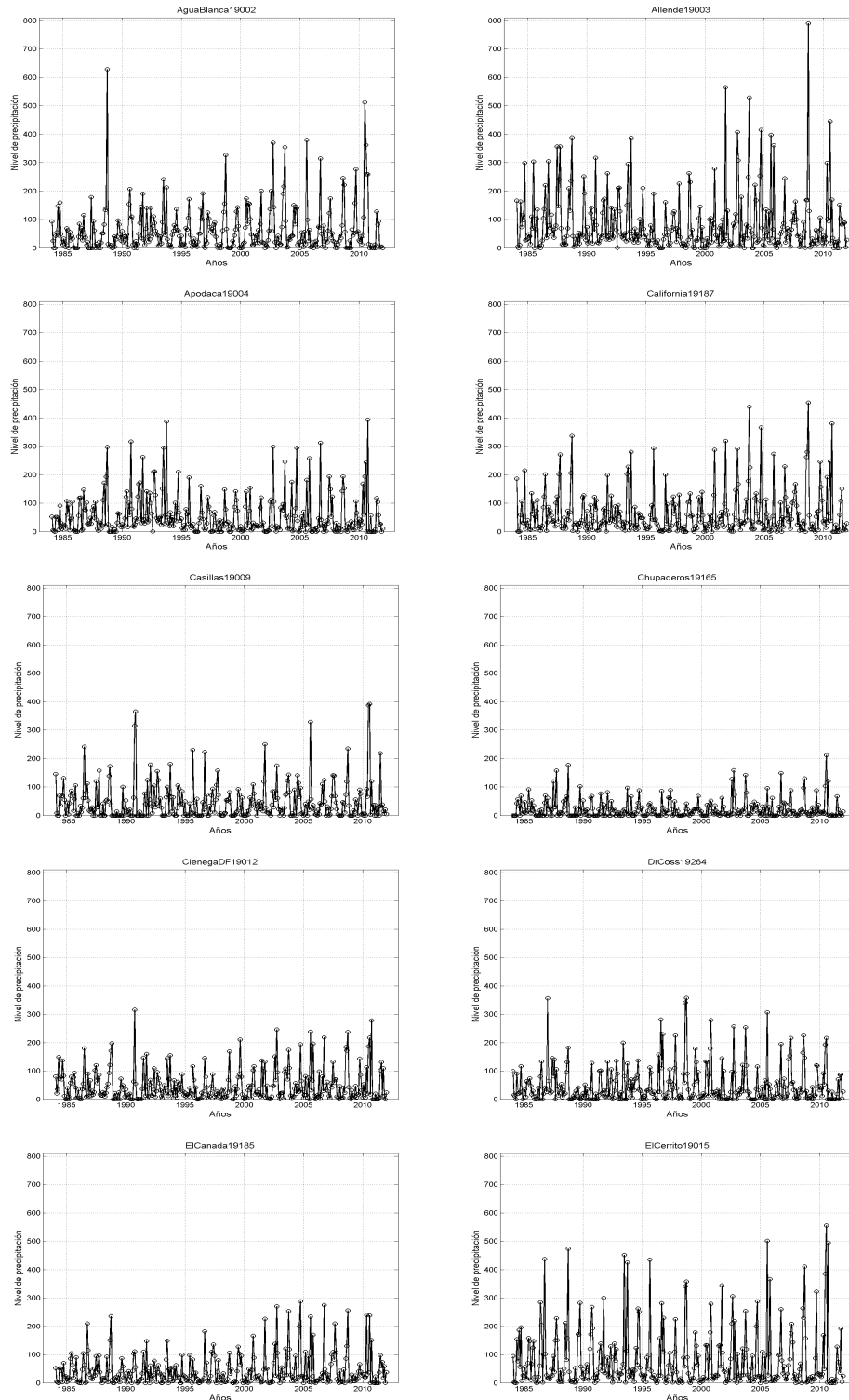
El Método de la persistencia, es la manera más simple de hacer un pronóstico, en él se asume que las condiciones atmosféricas no cambiarán en el tiempo. Si los patrones atmosféricos varían poco o se mueven lentamente en el tiempo, funcionan adecuadamente, pero, si las condiciones varían significativamente, de manera repentina este método falla y no es eficiente para pronosticar.

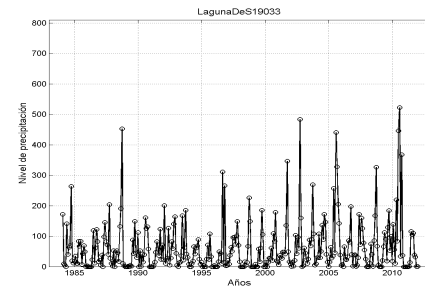
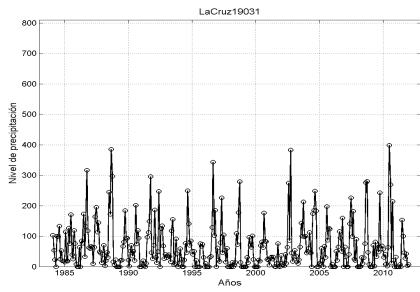
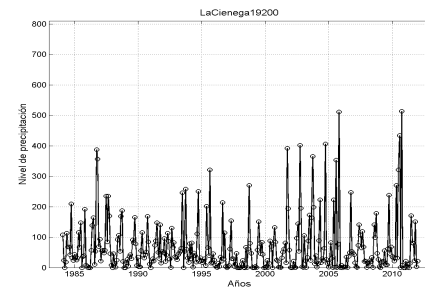
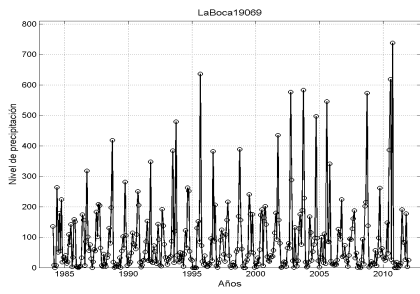
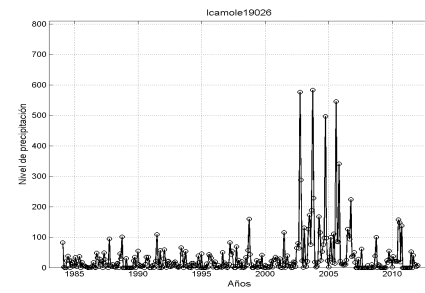
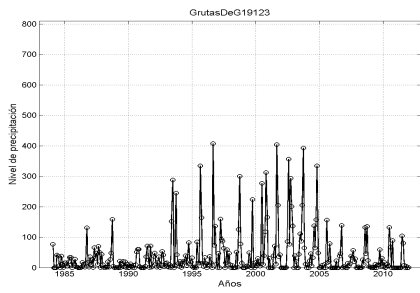
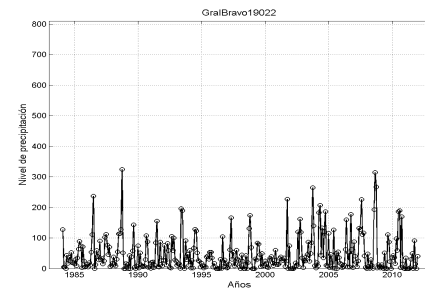
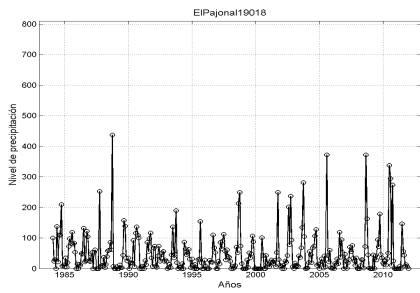
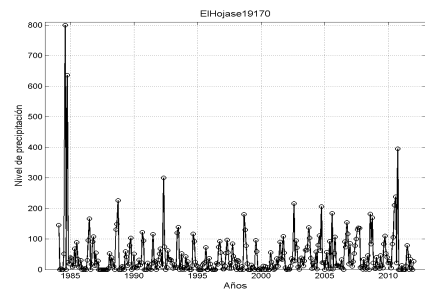
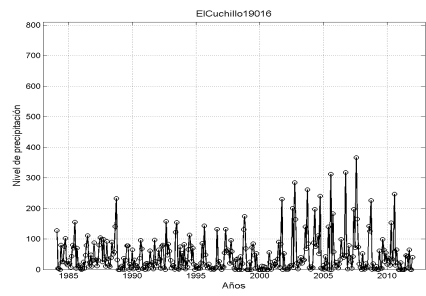
El Método de la tendencia, involucra diferentes cálculos como, medición de velocidad del viento, presiones atmosféricas, frentes, cúmulos de nubes y precipitación. Con esta información se puede predecir donde se esperan características semejantes en un tiempo posterior.

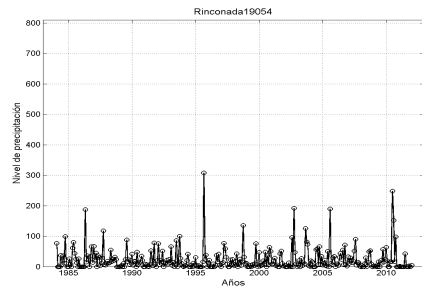
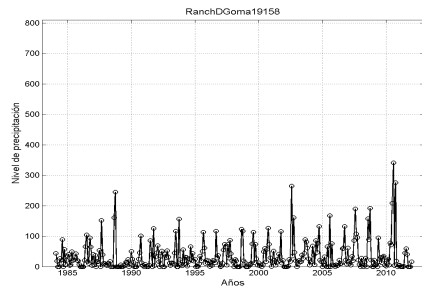
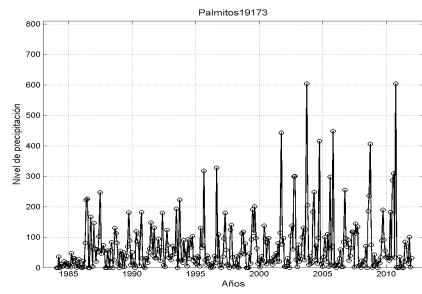
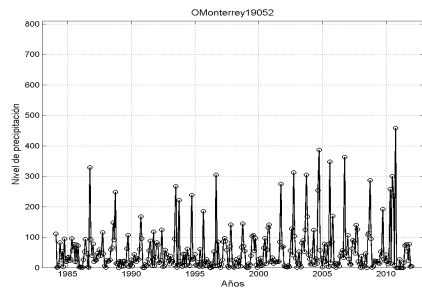
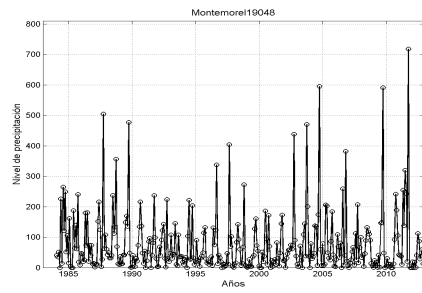
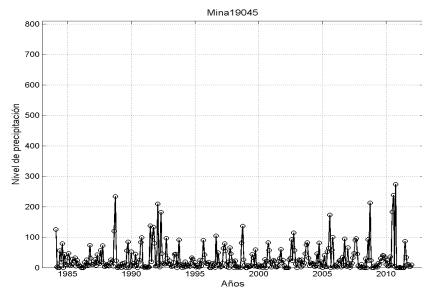
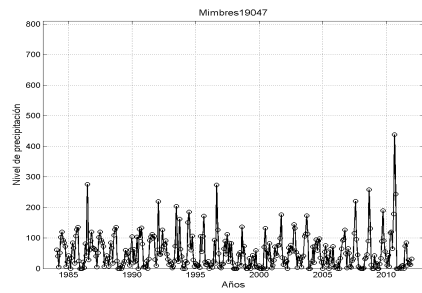
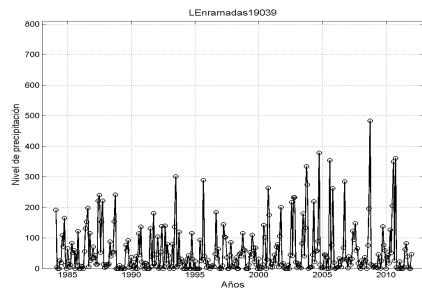
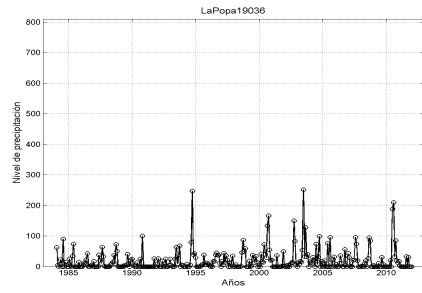
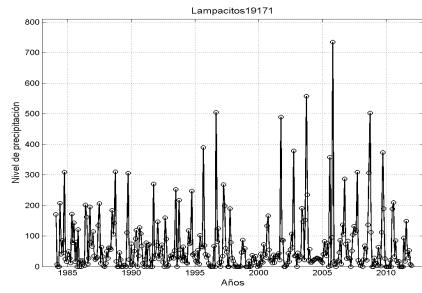
El Método climatológico, involucra el uso de promedios estadísticos de las diferentes variables atmosféricas, con registros históricos de años y algunos lo utilizan ya que parece simple realizar un pronóstico con él.

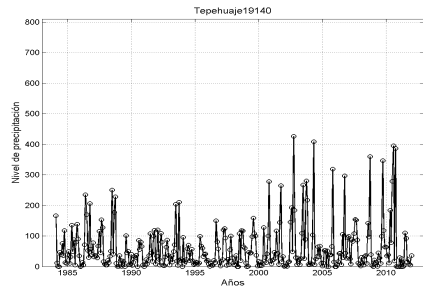
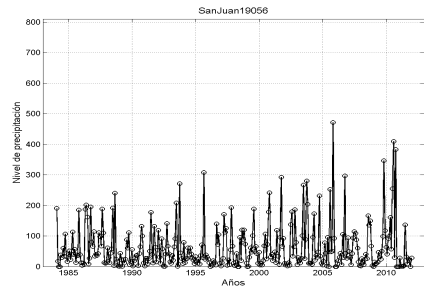
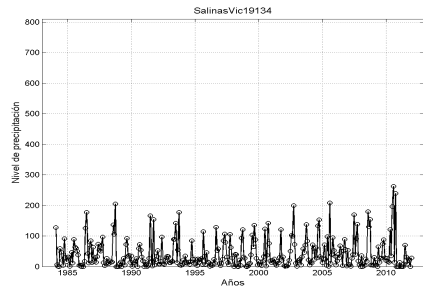
El Método análogo, es un poco más complicado, ya que se tienen que examinar condiciones actuales y recordar escenarios semejantes del pasado y su pronóstico futuro es equivalente al del pasado.

En este anexo se muestran las 33 series de tiempo de las estaciones pluviométricas analizadas.









Anexo 5. Noción de Fractales

A temprana edad estudiamos medidas de longitudes, Dimensión 1, para después adentrarnos en formas geométricas elementales, figuras que pueden ser dibujadas o representadas en un plano, triángulos, cuadrados, rectángulos, polígonos, círculos y su contorno: la circunferencia, cosas bidimensionales, Dimensión 2.

Posterior a esto, nos adentramos a la geometría de los cuerpos que ocupan un volumen: prismas, cilindros, conos, esferas, pirámides y sus diferentes combinaciones figuras en Dimensión 3.

Observamos que estos representan una geometría muy regular, pero del cálculo aparente a la realidad esto no es del todo correcto.

Para representar la naturaleza es necesario utilizar un tipo de geometría que se aproxime más a esas observaciones que ejecutamos y que las incorpore lo más cerca de su propia naturaleza.

De esta manera es, como se presenta una geometría, que quizá no sea la última, para analizar este tipo de situaciones: la geometría fractal.

Esta palabra puede adjudicarse a Benoit Mandelbrot (1924-2010) quien la acuñó por primera vez en su libro "La geometría Fractal de la naturaleza", recopilando un buen número de figuras geométricas y procesos naturales bajo un marco global que denominó fractales. Del adjetivo en latín fractus, derivada del verbo frangere: romper, quebrar en pedazos, desacoplar.

Este término no pudo ser más apropiado ya que uno de las primeras figuras fue la curva creada en 1904 por el matemático sueco N. F. H. Von Koch, la cual exhibe claramente un concepto clave en los fractales: la autosimilitud. La información geométrica que contiene la curva está contenida en cualquier trozo por minúsculo que sea, su longitud diverge. Su distancia no puede ser calculada.

La curva es continua y no derivable en numerosos puntos. No podemos determinar la tangente en incontables de sus puntos.

Lo que Benoit interpreta con esto es que todos los objetos cumplen con una serie de propiedades, las cuales pueden ser descritas con una nueva geometría, la de las formas irregulares, debido a que todo tiene demasiada irregularidad como para ser descrito en

términos geométricos tradicionales.

Un fractal, es por definición un conjunto cuya dimensión de Hausdorff-Besicovitch excede la dimensión topológica.

Desde la perspectiva de la física teórica y aplicada, para discutir una variedad de series espacio dimensional estamos sujetos a una comprensión del espacio multidimensional y que sólo puede quedarse en conceptos matemáticos.

Los fractales son un pensamiento pionero desde donde pueden ser abordados una gran diversidad de problemas, donde podemos incluir los de la naturaleza, y desde un punto crítico, analizar espacios no solamente unidimensionales, bidimensionales, tridimensional o espacio de 4 dimensiones. La primera, segunda o tercera dimensión o el mundo 1D, 2D o 3D, el de las mediciones, el de objetos planos, el de objetos sólidos y el espacio infinito, no ocupa pasar a un pensamiento mayor o menor sino solo aceptar lo que en él se presente o pueda ser analizado.

El tiempo, otra dimensión, un poco difícil de manejar, pero cuando lo empezamos a explorar, las cosas pueden ser más específicas, aunque en múltiples ocasiones la manipulación de esta dimensión, realmente es más difícil.

La ampliación de fractal, es generada por computadoras, mediante ecuaciones iteradas, no lineales, en diferentes dimensiones, incluso en el plano complejo. Estos pueden ser utilizados para desarrollar diversos procesos, como los pluricelulares. A. Lindenmayer, desarrolló los L-Sytems, en los que se analizan diferentes comportamientos encontrando relaciones fractales o dimensión fractal. Su potencia reside en la recursividad. En los 60's la física descubrió el caos determinista los cuales, estos, pueden generar dinámicas impredecibles, como son los conocidos atractores de Lorenz, Hénon y la curva logística de Feigenbaum, que no son otra cosa que fractales. El estudio de ellos, ha generado un avance en los sistemas dinámicos.

En los ochenta Barnsley, desarrolló fractales a partir de la aplicación iterada de contracciones afines. Aplicó rotaciones, contracciones, inversiones y traslaciones lo cual le permitió generar una gran cantidad de fractales, como el famoso helecho. Estos han sido utilizados para encriptación de imágenes, generación de estas y como una buena herramienta para ciberartistas.

Muchas de las figuras presentadas por Mandelbrot, ya habían sido propuestas como ejemplos paradójicos en la discusión de conceptos como curva o dimensión. Los atractores, que surgen de dinámicas que modelan sistemas reales, aparecen en el abstracto espacio de fases.

Estos conceptos pueden ser abordados con más detalle, cuando se utiliza el concepto de

fractalidad. Podemos decir que la naturaleza está plagada de estos.

La geometría fractales es una rama de las matemáticas, estudia los objetos que poseen una dimensión D no entera ó dimension fractal y que presentan propiedades de escala muy particulares. D , es un número real que garantiza el concepto de dimensión ordinaria para objetos que no permiten espacio tangente.

La fractalidad ha encontrado aplicación en diversos campos de la ciencia y la tecnología, proporcionando modelos matemáticos alternativos para obtener distribuciones y propiedades de diversos fenómenos, como propiedades de rocas, yacimientos, generando distribuciones con cierta heterogeneidad y cierto orden a la vez.

Esta dimensión, D , con exponente no entero, parece llenar espacios a escalas más y más finas.

Existen diferentes definiciones de esta, como la dimensión de Hausdorff-Besicovitch, la dimensión de Minkowski-Bouligand, la dimensión de empaquetado, la dimensión de homotecia y la dimension de Rényi. Estas no pueden ser tratadas como universales, ya que las diferencias en la estructura interna del fractal, ocasionan discrepancia entre ellas. Aunque para un buen número de fractales clásicos los valores de las diferentes definiciones de dimensión fractal coinciden en dimension, en general no son equivalentes.

6.1. Estimación de la dimensión fractal en la práctica

Si los fractales están definidos formalmente, como los mencionados anteriormente, puede calcularse su dimension fractal. No obstante, algunos fenómenos u objetos de la vida real pueden mostrar propiedades fractales, y es aquí donde puede ser útil obtener la dimensión fractal de un conjunto de datos de una muestra. Este cálculo no se puede obtener de forma exacta sino que debe estimarse.

Esto se usa en una variedad de áreas de investigación tales como la física, análisis de imagen, acústica, ceros de la función zeta de Riemann é incluso procesos electroquímicos.

En la práctica estas dimensiones fractales son muy sensibles al ruido numérico o experimental, y también a las limitaciones en la cantidad de datos. Las tesis basadas en estimaciones de dimensiones fractales deben tomarse con cuidado ya que hay un límite superior inevitable, a menos que se presenten cantidades muy grandes de datos.

Los más sencillos de implementar son, el conteo por cajas (box counting) y la dimensión de correlación (basada en generar un número de puntos aleatorios en un entorno del fractal y medir cuántos de ellos caen sobre el conjunto fractal). Una técnica que se ha hecho popular es utilizar la transformada de Fourier, para mediciones como el espectro de potencia o señales de comunicación.

El concepto de una dimensión fractal se basa en puntos de vista no convencionales de escala y dimensión. Las nociones tradicionales de geometría dan forma a la escala previsible de acuerdo a las ideas intuitivas y familiares sobre el espacio en el que están contenidas, de manera que, por ejemplo, la medición de una línea, usándola como una medida unitaria y luego tomar $\frac{1}{3}$ de su tamaño, dará para la primer medida un total de longitud 3 veces la cantidad de medidas que el particionado. Esto se mantiene en 2 dimensiones, también. Si se mide el área de un cuadrado a continuación, las medidas de uno nuevo, con una caja de longitud de lado $\frac{1}{3}$ del tamaño del original, se encontrará que es 9 veces el número de plazas que con la primera medida. Tales relaciones de escala familiares pueden definirse matemáticamente por la regla de escala general, donde la variable N , es el número de particiones, ϵ es el factor de escala, y D es la dimensión fractal:

$$N \propto \epsilon^{-D} \quad (6.1)$$

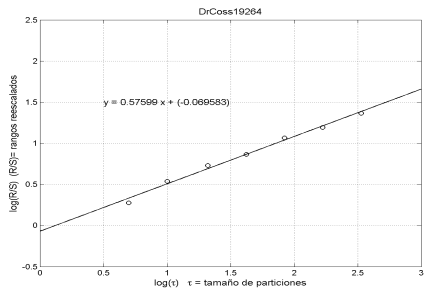
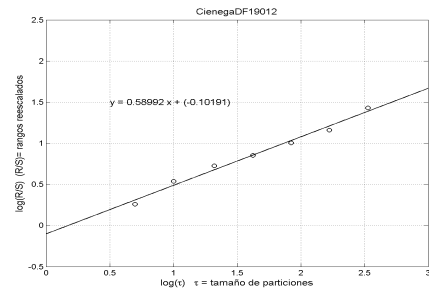
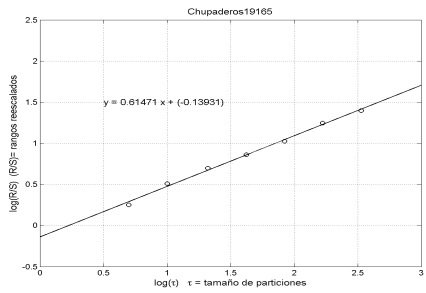
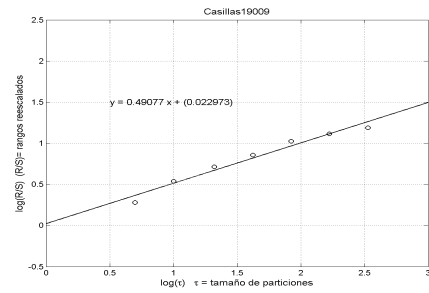
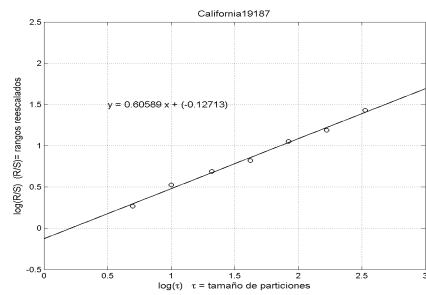
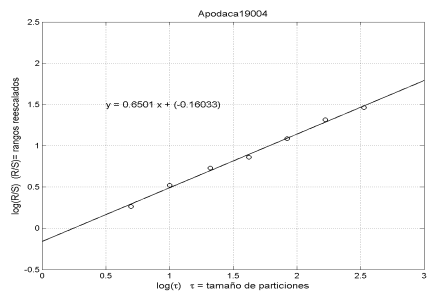
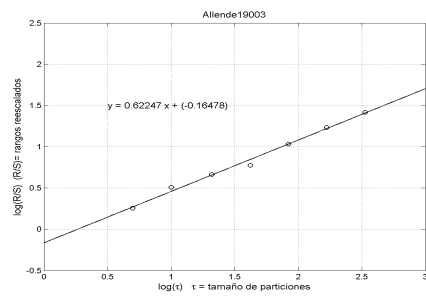
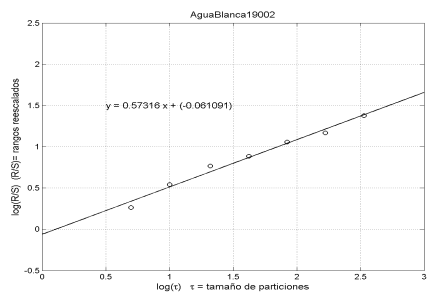
Esta regla de escala tipifica normas convencionales sobre la geometría y la forma para encontrar la dimensión de un fractal.

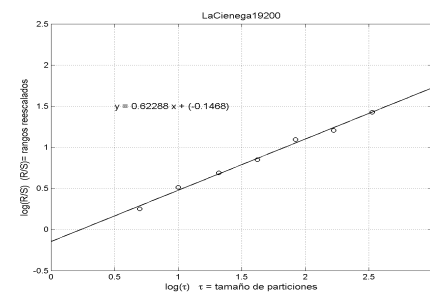
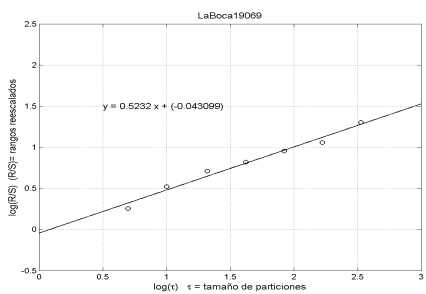
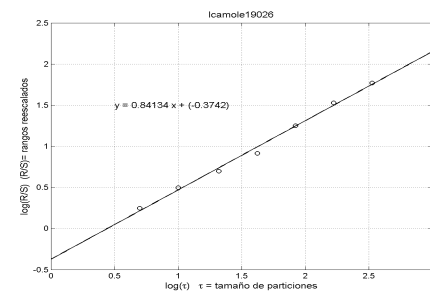
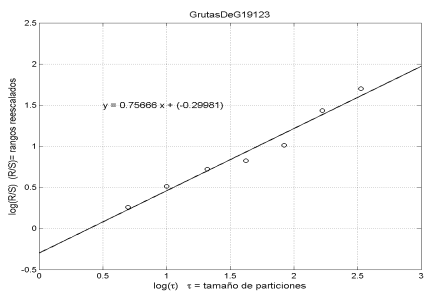
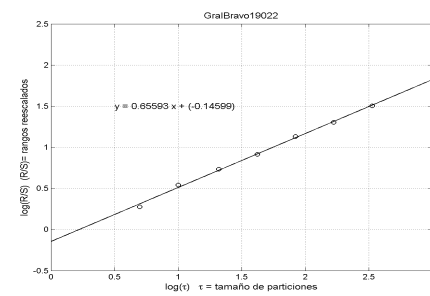
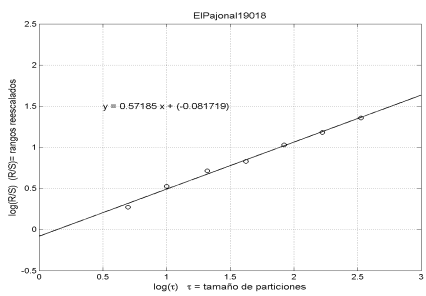
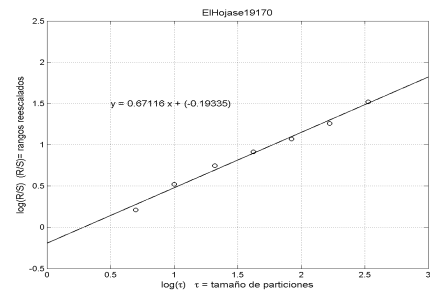
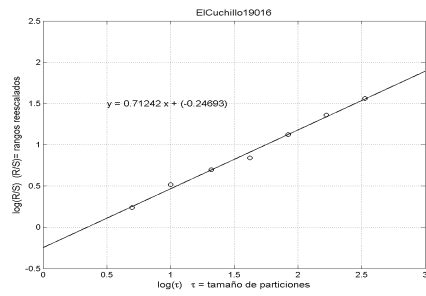
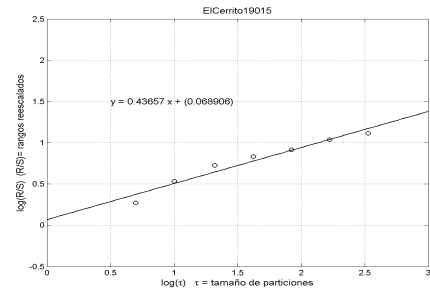
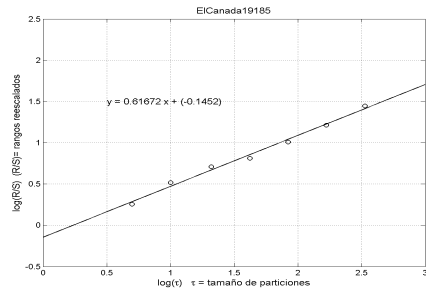
El concepto de dimensión fractal descrito es una vista básica de una construcción complicada.

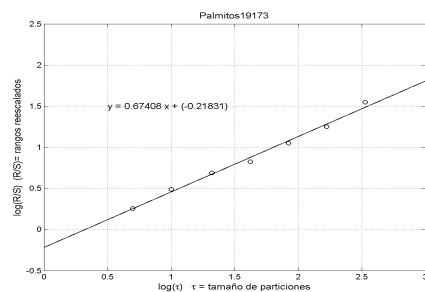
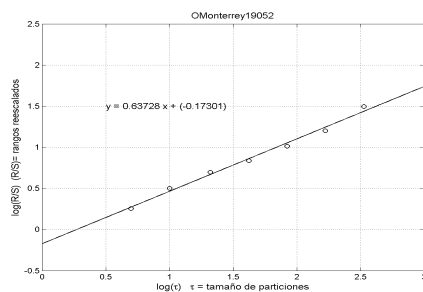
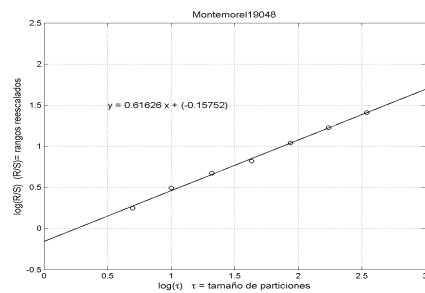
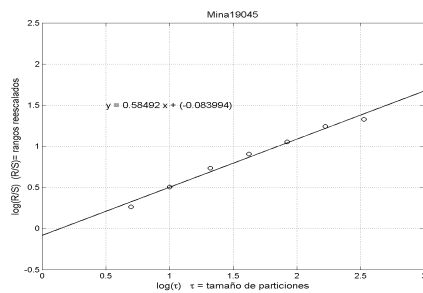
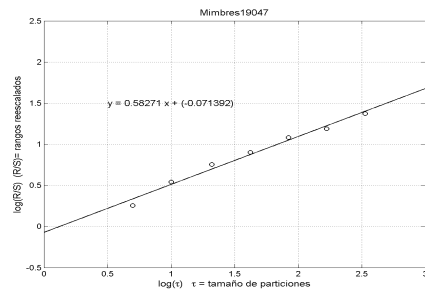
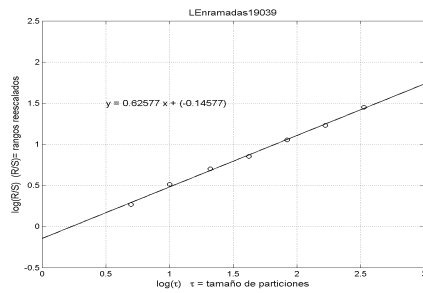
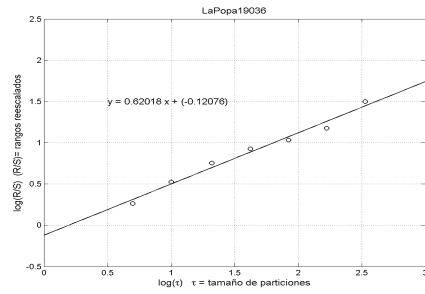
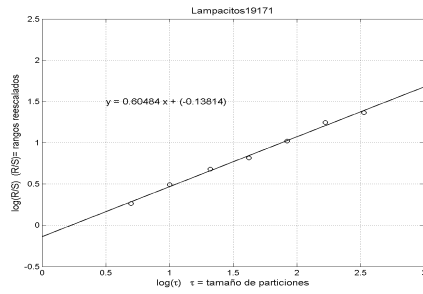
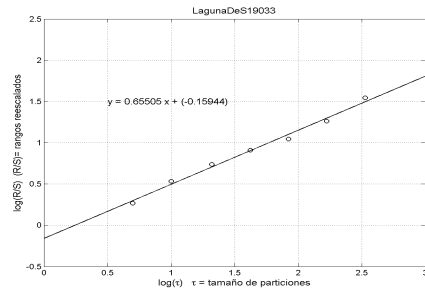
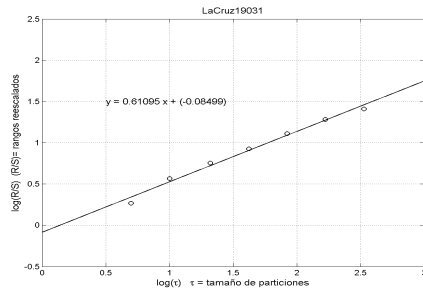
Una validación también se puede hacer mediante la comparación de distintas propiedades fractales, la cuales pueden estar implícitas en el modelo que utilicemos, y también acorde a los datos medidos. Incluso el comportamiento de nuestros datos, quizá este siendo impulsado por una compleja interacción entre la agregación y coalescencia donde dos o mas elementos pueden acompañarse uniendo sus dimensionalidades en un solo conjunto.

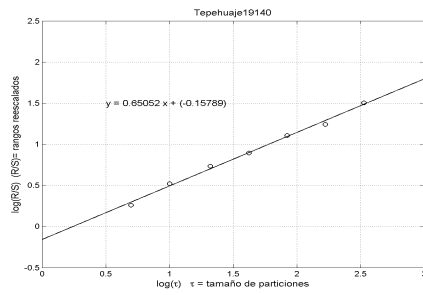
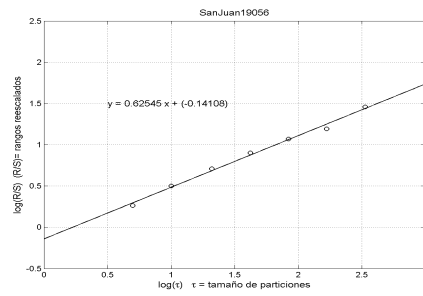
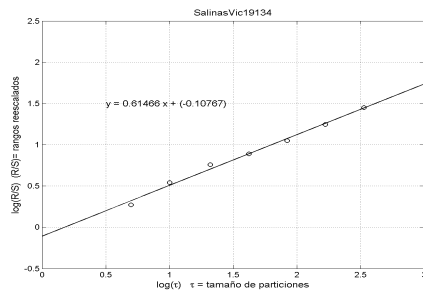
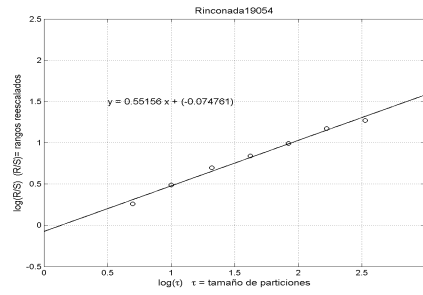
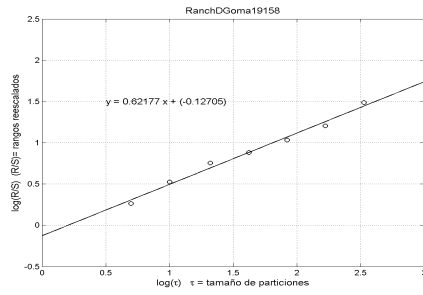
Anexo 6. Exponente fractal de Hurst

En este anexo se muestran las gráficas de las 33 rectas de mínimos cuadrados de las estaciones pluviométricas analizadas, la pendiente en cada ecuación es el denominador exponente fractal de Hurst.









Anexo 7. Variogramas y sus Gráficos

En este anexo se detalla la utilización de los variogramas y se muestran los gráficos de los 33 variogramas de las estaciones pluviométricas analizadas.

El semivariograma o variograma, $\gamma(h)$, es un ajuste o modelado espacial considerado como un estimador de la varianza poblacional y de análisis estructural, donde la población debe tener una tendencia de estacionalidad, se utiliza para describir la relación de observaciones pareadas separadas por una distancia h y en otros casos con una dirección.

Matemáticamente, el variograma estima la diferencia cuadrada prevista entre variables aleatorias vecinas, dando un soporte fundamental y permitiendo representar cuantitativamente esta relación. Este proceso continúa para cada punto de medición.

Teniendo en cuenta una serie de tiempo o procesos estocásticos $\{X_t, t \geq 0\}$, la función de autocovarianza en el punto $(t, t+h)$ se define como $C_X(t, t+h) = E[X_t X_{t+h}] - E[X_t]E[X_{t+h}]$ con $E[X_t]$ la media del proceso en tiempo t .

El semivariograma $\gamma(h)$ está dada por la mitad de la varianza de la diferencia entre pares de observaciones en diferentes “puntos” en el tiempo.

Si $X(t)$ y $X(t+h)$ son variables aleatorias independientes $\forall t$, de nuevo, aunque por una razón diferente, el semivariograma se reduce.

Por lo tanto, el valor medio de las diferencias $X_{t+h} - X_t$ se estima por,

$$E[X_{t+h} - X_t] = \frac{1}{n(h)} \sum_{i=1}^{n(h)} (x_{i+h} - x_i) \quad (6.2)$$

donde $n(h)$ es el número de diferencias con un lag h . Para $h = 1, 2, 3, \dots$ los promedios (6.2) son, respectivamente,

$$\begin{aligned}
 E[X_{t+1} - X_t] &= \frac{x_n - x_1}{n - 1} \\
 E[X_{t+2} - X_t] &= \frac{x_{n-1} + x_n - (x_1 + x_2)}{n - 2} \\
 E[X_{t+3} - X_t] &= \frac{x_{n-2} + x_{n-1} + x_n - (x_1 + x_2 + x_3)}{n - 3} \\
 &\vdots \\
 E[X_{t+h} - X_t] &= \frac{1}{n(1 - \frac{h}{n})} \left(\sum_{j=0}^{k-1} x_{n-j} - \sum_{l=1}^k x_l \right)
 \end{aligned} \tag{6.3}$$

De acuerdo a (6.3) para un valor máximo de h “relativamente moderado” ó $h/n < 1$, excepto en la presencia de valores extremos aislados, las dos sumas en (6.3) son aproximadamente del mismo orden, de manera que el valor medio empírico (6.2) puede ser aproximado por, $E[X_{t+h}] \approx E[X_t] = m = \text{constante}$. Esta es una característica observada en las series de tiempo de las estaciones pluviométricas.

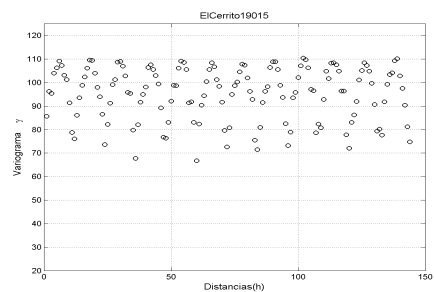
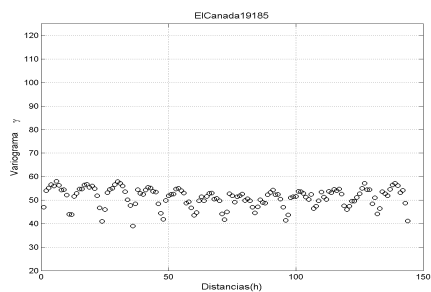
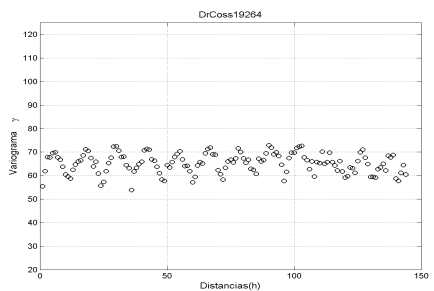
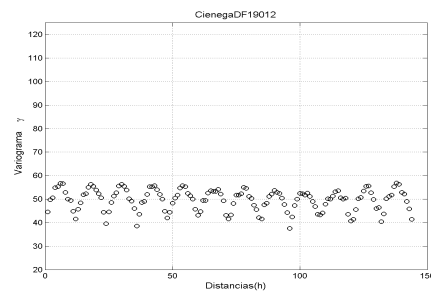
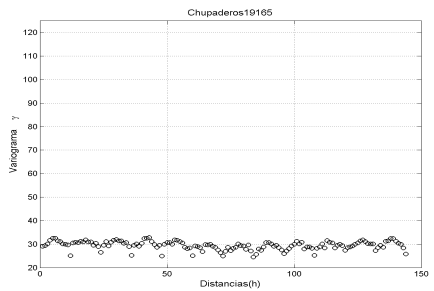
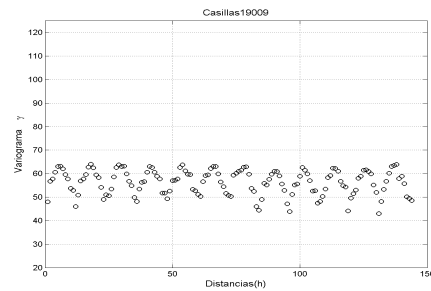
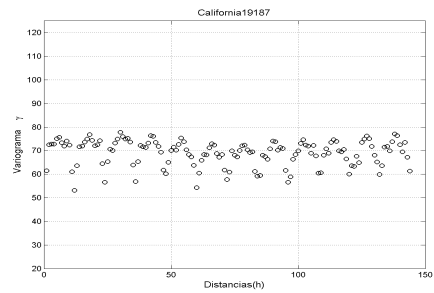
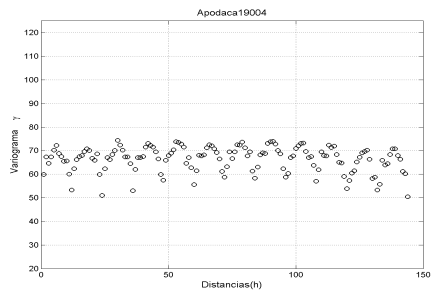
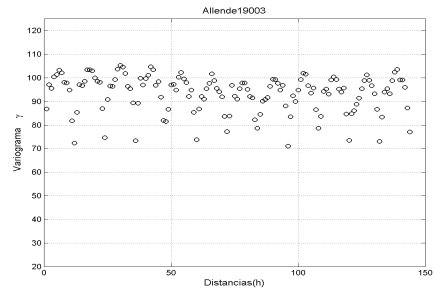
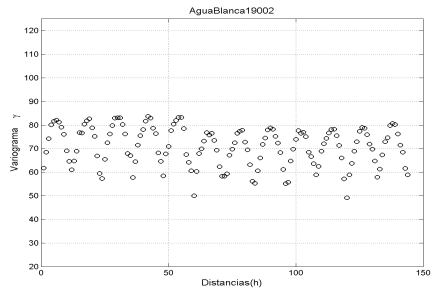
Por lo tanto, el correspondiente estimador de (??) es simplemente:

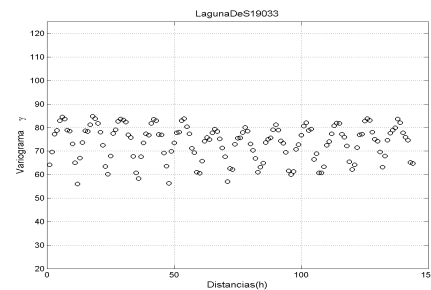
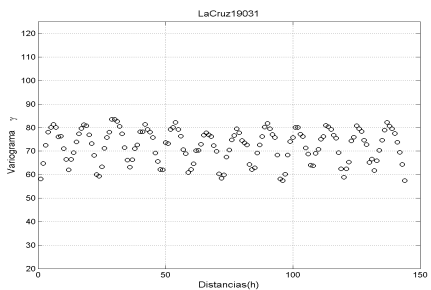
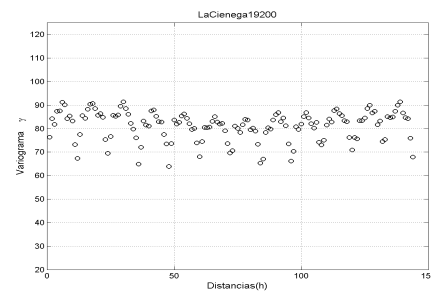
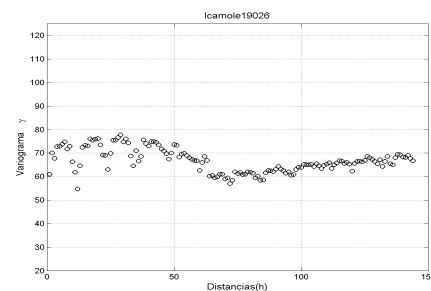
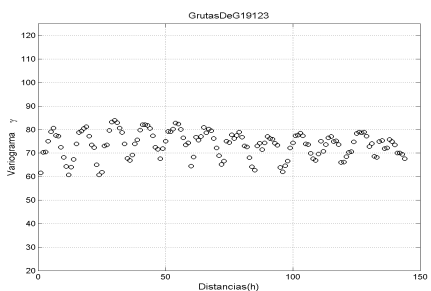
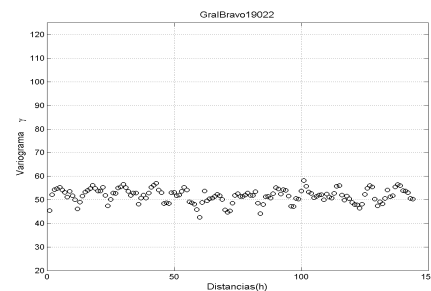
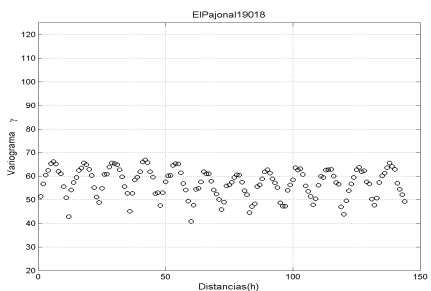
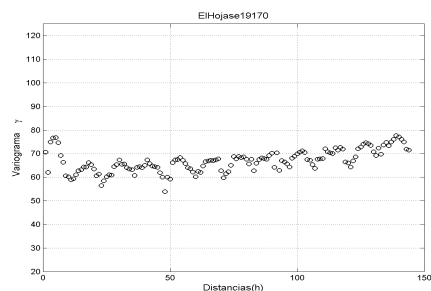
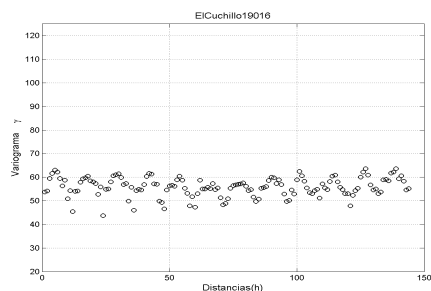
$$\gamma(h) = \frac{1}{2n(h)} \sum_{t=1}^{n(h)} (x_{t+h} - x_t)^2 \tag{6.4}$$

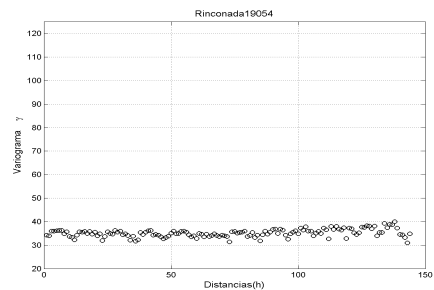
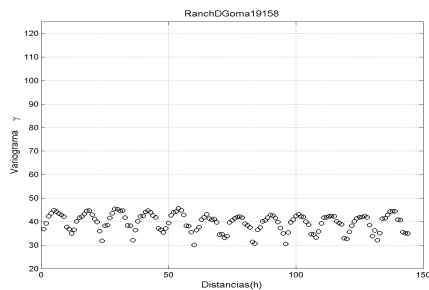
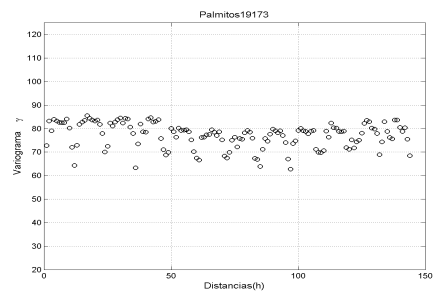
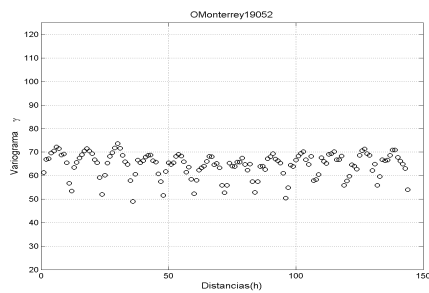
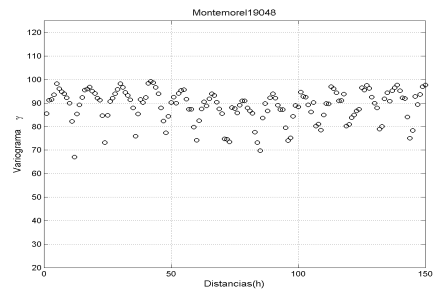
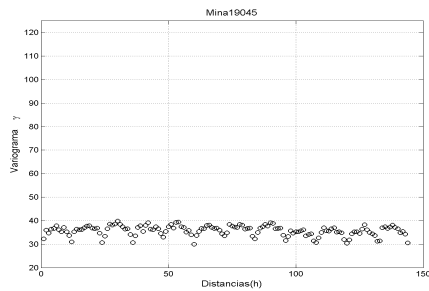
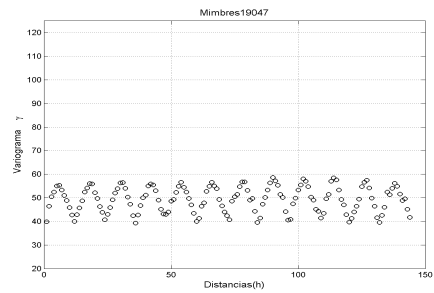
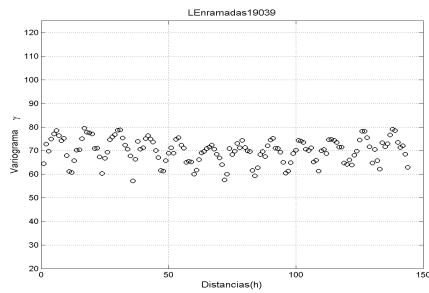
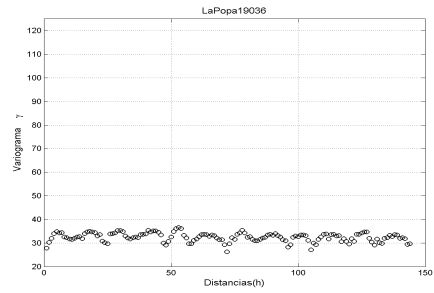
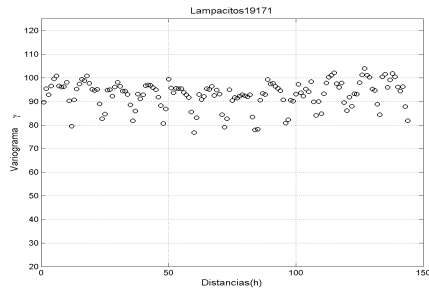
Este estimador de momentos, es un promedio de diferencias al cuadrado, que puede ser influenciado por un número pequeño de valores que ocasionan discrepancias al final de los cálculos debido a las particiones realizadas. Pero, se considera un estimador robusto, ya que disminuye la importancia de las diferencias grandes y al cuadrado.

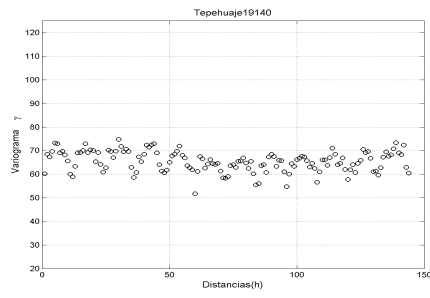
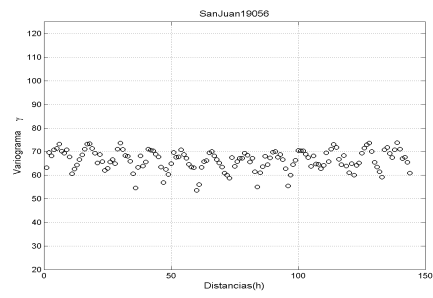
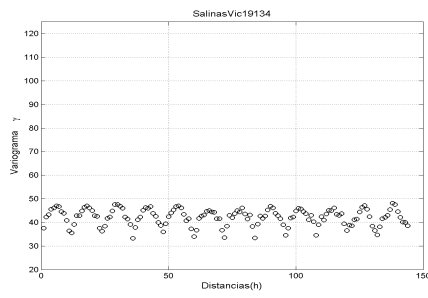
También se considera robusto en el sentido de que es resistente a distribuciones normales contaminadas y afloramientos posiblemente generados por distribuciones de colas pesadas. Esto puede verse por el uso de la raíz cuadrada de las diferencias, en lugar de diferencias de cuadrados, en el estimador insesgado.

La aplicación de variogramas a nuestros datos mostró periodos largos de comportamiento similar, con diferente duración, un comportamiento o patrón cíclico, pero no periódico.



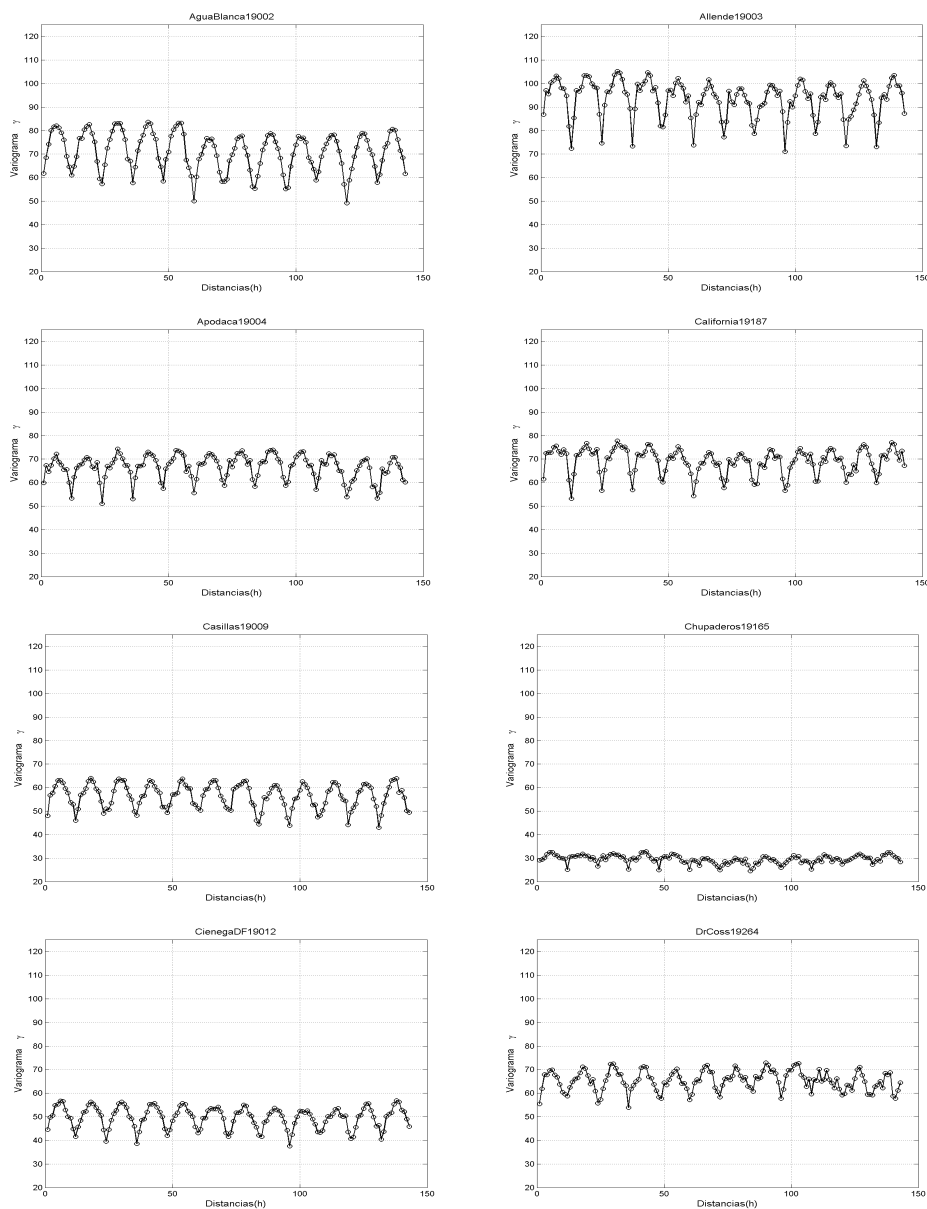


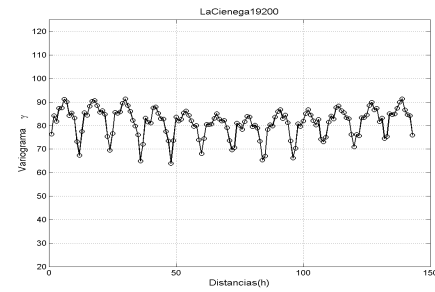
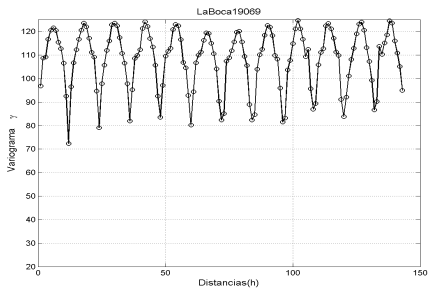
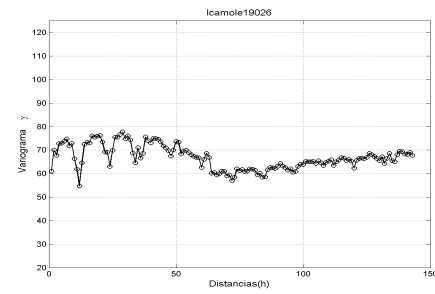
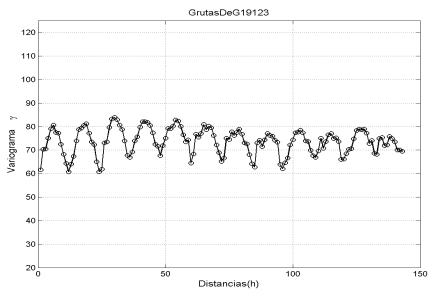
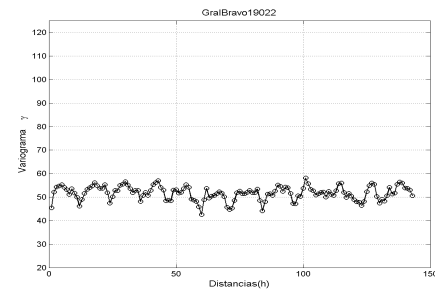
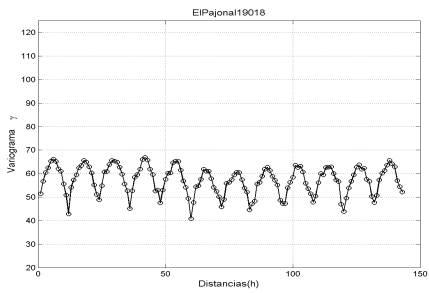
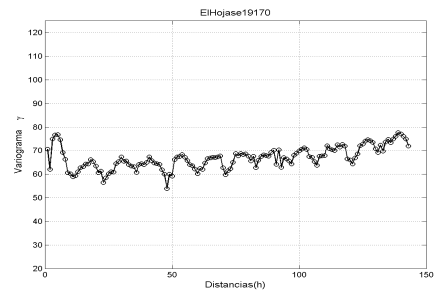
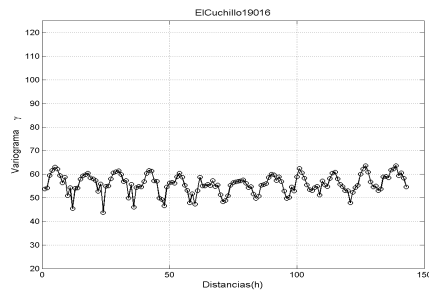
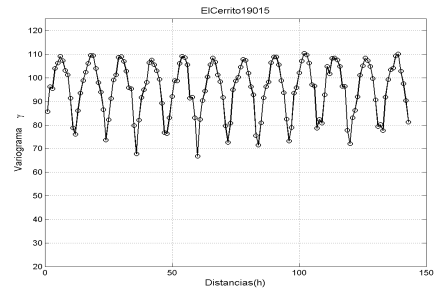
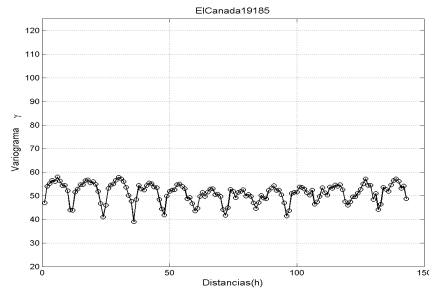


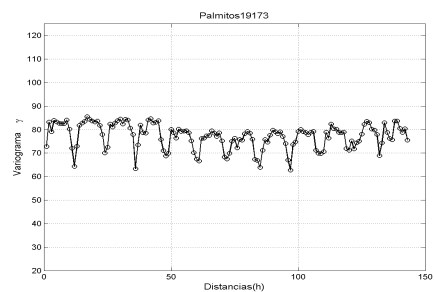
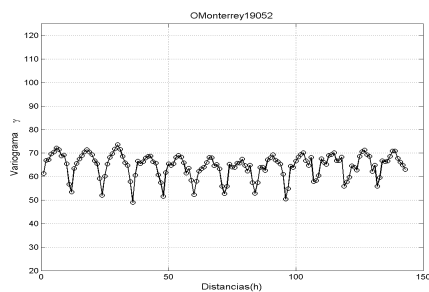
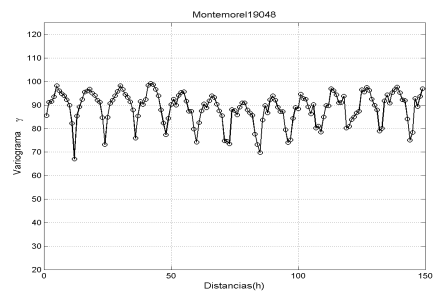
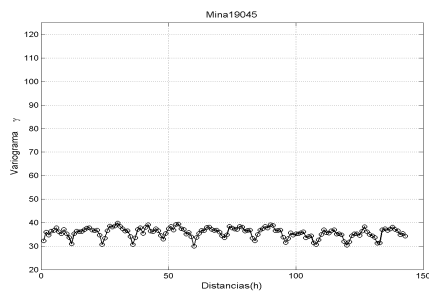
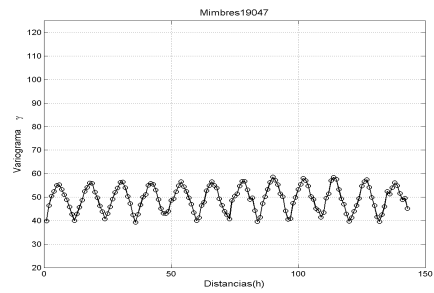
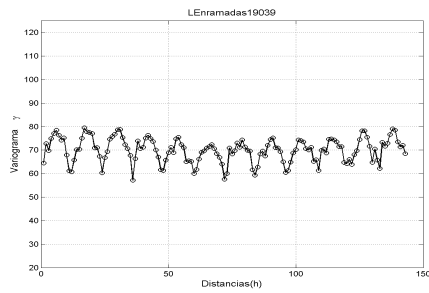
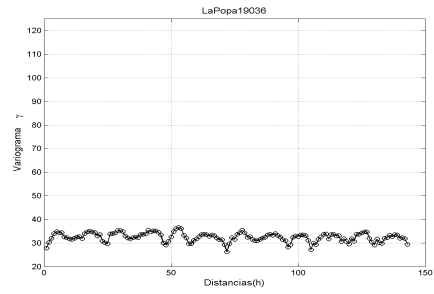
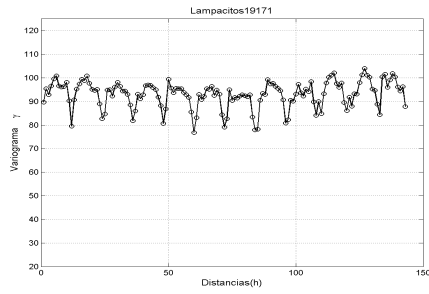
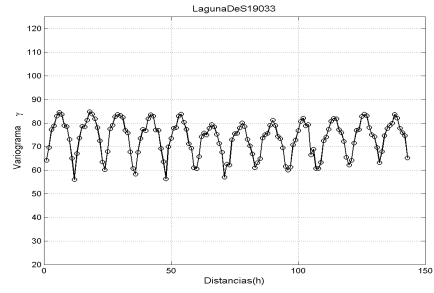
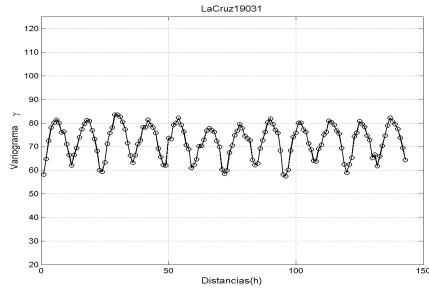


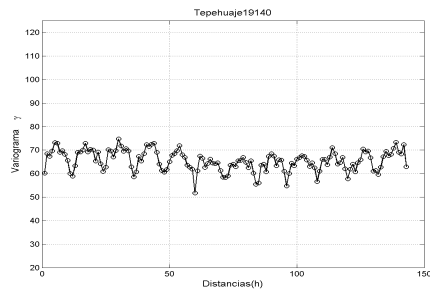
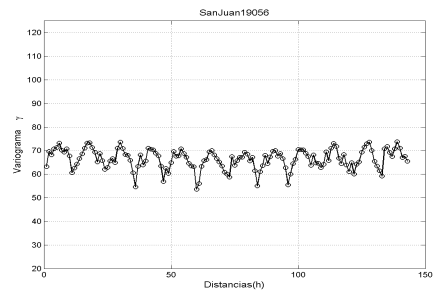
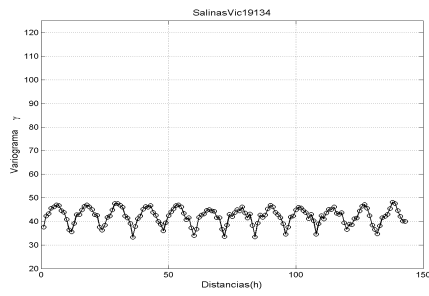
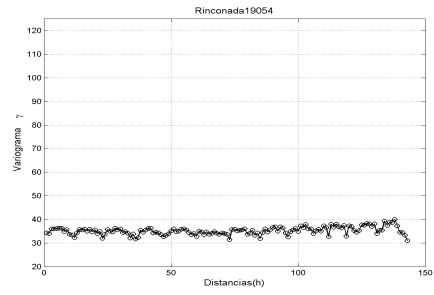
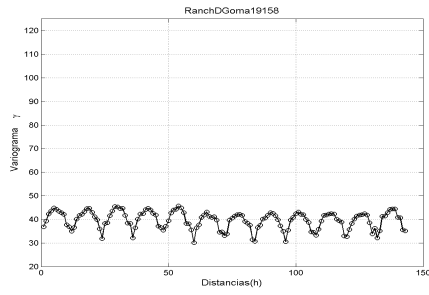
Anexo 8. Gráficos de Variogramas enlazados

En este anexo se muestran los gráficos de los 33 variogramas del anexo 4, enlazando cada uno de los valores calculados en los mismos de cada estación pluviométrica analizadas.









Anexo 9. Histogramas y Distribuciones de probabilidad

En este anexo se muestran los gráficos de histogramas y ajuste de distribución de las 33 estaciones pluviométricas analizadas y de la cual se calcularon los parámetros de ubicación, dispersión, asimetría y curtosis, mediante L-momentos, para ajustar la curva que se muestra en una de ellas.

