



UNIVERSIDAD DEL VALLE

MODELACIÓN DE LA CANTIDAD DE PRECIPITACIÓN TOTAL MENSUAL HACIENDO USO DE DISTRIBUCIONES CON ASIMETRÍA POSITIVA

Autores:

Angélica M. RESTREPO YUSTRE

William A. SANCLEMENTE S.

Directores:

Msc. Gabriel CONDE A.

Alvaro FLOREZ

9 de septiembre de 2014

Agradecimientos

Al finalizar un trabajo complicado y lleno de dificultades como este, es inevitable concentrar la mayor parte del merito en nosotros mismos. Sin embargo, llevando a cabo un análisis objetivo notamos inmediatamente que la magnitud de ese merito hubiese sido imposible sin la participación de personas que han facilitado las cosas para que este trabajo llegue a feliz término.

Gracias a nuestros compañeros y amigos por la compañía y apoyo brindado durante todos estos años. A todos nuestros maestros, gracias por su tiempo, así como por la sabiduría que nos transmitieron en el desarrollo de nuestra formación profesional, en especial a los profesores Alvaro J. Florez y Gabriel Conde, por haber guiado el desarrollo de este trabajo y llegar a la culminación del mismo.

El agradecimiento más profundo y sentido va para nuestras familias. Sin su apoyo, colaboración e inspiración habría sido imposible desarrollar esta dura tarea. Además por enseñarnos que la perseverancia y el esfuerzo son el camino para lograr objetivos. Por ellos y para ellos.

Y por supuesto a Dios, por la paciencia y sabiduría.

Gracias a todos.

William y Angélica

Abstract

This research has how aim the modelling of monthly rainfall amount for the automated meteorological station Melendez subscribed the metrological network CENICANA organization, which is located in the city of Santiago de Cali. Daily records involved in the analysis correspond to an observation period of 20 years, ranging from 1 September 1993 to 31 August 2013. For the development of this study positive skewed distributions such as Gamma, Log-normal and Weibull distributions widely used in the literature to describe the behavior of wet and dry seasons are used. This pattern is observed in a department like Valle del Cauca has two dry quarters (December-February and June-Aug) and two rainy (Mar-May and Sep-Nov).It also proposes the use of the distribution Leak (Loi des Fuites) as a method to describe the behavior of records considered dry months. The proposed distributions will be adjusted each month of the year, and by goodness of fit criteria relevant distribution is assigned to describe the behavior of the records in each of the 12 months. Using the Fourier Series is intended to describe the seasonal variation of the monthly mean and standard deviation, estimated models assigned to each of the months. Finally the validation of the assumptions of the Fourier series is made.

Keywords:*skewed distributions, Fourier Series, Kolmogorov-Smirnov, Criteria BIC, Q-Q plots, maximum likelihood estimation*

Resumen

Este trabajo tiene como objetivo la modelación de la cantidad de precipitación mensual para la estación Meteorológica Automatizada de Meléndez suscripta a la red meteorológica de la organización CENICAÑA, y la cual se encuentra ubicada en el municipio de Santiago de Cali. Los registros diarios involucrados en el análisis corresponden a un periodo de observación de 20 años, que van desde el 01 de septiembre 1993 al 31 de agosto de 2013. Para el desarrollo de este estudio se emplean distribuciones asimétricas positivas como la Gamma, Log-normal y Weibull, distribuciones ampliamente utilizadas en la literatura para describir el comportamiento de temporadas húmedas como secas. Este patrón es observado en un departamento como el Valle del Cauca que cuenta con dos trimestres secos (Dic-Feb y Jun-Ago) y dos lluviosos (Mar-May y Sep-Nov). Se propone además el uso de la distribución de Fugas (Loi des Fuites) como método para describir el comportamiento de los registros en meses considerados como secos. Las distribuciones propuestas serán ajustadas a cada uno los meses de año, y mediante criterios de bondad de ajuste se asignará la distribución pertinente para describir el comportamiento de los registros en cada uno de los 12 meses del año. Haciendo uso de las Series de Fourier se pretende describir la variación estacional de la media y desviación estándar mensual, estimadas en los modelos asignados a cada uno de los meses. Finalmente se efectúa la validación de los supuestos de la Serie de Fourier.

Palabras Claves: *Distribuciones Asimétricas, Series de Fourier, Prueba de Kolmogorov-Smirnov, Criterio BIC, Q-Q plots, estimación de máxima verosimilitud*

Tabla de Contenido

1. Introducción	1
1.1. Planteamiento del Problema	2
1.2. Objetivos	3
1.3. Justificación	4
1.4. Antecedentes	5
2. Marco Teórico	11
2.1. Marco Teórico Conceptual	11
2.1.1. El clima	11
2.1.2. Ciclo hidrológico	11
2.1.3. Definición de lluvia	12
2.1.4. Medición de la lluvia	12
2.1.5. Zona de Convergencia Intertropical (ZCIT)	12
2.1.6. Fenómeno del Niño y de la Niña	12
2.1.7. Régimen de la precipitación en Colombia	13
2.1.8. Régimen de la precipitación en el Valle del Cauca	13
2.2. Marco Teórico Estadístico	14
2.2.1. Distribución de probabilidad	14
2.2.2. Método de Máxima Verosimilitud (MLE)	16

2.2.3. Distribución Gamma	17
2.2.4. Distribución “loi des fuites”	19
2.2.5. Distribución Lognormal	21
2.2.6. Distribución Weibull	23
2.2.7. Pruebas de Bondad de Ajuste	24
2.2.8. Series de Fourier	27
2.2.9. Versión rankeada del test de razón de Von-Neumann	31
3. Metodología	33
3.1. Análisis exploratorio de datos	33
3.2. Ajuste de las distribuciones propuestas	34
3.3. Pruebas de bondad de ajuste	34
3.4. Ajuste de las series de Fourier	35
4. Resultados	41
4.1. Análisis Exploratorio de datos	42
4.2. Ajuste de las distribuciones propuestas	44
4.3. Pruebas de bondad de ajuste	45
4.4. Ajuste de las series de Fourier	52
4.4.1. Validación de los supuestos de Fourier	54
5. Conclusiones	61
A. Código en R project para hallar la función de distribución LDF	65
B. Código en R project para realizar las estimaciones de máxima verosimilitud de los parámetros de la distribución LDF	67

Índice de Tablas

4.1. Estadísticas Descriptivas de la cantidad total de lluvia mensual. Periodo multianual 1993-2013	42
4.2. Prueba de Independencia Von-Neumann Rankeada	44
4.3. Estimación de parámetros por máxima verosimilitud	44
4.4. Media ($\hat{\mu}$) y desviación estandar ($\hat{\sigma}$) estimadas	45
4.5. Valor p de la prueba de Kolmogorov-Smirnov y Criterio de Información Bayesiano (BIC)	48
4.6. Modelos seleccionados para cada uno de los meses	49
4.7. Prueba de razón de verosimilitud para el No. de Armónicos	52
4.8. Ajuste del modelo de Fourier para la media mensual estimada	53
4.9. Ajuste del modelo de Fourier para la desviación estandar mensual estimada	54
4.10. Oscilación de los modelos	54
4.11. Resumen test para la prueba de los supuestos del modelo de fourier de la media mensual estimada	56
4.12. Resumen test para la prueba de los supuestos del modelo de fourier de la desviación estandar mensual estimada	56

Índice de Figuras

2.1. Tipos de asimetría	16
2.2. Ejemplo Q-Q plot para la distribución Normal	27
3.1. Ejemplo Gráfica e_i vs tiempo. Tomado de Montgomery & Peck (2005) . . .	37
3.2. Ejemplo Gráfica e_i vs $\hat{y}_i$. Tomado de Montgomery & Peck (2005)	38
4.1. Estaciones meteorológicas CENICANÑA	41
4.2. Cantidad de precipitación total mensual	43
4.3. Función de Autocorrelación para la secuencia de precipitación mensual . . .	43
4.4. Histograma precipitación Sep, Oct, Nov y Dic vs ajuste modelos. Línea negra: Loi des fuites; Línea roja: Distribución Gamma; Línea verde: Distri- bución Lognormal; Línea azul: Distribución Weibull	46
4.5. Histograma precipitación Ene, Feb, Mar y Abr vs ajuste modelos. Línea negra: Loi des fuites; Línea roja: Distribución Gamma; Línea verde: Distri- bución Lognormal; Línea azul: Distribución Weibull	47
4.6. Histograma precipitación May, Jun, Jul y Ago vs ajuste modelos. Línea negra: Loi des fuites; Línea roja: Distribución Gamma; Línea verde: Distri- bución Lognormal; Línea azul: Distribución Weibull	47
4.7. Q-Q plot Lognormal vs Weibull para el mes de Septiembre	48
4.8. Q-Q plot Gamma vs Weibull para el mes de Noviembre	50
4.9. Q-Q plot Gamma vs Weibull para el mes de Febrero	50
4.10. Q-Q plot Gamma vs Weibull para el mes de Marzo	51

4.11. Q-Q plot Gamma vs Weibull para el mes de Agosto	51
4.12. Periodograma para la media mensual estimada	52
4.13. Periodograma para la desviación estandar mensual estimada	53
4.14. Ajuste de Fourier con dos armónicos para el modelo de media mensual estimada	55
4.15. Ajuste de fourier con un armónico para la desviación estandar mensual estimada	55
4.16. Q-Q plot Normal para los errores del modelo de media mensual estimada .	57
4.17. Q-Q plot Normal para los errores del modelo de desviación estandar mensual estimada	57
4.18. Diagrama de los errores vs valores ajustados para el modelo de media men- sual estimada	58
4.19. Diagrama de los errores vs valores ajustados para el modelo de desviación estandar mensual estimada	58
4.20. Errores a traves del tiempo para el modelo de media mensual estimada . . .	59
4.21. Errores a traves del tiempo para el modelo de desviación estandar mensual estimada	59

Capítulo 1

Introducción

La mayoría de actividades económicas e ingenieriles en una región pueden estar limitadas por el agua, principalmente en zonas áridas o semiáridas donde el aporte natural de este recurso es irregular o escaso. La evaluación del recurso hídrico permite una planificación con vista a la optimización del aprovechamiento del agua y su distribución entre los diferentes sectores demandantes: humano, agrícola, industrial, medioambiental, etc. (Roldán 1994). El estudio de procesos naturales relacionados con dicho elemento: flujos de ríos, evaporación natural y la precipitación han sido frecuentemente temática de investigación en trabajos desarrollados por diversas ramas de la ciencia. Cabe destacar el interés que los investigadores tienen hoy en día sobre el efecto que puede tener el cambio climático en ciclo hidrológico (Asín 2003).

La hidrología de una región depende principalmente del clima, de ahí la necesidad de conocer los procesos meteorológicos que lo determinan, entre estos cabe destacar la lluvia. Colombia por encontrarse ubicada sobre la zona tórrida o intertropical del planeta Tierra es altamente susceptible de presentar alteraciones climáticas debido a procesos de precipitación, los cuales generan reacciones adversas particularmente en el sector agrícola.

El estudio del fenómeno de la precipitación, considerado como una realización de un proceso estocástico involucra el análisis de dos componentes, cada uno de los cuales requiere un sub-modelo propio: la ocurrencia de lluvia, representada por una variable binaria y el segundo enfocado en modelar la cantidad de precipitación en los días lluviosos. El proceso de ocurrencia ha sido ampliamente estudiado haciendo uso de cadenas de Markov, modelos de renovación alternativa, además de modelos de regresión logística. Respecto de la cantidad de precipitación las técnicas de análisis incluyen el uso de distribuciones asimétricas positivas como la Log-normal, Weibull, Gamma, Exponencial, etc., además de distribuciones mixtas.

Específicamente, este documento se centra en el segundo componente del proceso de precipitación, es decir en la modelación de la cantidad de lluvia. El trabajo consta de 5 capítulos:

El capítulo 1, trata aspectos como el planteamiento de problema descrito de forma concreta, así como los diferentes objetivos considerados en el desarrollo del documento. Además de definir específicamente las razones que llevaron a la implementación de las metodologías propuestas para resolver el problema en cuestión. El capítulo 2, denominado Marco Teórico es dedicado a contextualización de términos necesarios para la comprensión de fenómeno objeto de estudio y la descripción teórica de las herramientas estadísticas utilizadas en la modelación de la cantidad de precipitación. En el capítulo 3, se describe el procedimiento metodológico propuesto para dar cumplimiento a los objetivos en este trabajo. En el capítulo 4, se lleva a cabo el ajuste de los modelos planteados para cada uno de los meses y mediante criterios de bondad de ajuste se selecciona el modelo que mejor describe el comportamiento de los registros para cada mes del año. Se utilizan las Series de Fourier para describir la variación estacional de la media y desviación estándar halladas en cada uno de los modelos asignados. Además de la verificación de los supuestos. Finalmente en el capítulo 5, se describe de forma concreta las conclusiones generadas con base a los resultados observados en el capítulo 4.

1.1. Planteamiento del Problema

La estimación y predicción de la cantidad de precipitación espacial y temporal es un problema de importancia fundamental en muchas aplicaciones en la agricultura, la hidrología y la ecología. Su variabilidad repercute, adversa o benéficamente, sobre casi todas las actividades del quehacer humano y de su conocimiento depende, en buena medida, el óptimo desarrollo social y económico del país (Bernal & et. al 1989).

En las regiones tropicales y subtropicales, la variabilidad de las precipitaciones constituye un factor limitante muy importante en la producción de alimentos, generación de energía hidroeléctrica, y cuestiones relacionadas con la salud (Sanso & Guenni 2000). La precipitación se presenta como un componente fundamental del sistema climático; en momentos actuales existe un amplio consenso sobre la necesidad de profundizar en el conocimiento sobre su variabilidad espacial y temporal dentro del contexto de cambio climático global (Citado por Guenni et al., 2008 de Solomon et al. 2007, p. 996).

A nivel nacional, la caracterización de la precipitación encarna un papel fundamental en el comportamiento del sistema agrícola de nuestro país, en la dinámica de enfermedades endémicas como el dengue y la malaria, así como en la distribución espacial y la frecuen-

cia de eventos hidrometeorológicos extremos causantes de inundaciones y corrimientos de tierra.

Este trabajo puede considerarse una herramienta importante para futuros estudios hidrológicos e hidráulicos que requieran el análisis de datos sobre precipitación, por ejemplo proyectos relacionados con la planificación para el uso de recurso hídrico, planes de ordenamiento urbano y agrícola principalmente para la optimización de la disponibilidad de agua con fines de riego en zonas de estudio, así como la determinación de la potencialidad de sequías climatológicas en lugares susceptibles (Dorado & et al. 2006).

La modelación de datos de precipitación desde una perspectiva estadística presenta el reto de hacer frente a una serie de tiempo que es no estacionaria y que además posee una distribución asimétrica que incluye un punto de masa en 0 (Sanso & Guenni 2000). Para afrontar dicho inconveniente han sido empleadas diversas técnicas, entre las cuales resaltan el uso de distribuciones tales como; *Exponencial*: (Todorovic & Woolhiser 1974); *mixta exponencial*: (Smith & Schreiber 1974), (Woolhiser & Pegram 1979), (Woolhiser & Roldán 1986), (Alcaide 1986); *gamma*: (Ison et al. 1971), (Katz 1977), (Ibáñez 1982), (Stern & Coe 1984); *uniforme*: (Haan et al. 1976), (Roldán 1979); *kappa*: (Mielke 1971,1973); *Weibull*: (Selker & Haith 1990), (Selker et al. 1998).

En este proyecto se hará uso de la distribución Gamma, Lognormal, Weibull y una denominada “loi des fuites”, las cuales han sido ampliamente usadas para modelar datos de precipitación. Esta última ha sido aplicada en lugares con climas áridos, semiaridos, mostrando buenos resultados en el ajuste en comparación con las tradicionales Gamma, Weibull, Lognormal, entre otras (Buishand 1977). La idea principal de este trabajo, es encontrar la mejor distribución que ajuste a la cantidad de precipitación total mensual para el periodo de registro 1993 – 2013.

Un factor importante a resaltar es la variación estacional de la media y desviación estandar mensual; debido a la variabilidad climática y la rotación del planeta sobre sus mismos ejes (Huynh, Anat & Apichai 1980). Por ende, se hará uso de las Series de Fourier, la cual es una metodología que permite describir dicha variación por medio de la suma de funciones trigonométricas Senos y Cosenos.

1.2. Objetivos

Objetivo General

Modelar el comportamiento de la cantidad de precipitación total mensual de la estación meteorológica Meléndez ubicada en la ciudad de Cali-Colombia para el periodo 1993-2013.

Objetivos Específicos

- Encontrar las estimaciones de los parámetros para los modelos haciendo uso del método de máxima verosimilitud.
- Evaluar la pertinencia de los modelos planteados haciendo uso de pruebas de bondad de ajuste.
- Encontrar las estimaciones de los coeficientes de la serie de Fourier, los cuales permiten describir la variación estacional de la media y desviación estandar mensual.

1.3. Justificación

En la actualidad, el uso de la información climática es de vital importancia para las actividades económicas y culturales de la sociedad, puesto que dichas actividades se encuentran altamente condicionadas por los fenómenos climatológicos. La repercusión del fenómeno climático de la precipitación, sobre diversos ámbitos del desarrollo humano genera gran expectativa por su modelación y pronóstico; puesto que conocer el comportamiento de esta variable permitirá de alguna forma establecer estrategias anticipadas que permitan mitigar los efectos adversos causados por este fenómeno.

El análisis de la información histórica, indica que las alteraciones que se producen en el régimen de lluvias en Colombia son causadas, en parte, por la variabilidad climática interanual, relacionada con el ciclo El Niño, La Niña - Oscilación del Sur, el cual ha sido causa de sequías extremas y lluvias extraordinarias en diferentes regiones geográficas del planeta (Montealegre & Pabon 2000).

“Colombia, por su localización en la Franja Intertropical de Convergencia (FIC) y por su accidentada fisiografía, es un país susceptible a catástrofes y desastres naturales, los cuales, en un 90 %, están asociados a fenómenos hidrometeorológicos. El Valle del Cauca es uno de los departamentos más afectados por su cercanía al Océano Pacífico, que es zona de influencia de fenómenos Océano-Atmosféricos cuasi periódicos, difíciles de comprender, como el fenómeno ENSO” (Carvajal, Jiménez & Materón 1998).

Las lluvias excesivas del segundo semestre del 2010 dieron lugar a inundaciones, deslizamientos y movimientos en masa que ocasionaron desastres de extraordinarias proporciones en el territorio nacional, en particular en las regiones Caribe y Andina, razón por la cual el gobierno tuvo que decretar la emergencia invernal en 28 de los 32 departamentos, exceptuando a San Andrés, Guaviare, Amazonas y Guainía, con la pérdida de 315 vidas humanas, afectación a más de 2 millones de personas y una inversión que supera los \$540.000 millones de pesos para atender la emergencia (León 2012).

Dado que el municipio de Santiago de Cali se encuentra a una altura de 995 msnm, consideraremos que su temperatura promedio anual es igual a los 22,6°C, según la información proporcionada por el IDEAM (2012c). Por otra parte, la clasificación de Richard Lang establecida en 1915 utiliza la precipitación anual en mm y la temperatura media anual en °C, relacionando dichos parámetros mediante el cociente entre la precipitación (P) y la temperatura (T), llamado factor de Lang (P/T). En base a dicho factor se obtienen seis clases de climas.

De acuerdo a la información anterior y sabiendo que la precipitación anual para la zona central del Valle del Cauca varía entre 1000 y 1500 mm total anual, se tiene que el clima para dicha zona estaría clasificada entre semiárido y semihúmedo. Debido a esto, se espera que el comportamiento de los datos sea más frecuente en cantidades de lluvia cercanas a cero, y algunas frecuencias correspondientes a cantidades de lluvias mayores, lo cual corresponde a distribuciones con curva en forma de J. De esta manera se justifica la utilización de las cuatro distribuciones involucradas en esta investigación.

1.4. Antecedentes

Buishand (1977). Modeló la cantidad de precipitación para periodos de 30 días (un mes) mediante la distribución gamma y una distribución denotada “loi des fuites” (LDF). Los datos usados consisten en registros de precipitación diaria para Winterswijk y Hoofddorp en Holanda con periodos de observación: 1880 – 1970 y 1861 – 1972 respectivamente. Consideró un día húmedo cuando la precipitación era mayor a 0.25 mm, la cual correspondía a la mínima cantidad registrada. Las estimaciones de los parámetros de las distribuciones propuestas fueron realizadas por máxima verosimilitud y usando el método de Newton-Raphson, tomando como valor semilla, las estimaciones de dichos parámetros encontradas por el método de los momentos.

Por otra parte, usó la serie finita de Fourier para describir la variación estacional de los parámetros, estimando los coeficientes de Fourier por mínimos cuadrados y probando la significancia del armónico mediante la estadística:

$$T_{n_0} = \frac{3(A_{n_0}^2 + B_{n_0}^2)}{\hat{\sigma}^2(n_0)}$$

Bajo el supuesto de independencia y normalidad de los errores, la distribución nula del estadístico T_{n_0} es $F(2, 11 - 2n_0)$. Si el n-ésimo armónico no es significativo, entonces la prueba es repetida sin tener en cuenta el armónico en cuestión.

Cabe resaltar que el autor empleó el coeficiente de correlación serial (scc) para verificar el supuesto de independencia entre cantidades de precipitación de días sucesivos, donde concluyó que no existe evidencia de correlación serial con un 95 % de confianza.

Finalmente realizó la comparación de los modelos propuestos mediante la prueba de bondad de ajuste chi-cuadrado, donde concluyo que la distribución LDF ajusta un poco mejor que la distribución gamma, en lugares con climas áridos.

Huynh et al. (1980). Usaron datos diarios de 56 estaciones meteorológicas ubicadas en el Noreste de Tailandia con registros de al menos 20 años, para modelar la cantidad de precipitación para cada uno de los meses del año. Los autores emplearon la distribución “lois des fuites” para tal fin, donde los parámetros fueron estimados por el método de los momentos. Usaron series de Fourier para describir la variación estacional de los parámetros, estimando los coeficientes de Fourier por mínimos cuadrados y probaron la significancia del armónico mediante el cociente:

$$(1/2)(A_k^2 + B_k^2)/\hat{\sigma}^2(v_j)$$

Y mediante el método empírico propuesto por Yevjevich (1972). Aquí v_j representa el armónico.

Por otra parte, compararon el modelo propuesto con la distribución gamma y la lognormal usando la prueba de Kolmogorv-Smirnov y la media y desviación estándar mensual empírica y la obtenida por los modelos. De los resultados obtenidos, encuentran que la distribución “lois des fuites” tiene un mejor ajuste en comparación con la distribución lognormal y gamma, particularmente en lugares con climas áridos o en precedencia de varios meses secos.

Stern & Coe (1984). Ajustan una cadena de Markov de primer orden con dos estados para el proceso de ocurrencia y una distribución gamma para la cantidad de precipitación en los días considerados húmedos. Los autores definieron un día húmedo cuando la cantidad de precipitación registrada fuera mayor a 0.1 mm, empleando registros de Morogoro, Tanzania con un periodo de observación de 1928 a 1981. Las estimaciones de los parámetros fueron realizadas por máxima verosimilitud para periodos de 5 días (73 periodos a lo largo del año) y describiendo la variación estacional de dichos parámetros mediante series de Fourier. Para estimar los coeficientes de la serie y probar la significancia de los armónicos, ellos usaron un modelo lineal generalizado con función de enlace log, tal como sigue:

$$\mu(t) = \exp[A_0 + \sum_{k=1}^m A_k \cos(\frac{2\pi tk}{26}) + B_k \sin(\frac{2\pi tk}{26})]$$

Donde μ corresponde a la media de la distribución gamma. El uso del modelo lineal generalizado es justificado debido al incumplimiento de la normalidad y varianza constante de los errores en los modelos de regresión simple.

Para determinar la significancia de los armónicos, los autores compararon la devianza de cada uno de los modelos incrementando el número de armónicos ajustados, mediante un análisis análogo al análisis de varianza usado con datos normales.

Ellos encontraron que tres armónicos son requeridos para describir la variación estacional de la media de la distribución gamma. Finalmente concluyeron que el modelo tanto de ocurrencia como de cantidad de precipitación enfocados como un modelo lineal generalizado tiene un buen ajuste a los datos de Morogoro, Tanzania. Resultados similares fueron obtenidos en Coe & Stern (1982), donde usaron registros de Zinder y Niger.

Roldán (1994). El autor en su trabajo “Tendencias actuales en el modelado de la precipitación diaria” efectúa una revisión de publicaciones realizadas por diversos autores sobre la modelación y análisis de la precipitación, lo anterior con el objetivo de identificar las técnicas más usadas para estudiar el comportamiento de dicha variable, dado que la mayoría de actividades económicas se encuentran limitadas por aspectos relacionados con la lluvia. Roldán afirman según estudios de Foufoula-Georgakalos & Georgakakos (1991) se deben establecer modelos de precipitación diaria basados en la metodología estocástica, diversos autores en trabajos posteriores concluyen que los procesos estocásticos describen tanto la ocurrencia como también la estructura de la cantidad de precipitación.

El autor efectúa además un análisis de la calidad de los datos recabados citando estudios como los de Woolhiser & Roldán (1986) y Woolhiser (1992) donde se identifican algunos factores que pueden incidir en la toma adecuada de los datos, tales como: la situación de la estación meteorológica, el tiempo de observación y los errores metodológicos de lectura de los datos.

Respecto de la modelación estocástica de la precipitación diaria, Roldan describe el proceso de precipitación mediante una secuencia de dos variables aleatorias:

$$\{(X_t, Y_t); \quad t = 1, 2, \dots\}$$

Donde X_t es una variable que representa el estado del proceso en el tiempo t . Su valor es igual a unidad si la lluvia es igual o mayor que un umbral T y a 0 en caso contrario. Y_t es una variable continua que representa la cantidad de lluvia en el día. Umbral para distinguir entre un día seco y uno lluvioso.

$$U(t) = Y(t) - T$$

Dado que el proceso de precipitación se describe en términos de la ocurrencia y la cantidad de lluvia. La estimación de los parámetros se llevó a cabo mediante la estimación de máxima verosimilitud para periodos de 14 días (26 periodos a lo largo del año), y usando series de Fourier para describir la variación estacional de los parámetros. Los coeficientes de Fourier también fueron estimados por máxima verosimilitud, usando procesos de optimización multivariados, tomando como valores semilla, las estimaciones por mínimos cuadrados de dichos coeficientes. Para probar la significancia de los armónicos, el autor empleó el test de razón de verosimilitud, encontrando tres armónicos significativos.

Finalmente, el autor concluye en su documento que los modelos de lluvia que describen la ocurrencia mediante una cadena de Markov de primer orden y la cantidad de precipitación a través de una distribución mixta exponencial, usando series de Fourier finitas para la variabilidad estacional con coeficientes estimados por técnicas de máxima verosimilitud, proporcionan una suficiente aproximación en una gran variedad de climas.

Selker, Duan & Grant (1998). Evaluaron la función de densidad de probabilidad para la precipitación diaria en el noroeste del pacifico en Estados Unidos. Para ello hicieron uso de tres distribuciones comúnmente usadas para tal fin, tales como, la distribución weibull, la gamma y la kappa.

Los autores emplearon los datos de precipitación diaria registrados por 99 estaciones meteorológicas a lo largo del noroeste del pacifico, con registros de más de 40 años. Además dividieron el registro histórico en periodos mensuales, con el fin de seleccionar para cada mes la distribución con mejor ajuste, usando la prueba chi-cuadrado y el diagrama de razón de los L-momentos. Finalmente encontraron que la distribución weibull tuvo el mejor ajuste para la mayoría de los meses, debido a sus grandes propiedades tales como, invertibilidad, integrabilidad, y estimaciones cercanas a los momentos probabilísticos; además de su flexibilidad y maleabilidad.

Suhaila & Jemain (2007b). Presentan un estudio estadístico de la precipitación, en la cual se compara diferentes tipos de distribución normal transformada que describen la distribución de la cantidad de lluvia en Malasia, considerando un periodo multi-año (21 a 35 años). Las distribuciones lognormal, normal asimétrica y la mixta lognormal fueron propuestas y probadas para identificar el modelo óptimo para la cantidad de precipitación diaria en distintas estaciones pluviométricas ubicadas en diversos lugares de Malasia. Las estimaciones de los parámetros para las distribuciones propuestas se realizaron por el método de máxima verosimilitud y empleando el algoritmo de Newton-Raphson. Por otra parte, para determinar el mejor modelo, los autores hicieron uso de pruebas de bondad de ajuste, específicamente las pruebas de Kolmogorov-Smirnov, Cramér-von Mises y la de Anderson-Darling, además de utilizar unas pruebas de bondad de ajuste más potentes que las tradicionales propuestas por Zhang (2002) y Zhang & Wu (2005). Finalmente, los

autores concluyeron que la mixta lognormal fue el modelo más apropiado para describir la cantidad de lluvia en la Malasia.

Liu et al. (2011). Modelan tanto la ocurrencia como la cantidad de precipitación diaria, usando datos recolectados en 10 estaciones ubicadas en la cuenca del río Yishu, para un periodo de observación de 46 años (1962 a 2007). Con respecto a la modelación de la ocurrencia de lluvia, los autores utilizan cadenas de Markov de primer orden, mientras que para la cantidad de precipitación en los días húmedos (consideraron un día húmedo cuando la cantidad de precipitación era mayor a 0.25 mm), plantean 4 modelos: La distribuciones exponencial, gamma, lognormal, y por último, la mezcla de dos distribuciones exponencial (mixta exponencial). Los parámetros de estos modelos fueron estimados para cada uno de los meses del año, empleando técnicas de máxima verosimilitud, desarrolladas con algoritmos genéticos, y mediante Series de Fourier para describir la variación estacional de los parámetros. Por otra parte, para la comparación de los modelos propuestos a la cantidad de lluvia, los autores hacen uso del criterio de información Bayesiano (BIC), y la estimación de la varianza mensual de precipitación, en donde finalmente concluyen que las distribuciones lognormal y mixta exponencial ajustan de mejor manera a los datos de la cantidad de lluvia, en comparación con las distribuciones gamma y exponencial.

Capítulo 2

Marco Teórico

La modelación de la cantidad de precipitación diaria requiere contextualizar ciertas definiciones y teorías estadísticas necesarias para abordar el proceso de análisis de dicha variable.

2.1. Marco Teórico Conceptual

2.1.1. El clima

El clima es el conjunto fluctuante de las condiciones atmosféricas, caracterizado por los estados y evoluciones del estado del tiempo, durante un periodo de tiempo y en un lugar o región dada, y controlado por los denominados factores forzantes, factores determinantes y por la interacción entre los diferentes componentes del denominado sistema climático (atmósfera, hidrosfera, litosfera, criósfera, biosfera y antropósfera) (IDEAM 2012*b*).

2.1.2. Ciclo hidrológico

El ciclo hidrológico está gobernado por los siguientes procesos naturales que ocurren de manera continua:

El sol calienta las aguas de la superficie terrestre, una cantidad del vapor de agua producido es enviado a la atmósfera mediante corrientes de aire ascendentes. El vapor se condensa en forma de nubes a medida que las temperaturas descienden, esto desencadena la precipitación de agua en forma de lluvia o nieve sobre la superficie terrestre. La parte del agua que escurre a lo largo de las laderas puede ser interceptada por las depresiones

naturales del terreno, en donde se evapora o infiltra; o, por último, se mueve a través de los drenajes naturales de la cuenca y forma el flujo superficial. Estos tres flujos superficial, subsuperficial y subterráneo conforman la escorrentía, que integra los cauces de las corrientes, alimenta los diferentes almacenamientos y drena finalmente al mar, donde el ciclo se renueva (SIAC 2013).

2.1.3. Definición de lluvia

Fenómeno atmosférico consistente en una precipitación acuosa en forma de gotas líquidas, cuyo diámetro se halla generalmente comprendido entre 0,5 y 7 mm, y que caen a una velocidad del orden de los 3 m/s (IDEAM 2012c).

2.1.4. Medición de la lluvia

El volumen de lluvia se mide en milímetros. Un milímetro equivale a un litro de agua por metro cuadrado. Los dos principales aparatos de medición de la precipitación son el pluviómetro y el pluviógrafo. El primero mide el volumen total de lluvia caída durante el día meteorológico (7 A.M. a 7 A.M. del día siguiente). El segundo es un aparato de registro continuo que permite determinar la intensidad de la precipitación en un intervalo de tiempo dado (IDEAM 2012c).

2.1.5. Zona de Convergencia Intertropical (ZCIT)

Es una franja de bajas presiones ubicada en la zona ecuatorial, en esta confluyen los vientos alisios del sureste y del noreste. Debido a las altas temperaturas las masas de aire son forzadas a ascender originando abundante nubosidad y fuertes precipitaciones, algunas acompañadas de descargas eléctricas. La ZCIT no es uniforme ni continua, se puede interrumpir y su grosor variar de un sitio a otro, también su comportamiento en zonas marítimas y continentales. (Citado por Gómez & Hernández (2013) de Pabón et al. (1998))

2.1.6. Fenómeno del Niño y de la Niña

El ciclo conocido como “el Niño” y su fase opuesta “la Niña” son la causa de la mayor señal de variabilidad climática en la franja tropical del océano Pacífico, en la escala interanual. Son las componentes oceánicas del ENOS (Oscilación del Sur) que corresponde a la aparición, de tiempo en tiempo, de aguas superficiales relativamente más cálidas (El

Niño) o más frías (La Niña) por lo normal en el Pacífico tropical central y oriental, frente a las costas del norte de Perú, Ecuador y sur de Colombia (SIAC 2013).

2.1.7. Régimen de la precipitación en Colombia

El patrón de la lluvia total anual en el país puede esquematizarse en la siguiente forma (IDEAM 2012*c*):

- Las menores lluvias se presentan en la Alta Guajira con totales de 500 mm y menos; los núcleos máximos se han registrado en la región Pacífica con totales anuales de más de 10000 mm.
- La región Caribe registra lluvias entre 500 y 2000 mm con un gradiente muy definido en dirección Sur.
- La región Andina posee una gran diversidad pluviométrica, con lluvias relativamente escasas (hasta 2000 mm) a lo largo de la Cordillera Oriental y en los valles del Alto Magdalena y Alto Cauca y núcleos máximos (de 3000 a 5000 mm) en las cuencas del Medio Magdalena y Medio Cauca.
- En la Orinoquia generalmente predominan las lluvias altas de 2000 a 3000 mm en su parte central y Oriental, aun cuando hacia el piedemonte pueden observarse hasta 6000 mm y por el contrario en el extremo Norte de Arauca las lluvias pueden ser menores de 1500 mm.
- La mayor parte de la Amazonia recibe entre 3000 y 4500 mm por año, mientras que en la región Pacífica se reciben entre 3000 y 12000 mm.

2.1.8. Régimen de la precipitación en el Valle del Cauca

En el departamento del Valle del Cauca se presenta una distribución de lluvias de tipo bimodal, con dos períodos húmedos o lluviosos (Marzo a Mayo y Septiembre a Noviembre) y dos períodos secos y/o de lluvias moderadas (Diciembre a Febrero y Junio a Agosto). Este régimen es característico de la Zona de Convergencia Intertropical (ZCIT).

- En la vertiente del Pacífico de la Cordillera Occidental la precipitación media anual puede fluctuar entre 4000 mm en la parte baja y unos 2500 en la parte alta, exceptuando el valle alto y medio del Río Dagua, con una precipitación media inferior a 2000 mm (debido al efecto abrigo).

- La vertiente oriental o interior de la Cordillera Occidental presenta una precipitación media que puede oscilar entre 1500 y 2000 mm anuales.
- La parte plana del valle del Río Cauca presenta una precipitación media anual inferior a los 1500 mm, debido a que buena parte de las nubes que logran pasar la Cordillera Occidental son arrastradas por los vientos occidentales hacia la Cordillera Central o posiblemente son disipadas por el calor del fondo del valle.
- En la Cordillera Central la precipitación media anual es ligeramente superior a los 1500 mm entre los 1800 msnm y los 3500 msnm de altitud. En las partes más elevadas, por encima de los 3500 msnm, la precipitación se reduce gradualmente hasta unos 1000 mm anuales.

2.2. Marco Teórico Estadístico

A continuación se describirán los métodos que serán empleados en la modelación de la cantidad de precipitación diaria en el Valle del Cauca.

2.2.1. Distribución de probabilidad

La distribución de probabilidad de una variable aleatoria, es una función que asigna a cada suceso definido sobre la variable, la probabilidad de que dicho suceso ocurra. Esta se encuentra completamente especificada por la función de distribución, cuyo valor en cada real x es la probabilidad de que la variable aleatoria sea menor o igual que x .

Dada una variable aleatoria X , su función de distribución, $F_X(x)$ es (Maronna 1995):

$$F_X(x) = P(X \leq x)$$

Distribución de variable continua

Una variable X tiene distribución absolutamente continua si existe una función $f_X : R \rightarrow R_+$ (llamada densidad de X) tal que (Maronna 1995):

$$P(X \in A) = \int_A f_X(x)dx \quad \forall A \subseteq R$$

En particular, tomando $A = (a, b]$, resulta para todo intervalo:

$$P(a < X \leq b) = \int_a^b f_X(x)dx$$

Si se hace $a = -\infty$ queda

$$F_X(x) = \int_{-\infty}^{\infty} f_X(t)dt \quad \forall x;$$

y por lo tanto

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

Aplicando a $F_X(x)$ el Teorema Fundamental del Cálculo Integral, se obtiene que para una distribución absolutamente continua, $F_X(x)$ es una función continua para todo x , y su derivada es $f_X(x)$ en todos los x donde f_X es continua.

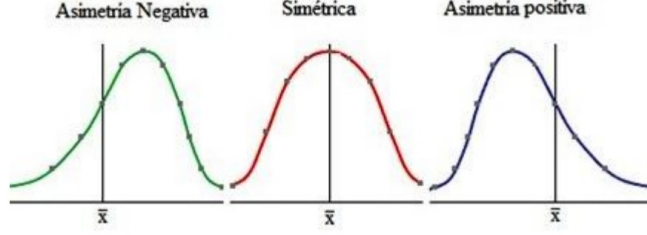
Entre las distribuciones de variable continua se encuentran: Distribución Gamma, Distribución Lognormal, Distribución Weibull, Distribución Normal, Distribución t de Student, Distribución Exponencial, entre otras.

Asimetría

La asimetría es una medida de forma de una distribución que permite identificar y describir la manera como los datos tiende a reunirse de acuerdo con la frecuencia con que se hallen dentro de la distribución. Permite identificar las características de la distribución de datos sin necesidad de generar el gráfico. La asimetría presenta las siguientes formas (Joanest & Gill 1998):

- **Asimetría negativa o a la izquierda:** Se da cuando en una distribución la minoría de los datos está en la parte izquierda de la media. Este tipo de distribución presenta un alargamiento o sesgo hacia la izquierda, es decir, la distribución de los datos tiene a la izquierda una cola más larga que a la derecha.
- **Simétrica:** Se da cuando en una distribución se distribuyen aproximadamente la misma cantidad de los datos a ambos lados de la media aritmética.
- **Asimetría positiva o a la derecha:** Se da cuando en una distribución la minoría de los datos está en la parte derecha de la media aritmética. Este tipo de distribución presenta un alargamiento o sesgo hacia la derecha, es decir, la distribución de los datos tiene a la derecha una cola más larga que a la izquierda.

Figura 2.1: Tipos de asimetría



Definición. Sea $\{x_i\}$ una muestra de n observaciones, y μ, σ^2 denotan la media y varianza de $\{x_i\}$. Los momentos centrales m_j son definidos por (Joanest & Gill 1998):

$$m_j = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^j$$

Luego la asimetría γ_1 está definida como sigue:

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}$$

Mientras que la asimetría muestral se calcula como:

$$skew = \frac{m_3}{m_2^{3/2}} = \frac{1/n \sum_{i=1}^n (x - \bar{x})^3}{(1/n \sum_{i=1}^n (x - \bar{x})^2)^{3/2}} \quad (2.1)$$

2.2.2. Método de Máxima Verosimilitud (MLE)

De acuerdo con Gujarati (2001), un método de estimación puntual, con algunas propiedades teóricamente más fuertes que las del método de mínimos cuadrados ordinarios, es el método de Máxima Verosimilitud (MV), también conocido como método de “razón de varianza mínima”. La evaluación por el método de máxima verosimilitud procura encontrar los valores más probables para un conjunto de datos, maximizando el valor de lo que se conoce como *función de verosimilitud*.

Supónaga que se tiene una muestra $x_1, x_2, \dots, x_n$ de n observaciones independientes extraídas de una función de distribución desconocida con función de densidad $f_0(\cdot)$. Se sabe, sin embargo, que f_0 pertenece a una familia de distribuciones $\{f(\cdot|\theta), \theta \in \Theta\}$, llamada modelo paramétrico, de manera que f_0 corresponde a $\theta = \theta_0$, el cual es el verdadero valor del parámetro. Se desea encontrar el valor de $\hat{\theta}$ que esté lo más cercano posible al verdadero valor θ_0 .

La idea de este método es el de encontrar primero la función de densidad conjunta de todas las observaciones, que bajo condiciones de independencia, es:

$$f(x_1, x_2, \dots, x_n | \theta) = f(x_1 | \theta) f(x_2 | \theta) \dots f(x_n | \theta)$$

Esta es la función de verosimilitud, y se denota como sigue:

$$L(\theta | x_1, \dots, x_n) = \prod_{i=1}^n f(x_i | \theta)$$

En la práctica, se suele utilizar el logaritmo de esta función:

$$\hat{L}(\theta | x_1, \dots, x_n) = \ln(L) = \sum_{i=1}^n \ln f(x_i | \theta)$$

El método de máxima verosimilitud estima θ_0 buscando un valor de θ que maximiza $\hat{\mathcal{L}}(\theta | x)$. Este es el llamado **estimador de máxima verosimilitud** (MLE) de θ_0 .

2.2.3. Distribución Gamma

La distribución Gamma es una distribución adecuada para modelizar el comportamiento de variables aleatorias continuas con asimetría positiva. Es decir, variables que presentan una mayor densidad de sucesos a la izquierda de la media que a la derecha. En su expresión se encuentran dos parámetros, siempre positivos, (k) y (λ) de los que depende su forma y alcance por la derecha, y también la función Gamma $\Gamma(k)$, responsable de la convergencia de la distribución (Buishand 1977).

Una variable aleatoria X sigue una distribución Gamma, si su función de densidad está dada como sigue:

$$f(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)} \quad (2.2)$$

$$x > 0, \quad k > 0, \quad \lambda > 0$$

Donde $\Gamma(k)$ es la función Gamma, definida como sigue para valores no enteros:

$$\Gamma(k) = \int_0^{\infty} x^{k-1} e^{-x} dx$$

λ denota el parámetro de escala y k el parámetro de forma. Si $k > 1$ se deduce por diferenciación de 2.2 que existe una moda en $(k - 1)/\lambda$. Si $k < 1$ La curva de densidad tiene forma de J y es infinita en el origen. Para $k = 1$ se obtiene la distribución Exponencial, la cual tiene función de densidad:

$$f(x) = \lambda e^{-\lambda x}$$

Otro caso especial de la distribución Gamma es la distribución Chi-cuadrado.

Los momentos de la distribución Gamma son (Buishand 1977):

$$\mu_1' = k/\lambda \quad (2.3)$$

$$\mu_2 = k/\lambda^2 \quad (2.4)$$

$$\mu_3 = 2k/\lambda^3 \quad (2.5)$$

$$\mu_4 = (6k + 3k^2)/\lambda^4 \quad (2.6)$$

$$C = \sqrt{\mu_2}/\mu_1' = 1/\sqrt{k} \quad (2.7)$$

$$\gamma = 2/\sqrt{k} \quad (2.8)$$

De 2.7 y 2.8 Se deduce que el cociente γ/C es siempre 2, sin importar los parámetros de la distribución. Las estimaciones de k y λ pueden ser obtenidas por el método de los momentos o el método de máxima verosimilitud (ML) (Thom 1958).

Las estimaciones ML de λ y k se obtienen a partir de la función de log-verosimilitud (Thom 1958):

$$-(nk)\ln(\lambda) - n\ln(\Gamma(k)) + (k - 1) \sum \ln(x_i) - \lambda \sum x_i \quad (2.9)$$

Maximizando la expresión en 2.9 con respecto a λ se tiene:

$$\hat{\lambda} = k/\bar{x} \quad (2.10)$$

Mientras que para el parámetro k :

$$\ln(\hat{k}) - \Psi(\hat{k}) = \ln(\bar{x}) - \frac{1}{n} \sum \ln(x) \quad (2.11)$$

Esta ecuación es implícita en $\hat{k}$ pero puede resolverse con algunas dificultades usando una aproximación a $\ln(\hat{k}) - \psi(\hat{k})$ desarrollada por (Thom 1958).

Siguiendo la metodología planteada por Thom (1958) se tiene que el estimador de máxima verosimilitud para k es:

$$\hat{k} = \frac{1 + \sqrt{1 + 4A/3}}{4A} \quad (2.12)$$

Donde $A = \ln(\bar{x}) - \frac{1}{n} \sum \ln(x)$

2.2.4. Distribución “loi des fuites”

La segunda distribución de probabilidad la cual es considerada en este proyecto puede ser descrita como sigue. Suponga que la precipitación ocurre en lluvias instantáneas según un proceso Poisson con intensidad media o tasa $1/\mu$, es decir, el número de lluvias en un intervalo de tamaño t se distribuye Poisson con media t/μ . La cantidad de precipitación en una lluvia en particular se asume:

- Independiente del proceso de ocurrencia.
- Mutuamente independiente.
- Distribuida exponencial con media $1/\rho$.

El proceso descrito aquí fue sugerido como un modelo para precipitaciones en periodos áridos por Fisher & Cornish (1960). Bernier & Fandoux (1970) aplicaron este proceso exitosamente al ajuste de la distribución del total mensual para las series de precipitación en Francia y anteriormente había sido usada para describir la distribución de los flujos de escape de los conductos de gases, por lo que la denominaron “loi des fuites” (LDF).

Sea x_t la cantidad de precipitación total en un periodo de tamaño t . La distribución de probabilidad de x_t esta dada como sigue (Ribstein 1983):

$$P(x_t \leq x) = e^{-\theta} + \sum_{k=1}^{\infty} \frac{\theta^k e^{-\theta}}{k!} (1 - \sum_{i=0}^{k-1} \frac{(\rho x)^i}{i!} e^{-\rho x}) \quad (2.13)$$

De 2.13 se deduce por diferenciación (Ribstein 1983):

$$\begin{cases} P(x_t = 0) = e^{-\theta}, \\ P(x < x_t < x + dx) = e^{-\theta - \rho x} \sqrt{\frac{\rho\theta}{x}} I_1(2\sqrt{\rho\theta x}), \quad x > 0 \end{cases} \quad (2.14)$$

Donde I_1 representa una función de Bessel modificada de orden 1.

Por otra parte, para derivar los momentos de x_t se puede hacer uso de la función generadora de momentos (Cox 1962):

$$f(s) = E(e^{-sx_t}) = e^{-\theta} \exp\left(\frac{\rho\theta}{\rho + s}\right) \quad (2.15)$$

Con $\theta = t/\mu$. De 2.15 se deduce:

$$\log(f(s)) = \theta(-1 + \frac{1}{1+s/\rho}) = \theta \sum_{m=1}^{\infty} (-1)^m \left(\frac{s}{\rho}\right)^m$$

O escrito de otra manera:

$$\log(f(s)) = \sum_{m=1}^{\infty} (-1)^m \chi_m \frac{s^m}{m!}$$

Donde χ_m es, por definición, el m-ésimo cumulante de x_t . Así χ_m satisface la relación:

$$\chi_m = \frac{m!\theta}{\rho^m} \quad (2.16)$$

De la expresión 2.16 las expresiones para los momentos y momentos centrales pueden ser obtenidos:

$$\mu_1' = \chi_1 = \theta/\rho \quad (2.17)$$

$$\mu_2 = \chi_2 = 2\theta/\rho^2 \quad (2.18)$$

$$\mu_3 = \chi_3 = 6\theta/\rho^3 \quad (2.19)$$

$$\mu_4 = \chi_4 + 3\mu_2^2 = (24\theta + 12\theta^2)/\rho^4 \quad (2.20)$$

$$C = \sqrt{2/\theta} \quad (2.21)$$

$$\gamma = 3/\sqrt{2\theta} \quad (2.22)$$

De la ecuación 2.21 y 2.22 se deduce que el cociente γ/C es siempre 1.5, sin importar los parámetros de la distribución.

Las estimaciones de los parámetros de la distribución LDF por el método de máxima verosimilitud es complicado y se requiere el uso del algoritmo de Newton-Raphson. La función de verosimilitud $L(\rho, \theta)$ está dada por (Buishand 1977):

$$L(\rho, \theta) = e^{-N\theta} \prod_{i=1}^m e^{-\rho x_i} \left\{ \sum_{k=1}^{\infty} \frac{(\rho\theta)^k x_i^{k-1}}{k!(k-1)!} \right\} \quad (2.23)$$

2.2.5. Distribución Lognormal

Una variable aleatoria X se dice que tiene una distribución Lognormal con parámetros $\mu \in \mathbb{R}$ y $\sigma > 0$ si $\ln(X)$ sigue una distribución Normal con media μ y desviación estándar σ . Equivalentemente, $X = \exp(Y)$ donde Y se distribuye Normal con media μ y desviación estándar σ . La distribución Lognormal es usada para modelar cantidades aleatorias continuas con distribución de frecuencia asimétrica, tales como ingresos, tiempos de vida, etc (Limpert et al. 2001).

La función de densidad de la Lognormal está dada por (Suhaila & Jemain 2007b):

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left[\frac{\ln x - \mu}{\sigma}\right]^2\right) \quad x > 0, \quad \sigma > 0, \quad -\infty < \mu < \infty \quad (2.24)$$

Note que la distribución lognormal es unimodal y es asimétrica a la derecha. Específicamente, si $m = \exp(\mu - \sigma^2)$ entonces:

- $f(x)$ Se incrementa en $(0, m)$ y decrece en (m, ∞) , así que la moda ocurre en $x = m$.
- $f(x) \longrightarrow 0$ cuando $x \longleftarrow \infty$.
- $f(x) \longrightarrow 0$ cuando $x \downarrow 0$.

El valor esperado y la varianza son:

$$E(X) = e^{\mu + \frac{\sigma^2}{2}} \quad (2.25)$$

$$V(X) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2} \quad (2.26)$$

Por otra parte, la función de distribución de la Lognormal está dada por:

$$F(x) = \phi\left(\frac{\ln(x) - \mu}{\sigma}\right), \quad x > 0 \quad (2.27)$$

Y la función cuantil:

$$F^{-1}(p) = \exp(\mu + \sigma\phi^{-1}(p)), \quad 0 < p < 1 \quad (2.28)$$

Los momentos de la distribución Lognormal pueden ser calculados de la función generadora de momentos de la distribución normal:

$$E(X^k) = \exp(k\mu + \frac{1}{2}k^2\sigma^2), \quad k \in \mathbb{N}$$

Los parámetros son estimados por máxima verosimilitud; las expresiones son presentadas a continuación (Suhaila & Jemain 2007b):

$$\hat{\mu} = \frac{\sum_{i=1}^n \ln x_i}{n} \quad (2.29)$$

y

$$\hat{\sigma} = \left[\frac{\sum_{i=1}^n (\ln x_i - \hat{\mu})^2}{n} \right]^{1/2} \quad (2.30)$$

2.2.6. Distribución Weibull

La función de densidad $f(x)$ y la función de distribución $F(x)$ de la distribución Weibull de 2 parámetros son definidas como sigue (Wong Raymond 1977):

$$f(x) = \frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{x}{\beta}\right)^\alpha\right), \quad x > 0, \alpha > 0, \beta > 0 \quad (2.31)$$

$$F(x) = 1 - \exp\left(-\left(\frac{x}{\beta}\right)^\alpha\right), \quad x > 0, \alpha > 0, \beta > 0 \quad (2.32)$$

Donde α y β son los parámetros de forma y escala respectivamente. El r -ésimo momento alrededor de cero de la variable aleatoria Weibull X esta dado por:

$$E(X^r) = \beta^r \Gamma(1 + r/\alpha)$$

y

$$\Gamma(t) = \int_0^\infty \xi^{t-1} \exp(-\xi) d\xi$$

Es la función gamma. En particular, la media y varianza de la distribución Weibull son $\mu = \beta \Gamma(1 + 1/\alpha)$ y $\sigma^2 = \beta^2 (\Gamma(1 + 2/\alpha) - (\Gamma(1 + 1/\alpha))^2)$, respectivamente.

Por otra parte, los parámetros son estimados usualmente por máxima verosimilitud, sin embargo se requiere el uso del algoritmo Newton-Raphson para resolver las ecuaciones de verosimilitud que están dadas en su forma implícita (Wong Raymond 1977).

El método de los momentos es usado para proveer estimaciones iniciales de α y β para el proceso iterativo de las estimaciones máximo verosímil. Considerando las relaciones:

$$\mu_2/(\mu_1')^2 = \Psi(\alpha) = \Gamma(1 + 2/\alpha)/(\Gamma(1 + 1/\alpha))^2$$

$$\beta = (\mu_2/\mu_1') \chi(\alpha)$$

Donde $\chi(\alpha) = \Gamma(1 + 1/\alpha)/\Gamma(1 + 2/\alpha)$. El primer y segundo momento muestral ($\tilde{\mu}_1'$ y $\tilde{\mu}_2'$) pueden ser usados para obtener las estimaciones de $\tilde{\alpha}$ y $\tilde{\beta}$ por el método de los momentos.

Ahora bien, las estimaciones de máxima verosimilitud de α y β son las soluciones de (Wong Raymond 1977):

$$\frac{\partial \ln L}{\partial \alpha} = 0 \quad (2.33)$$

$$\frac{\partial \ln L}{\partial \beta} = 0 \quad (2.34)$$

Donde $\ln L = N \ln \alpha - N \alpha \ln \beta + (\alpha - 1) \sum_{i=1}^N \ln x_i - \sum_{i=1}^N (\frac{x_i}{\beta}) \alpha$. Para más detalles sobre el proceso de estimación de los parámetros ver Wong Raymond (1977)

2.2.7. Pruebas de Bondad de Ajuste

Las pruebas de bondad de ajuste miden la compatibilidad de una muestra aleatoria con una función de distribución teórica. En otras palabras, estos tests muestran que tan bueno es el ajuste de la distribución seleccionada al conjunto de datos. A continuación se describirán algunas de las pruebas que existen, las cuales serán de utilidad para medir el ajuste de los modelos seleccionados a los datos de cantidad de precipitación diaria.

Prueba de Kolmogorov-Smirnov

El test de Kolmogorov-Smirnov (KS) es una prueba no paramétrica que se utiliza para medir el grado de concordancia existente entre la distribución de un conjunto de datos ($F_n(x)$) y una distribución teórica especificada ($F_0(x)$).

Suponga que se tiene una muestra aleatoria $x_1, x_2, \dots, x_n$ proveniente de una distribución F y se desea probar $H_0 : F = F_0$ para una cierta distribución F_0 . El estadístico de prueba D está basado sobre la mayor diferencia (vertical) entre $F_0(x)$ y $F_n(x)$, el cual se define como sigue:

$$D_n = \sup_{x \in R} |F_n(x) - F_0(x)| \quad (2.35)$$

Para estudiar el comportamiento de D_n , el cual toma valores en el intervalo $(0,1)$, y que a medida que el tamaño de muestra aumenta tiende a tomar valores más próximos a cero (Teorema de Glivenko-Canteli), se utilizan los estadísticos (Suhaila & Jemain 2007 a):

$$D_n^+ = \sup_{x \in R} (F_n(x) - F_0(x)) \quad (2.36)$$

$$D_n^- = \sup_{x \in R} (F_0(x) - F_n(x)) \quad (2.37)$$

que permiten comprobar que $D_n = \max(\frac{i}{n} - F_0(x_{(i)}), F_0(x_{(i)}) - \frac{i-1}{n})$, $i = 1, \dots, n$. La distribución de D_n es independiente de la distribución formulada en la hipótesis nula, ya que la transformación de los estadísticos ordenados de una variables continua por su función de distribución da lugar a los estadísticos ordenados de una variable distribuida $U(0, 1)$.

Criterio de Información Bayesiano (BIC)

El criterio de información bayesiano (BIC) o también conocido como el criterio de Schwarz, es un criterio para la selección de modelos entre un conjunto finito de modelos. En el contexto de procedimientos basados en la verosimilitud, Schwarz (1978) sugirió que el AIC podría no ser asintóticamente justificable, y presentó un criterio de información alternativo a partir de un enfoque bayesiano. Con este criterio, se penaliza el número de parámetros con $\ln(n)$, así:

$$BIC(k) = -2\ln(L(\hat{\theta}(k))) + k\ln(n)$$

Donde $L(\hat{\theta}(k))$ es la función de verosimilitud de las observaciones, $\hat{\theta}(k)$ es la estimación máximo verosímil del vector de parámetros $\theta = (\mu, L, \Psi)$ y k es el número de parámetros estimados dentro del modelo, mientras que n es el tamaño muestral.

El criterio es obtenido analizando el comportamiento de la probabilidad a posteriori del j -ésimo modelo cuando n tiende a infinito, bajo la hipótesis de algunas distribuciones a priori de los parámetros.

$$\ln f(X|M_j) = L_j(\hat{\theta}_j|X) + \ln P(\hat{\theta}_j|M_j) + \frac{k_j}{2}\ln(2\pi) - \frac{k_j}{2}\ln(n) + \frac{1}{2}\ln|R_j|$$

Donde:

- $f(X|M_j)$ es la verosimilitud marginal de los datos para el modelo M_j .
- $L_j(\hat{\theta}_j|X)$ es la verosimilitud del modelo M_j , donde $\hat{\theta}_j$ es el estimador máximo verosímil del parámetro θ_j .
- $P(\hat{\theta}_j|M_j)$ es la probabilidad a priori para los parámetros.
- k_j es el número de parámetros libres estimados dentro del modelo M_j .
- $|R_j|$ es el determinante de $R_j = n.S_j$, donde S_j es la matriz de covarianzas del estimador $\hat{\theta}_j$.

Para un n suficientemente grande,

$$\ln f(X|M_j) \simeq L_j(\hat{\theta}_j|X) - \frac{k_j}{2} \ln(n)$$

Esta expresión fue obtenida por primera vez por Schwarz (1978), que propuso escoger el modelo que conduzca a un valor máximo de esta cantidad. Una forma equivalente de este criterio es calcular para cada modelo la cantidad

$$BIC(M_j) = -2\ln L_j(\hat{\theta}_j|X) + k_j \ln(n) \quad (2.38)$$

y seleccionar aquel modelo para el cual esta cantidad sea mínima. De esta manera, este criterio pondera la desviación del modelo, medida por $-\ln L_j(\hat{\theta}_j|X)$, con el número de parámetros. Si introducimos más parámetros en el modelo, mejorará el ajuste, con el que aumentará el soporte o disminuirá la desviación, y este efecto quedará compensado por el aumento del número de parámetros que aparece en $k_j \ln(n)$.

Gráfico Cuantil-Cuantil (Q-Q plot)

Un gráfico cuantil-cuantil es un método gráfico para diagnosticar las diferencias entre la distribución de probabilidad de una población de la que se ha extraído una muestra aleatoria (función de distribución empírica) y una distribución usada para la comparación. Una forma básica de gráfico surge cuando la distribución para la comparación es una distribución teórica (Wilk & Gnanadesikan 1968).

El cuantil q de $F(x)$ es el número x_q tal que

$$F(x_q) = q$$

Si F^{-1} es la inversa de F , entonces

$$x_q = F^{-1}(q)$$

Un estimado razonable de la función de distribución $F(x)$ de una variable aleatoria X es $F_n(x)$ definida como una proporción de las muestras que son más pequeñas que x_i , es decir:

$$F_n(x_i) = \frac{i}{n}$$

No obstante, i/n , es igual a 1 cuando $i = n$, mientras que la mayoría de las distribuciones teóricas tienden a 1 cuando x va a infinito. De modo que se usa:

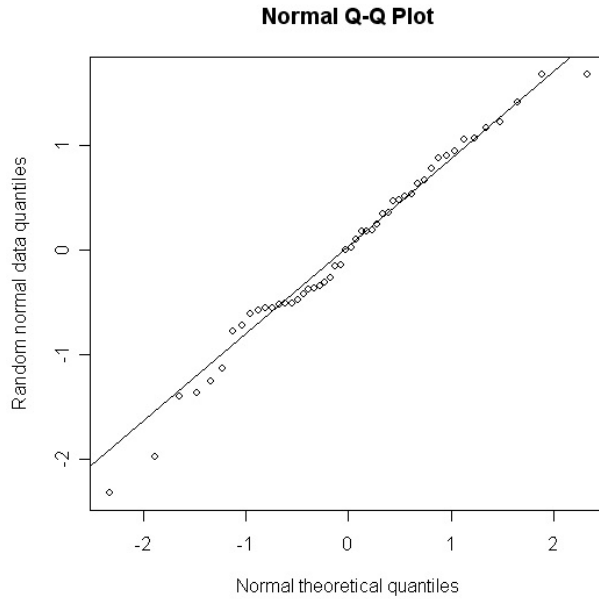
$$F_n(x_i) = \frac{i - 0,5}{n} \quad (2.39)$$

Una forma fácil para ver si los datos x_i se ajustan a una función de distribución teórica determinada es graficando los puntos $(x_1, 0,5/n), (x_2, 1,5/n), \dots, (x_n, (n - 0,5)/n)$ y comparando este resultado con el gráfico de la distribución teórica que se cree ajusta al conjunto de datos. El problema radica en que muchas de las funciones de distribución tienen forma de S y es muy difícil ver sus similitudes a ojo (Wilk & Gnanadesikan 1968).

La segunda forma es usando un q-q plot. Para cualquier conjunto de datos continuo un q-q plot es un gráfico de los cuantiles de la función de distribución teórica $x_{q_i}^T$ que se obtienen como $x_{q_i}^T = F^{-1}(q_i)$, versus los cuantiles de la distribución muestral $F_n(x)$, denotados por $x_{q_i}^M = F_n^{-1}(q_i)$, para $i = 1, 2, \dots, n$.

Si las funciones de distribución tienen una inversa, es muy fácil determinar los x_{q_i} , en otros casos, se tienen que usar tablas e interpolar los valores si es necesario.

Figura 2.2: Ejemplo Q-Q plot para la distribución Normal



2.2.8. Series de Fourier

La Transformación de Fourier fue introducida por primera vez en 1807 por el matemático francés, Jean Baptiste Joseph Fourier (1768 – 1830), como resultado de sus trabajos en

transmisión de calor donde empleaba estas transformaciones para solucionar las ecuaciones involucradas. Desde su presentación, esta técnica tuvo algunos opositores, como fueron: Euler, Laplace y Lagrange, entre otros. La publicación formal de la Transformación de Fourier fue con el libro *La teoría analítica del calor* (Fourier 1822) donde se afirma la convergencia de una transformación de funciones periódicas en sumas infinitas de funciones trigonométricas.

Transformación de Fourier o Serie de Fourier

Una serie de la forma:

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} \{a_k \cos(kx) + b_k \sin(kx)\} \quad (2.40)$$

Es llamada serie trigonométrica o serie o sinusoidal. Dado $f(x)$ una función definida e integrable en el intervalo $[-\pi, \pi]$, entonces la serie de Fourier asociada a la función $f(x)$ es una serie trigonométrica de la forma 2.40 y los parámetros a_k y b_k viene dados por:

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(kx) dx, \quad n = 0, 1, 2, \dots, \quad (2.41)$$

$$b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(kx) dx, \quad n = 0, 1, 2, \dots, \quad (2.42)$$

Donde a_k y b_k son llamados los coeficientes de Fourier de la función $f(x)$.

Serie de Fourier Simple

Suponga un proceso Estocástico $x_1, x_2, \dots, x_n$ indexado en el tiempo y suponiendo que tiene un comportamiento periódico o cíclico. Entonces el proceso se puede modelar mediante un Sinusoide de la forma.

$$S_t = \mu + A \cos\left(\frac{2\pi t}{T} + \theta\right) \quad (2.43)$$

Donde T es el período o intervalo de tiempo que transcurre entre repetición de los valores de la variable, A es la amplitud y θ es la fase o desplazamiento en el origen respecto al valor 1.

La representación de la serie de Fourier de la ecuación 2.43 se puede expresar de una forma alternativa siguiendo una identidad trigonométrica:

$$\cos(r + t) = \cos(r)\cos(t) - \sin(r)\sin(t)$$

Entonces:

$$S_t = A\cos(wt + \theta) = A\cos(wt)\cos(\theta) - A\sin(wt)\sin(\theta) \quad (2.44)$$

$$S_t = \mu + a\cos(wt) + b\sin(wt) \quad (2.45)$$

Donde $w = \frac{2\pi}{T}$ se le conoce como la frecuencia armónica; $a = A\cos(\theta)$ y $b = -A\sin(\theta)$. La amplitud se puede obtener de la siguiente manera:

$$a^2 + b^2 = A^2\cos^2(\theta) + A^2\sin^2(\theta) = A^2 \underbrace{(\cos^2(\theta) + \sin^2(\theta))}_1 = A^2 \quad (2.46)$$

En consecuencia se puede obtener la amplitud de la serie periódica a través de $A = \sqrt{a^2 + b^2}$ y la fase (θ) se determina según la expresión (Bloomfield 2000) mencionado por (Gómez & Hernández 2013):

$$\theta = \begin{cases} \arctan(-b/a), & a > 0; \\ \arctan(-b/a) - \pi, & a > 0, b > 0; \\ \arctan(-b/a), & a > 0, b < 0; \\ -\pi/2, & a = 0, b > 0; \\ -\pi/2, & a = 0, b < 0; \\ \text{Arbitrario}, & a = 0, b = 0. \end{cases} \quad (2.47)$$

Kupper (1972) propone generalizar el uso de la serie para ajustar curvar armónicas, mediante una suma finita de la serie de Fourier de k componentes armónicos.

$$S_t = \mu + \sum_{i=1}^k \{a_i \cos(iwt) + b_i \sin(iwt)\} + e_t \quad (2.48)$$

Donde w se le denomina frecuencias armónicas fundamental y e_t corresponde al residual en el tiempo t , el cual se puede considerar como un ruido. Para este caso de la serie de Fourier, se deben de estimar $2k + 1$ parámetros; $\mu, a_1, a_2, \dots, a_k$ y $b_1, b_2, \dots, b_k$.

Estimación por Mínimos Cuadrados

Las estimaciones de los $2k + 1$ parámetros del modelo de la Serie de Fourier se puede realizar por medio de Mínimos Cuadrados y este consiste básicamente en minimizar una función $S(\mu, a_1, \dots, a_m, b_1, \dots, b_k)$ con respecto a los parámetros $\mu, a_1, \dots, a_m, b_1, \dots, b_k$.

$$\begin{aligned}
 S(\mu, a_1, \dots, a_m, b_1, \dots, b_m) &= \sum_{j=1}^n e^2 \\
 &= \sum_{j=1}^n [y_i - \mu - \sum_{i=1}^k a_i \cos(wit) + b_1 \sin(wit)]^2 \\
 \frac{\delta S}{\delta \mu} &= -2 \sum_{j=1}^n [y_i - \mu - \sum_{i=1}^k a_i \cos(wit) + b_1 \sin(wit)] = 0 \\
 &\vdots \\
 \frac{\delta S}{\delta a_k} &= -2 \sum_{j=1}^n [y_i - \mu - \sum_{i=1}^k a_i \cos(wit) + b_1 \sin(wit)] \cos(wit) = 0 \\
 &\vdots \\
 \frac{\delta S}{\delta b_k} &= -2 \sum_{j=1}^n [y_i - \mu - \sum_{i=1}^k a_i \cos(wit) + b_1 \sin(wit)] \sin(wit) = 0
 \end{aligned}$$

Esto corresponde a un sistema de $2m + 1$ ecuaciones igualadas a cero (0). Khuri (2003) presenta una expresión de los estimadores por Mínimos Cuadrados.

$$\begin{aligned}
 \hat{\mu} &= \frac{1}{n} \sum_{t=1}^n y_t \\
 \hat{a}_i &= \frac{2}{n} \sum_{t=1}^n y_t \cos(wit), \quad i = 1, 2, \dots, k \\
 \hat{b}_i &= \frac{2}{n} \sum_{t=1}^n y_t \sin(wit), \quad i = 1, 2, \dots, k
 \end{aligned}$$

Periodograma

El periodograma es una metodología que se emplea para encontrar “periodocidades ocultas” o frecuencias importantes en una serie que no son percibidas de forma inmediata. Esta metodología consiste en graficar las amplitudes de las ondas cosenosoidales en función de las frecuencias $w_i = iw$, donde $w = 1/T$, $i = 1, 2, 3, \dots, T$ se le conoce como la frecuencia principal (Carcedo 2000). La expresión del periodograma es presentada por (Khuri 2003):

$$I_n(w_i) = \frac{n}{2} (\hat{a}_i^2 + \hat{b}_i^2) \quad (2.49)$$

Prueba de Razón Generalizada de Verosimilitud

Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria de una población con función de densidad de probabilidad $f(x, \theta)$. Si $\theta \in \Theta_0$ donde Θ_0 es el correspondiente espacio paramétrico bajo la H_0 , entonces el estadístico $T(x) = -2\ln v(x)$ tiene distribución asintótica χ^2_{p-q} , donde q es el número de parámetros de θ que ha sido completamente especificado en la hipótesis nula y $v(x)$ tiene la siguiente expresión:

$$v(X) = \frac{\sup_{\theta \in \Theta} f_n(x, \theta)}{\sup_{\theta \in \Theta} f_n(x, \hat{\theta})} = \frac{\sup L_0}{\sup L} = \frac{f_n(x, \hat{\theta}_0)}{f_n(x, \hat{\theta})}$$

Y por lo tanto el estadístico $T(x)$ se expresaría de la siguiente manera:

$$T(x) = -2\ln \frac{f_n(x, \hat{\theta}_0)}{f_n(x, \hat{\theta})} \quad (2.50)$$

Un contraste de la hipótesis H_0 puede obtenerse comparando el estadístico $T(x)$ con el cuantil $\chi^2_{(p-q), \alpha}$; valores del estadístico mayores que dicho cuantil, conducirán al rechazo de la hipótesis nula (Manríquez & Valdez 2012).

2.2.9. Versión rankeada del test de razón de Von-Neumann

El principal supuesto para llevar a cabo el estudio es el de independencia entre las cantidades de precipitación de días sucesivos. Por tal motivo se presenta el test de razón de Von-Neumann rankeado, el cual evalúa la aleatoriedad (independencia) de una variable X .

Sea $R(i)$ el rango asociado con la i -ésima observación en una secuencia de N observaciones; la versión rankeada de la razón de Von-Neumann está definido como sigue (Cromwell, Labys & Terraza 1994):

$$RNV = \frac{\sum_{i=1}^{N-1} [R(i+1) - R(i)]^2}{\sum_{j=1}^N [R(j) - \mu_R]^2} \quad (2.51)$$

Donde μ_R es la media de los rangos de X . (Bartels 1982) proporciona una explicación relativamente sencilla, según la cual, la prueba de rachas debería ser menos potente que una prueba basada en rangos, debido a que el test de rachas ignora completamente las magnitudes de las observaciones. Los valores críticos para esta prueba están dados en (Bartels 1982).

A continuación se enuncian los pasos que se llevan a cabo para la realización de la prueba de Bartels:

Paso 1: Rankear los valores de X en forma ascendente, y denominar a la nueva variable $R(i)$.

Paso 2: Las hipótesis a probar son.

H_0 : las observaciones de X son independientes

H_1 : Las observaciones de X no son independientes

La prueba de estas hipótesis está basada en la estadística presentada en la ecuación 2.51:

Paso 3: Los valores críticos, τ , para comparar con RNV se encuentran en (Bartels 1982).

Paso 4. Se rechaza la hipótesis de independencia de X si

$$RNV > \tau$$

Capítulo 3

Metodología

La metodología propuesta para dar cumplimiento a los objetivos planteados en este proyecto, se desarrolló pretendiendo mantener una revisión bibliográfica continua, seguido de un análisis exploratorio de las observaciones, con el propósito de realizar una descripción del proceso de la precipitación (mm). Los datos fueron proporcionados por CENICAÑA, los cuales fueron registrados por la estación Melendez Cali y comprenden un periodo de estudio que va desde Septiembre 1993 hasta Agosto 2013, para un total de 20 años. Cabe resaltar que el estudio fue realizado en base a los días húmedos. Se consideró un día húmedo cuando la precipitación fue mayor a $0,1mm$ (la mínima cantidad de precipitación registrada) para filtrar en todo lo posible casos de precipitación horizontal y error de precisión del pluviómetro (Mormeneo & F. 2001).

$$U_j = Y_j - 0,1 \quad j = 1, 2, \dots, 365 \quad (3.1)$$

Donde Y_j denota la cantidad de precipitación en un día humedo ($X_j = 1$).

La cantidad de precipitación total en un mes se obtuvo mediante la suma de todas las cantidades diarias en ese mes ($Z_t = u_1 + u_2 + \dots$), y la colección de tales valores mensuales sobre todo el periodo de registro formaron la secuencia de precipitación mensual en dicha estación. Esta metodología es similar a la usada por Buishand (1977), Huynh et al. (1980), entre otros.

3.1. Análisis exploratorio de datos

Dentro del análisis exploratorio se tuvieron en cuenta estadísticas tales como la precipitación media, desviación estandar y asimetría mensual, número de días húmedos promedio

por mes, entre otras. Además, se examinó la influencia de precipitación en un mes sobre la precipitación en el siguiente mes, mediante el coeficiente de correlación serial (scc), definido como sigue (Buishand 1977):

$$r_k = \frac{\sum_{i=1}^{N-k} (v_i - \bar{v})(v_{i+k} - \bar{v})}{\sum_{i=1}^N (v_i - \bar{v})^2} \quad (3.2)$$

Donde $N = 12 * l$ (l , número de años) y v_i son las observaciones estandarizadas:

$$v_i = \frac{z_{lt} - \bar{z}_t}{s_t}$$

Con $l = 1, \dots, 20$, $t = 1, 2, \dots, 12$ y $\bar{z}_t$ y s_t , son la media y varianza muestral para el mes t . Se estandarizaron las secuencias de precipitaciones mensuales debido a la variación estacional en la media y desviación estándar para cada mes.

Puesto que las observaciones no son normales, se aplicó el test de razón de Von-Neumann rankeado, propuesto por Bartels (1982) definido en la ecuación 2.51 el cual es una prueba que no está basada en ninguna distribución de probabilidad, a diferencia de pruebas como Durbin-Watson y Ljung-Box que requieren normalidad.

3.2. Ajuste de las distribuciones propuestas

Para cada uno de los periodos mensuales se realizaron las estimaciones de máxima verosimilitud de los parámetros para las distribuciones loi des fuites (LDF), Gamma, Lognormal y Weibull, tal como se explicó en la sección 2.

3.3. Pruebas de bondad de ajuste

Con el fin de determinar que tan bien ajustaron los modelos propuestos a las observaciones totales mensuales, se llevaron a cabo pruebas de bondad de ajuste. Las pruebas de bondad de ajuste pueden ser probadas tanto cualitativamente como cuantitativamente. La prueba cuantitativa más usada es la prueba de Kolmogorov-Smirnov, la cual compara el estadístico D_n (la mayor diferencia, en valor absoluto, entre la distribución empírica y la distribución teórica) con un valor critico determinado por el número de observaciones en el conjunto de datos y un nivel de significancia seleccionado (Wilks 1998).

Además de la prueba K-S, se usó el criterio BIC, el cual penaliza la función de verosimilitud maximizada por el número de parámetros usados en la estimación. El mejor modelo

será aquel que obtenga un BIC más pequeño. Por otra parte, las pruebas cualitativas a menudo usadas involucran simples histogramas con la curva de distribución ajustada sobrepuesta, Q-Q plots, entre otros. También se compararon las medias y desviaciones estándar de cada uno de los modelos con respecto a la media y desviaciones muestrales para cada mes.

3.4. Ajuste de las series de Fourier

Luego de haber seleccionado la mejor distribución para cada uno de los meses, se procedió a describir la variación estacional de la media y desviación estandar mensual, usando las estimaciones dadas por los modelos para dichas medidas de tendencia central y de forma. Esta metodología es similar a la trabajada por Coe & Stern (1982) y Stern & Coe (1984), donde los autores describieron la variación presentada por la media de la distribución Gamma mediante una serie de Fourier con un enfoque de modelo lineal generalizado.

La variación de la media y desviación estandar mensual es descrita mediante la serie finita de Fourier:

$$\theta_j(t) = A_{j0} + \sum_{k=1}^{m_j} \left\{ A_{jk} \cos\left(\frac{2\pi tk}{12}\right) + B_{jk} \sin\left(\frac{2\pi tk}{12}\right) \right\} + \epsilon_t \quad (3.3)$$

Donde $j = 1, 2$ indica los datos a usar de la media y desviación estandar mensual respectivamente, m_j es el número máximo de armónicos requeridos para especificar dichos parámetros, A_{j0} , A_{jk} y B_{jk} son los coeficientes de la serie de Fourier y ϵ_t son los errores aleatorios los cuales son independientes y distribuidos normal con media cero y varianza σ_t^2 . Los parámetros de la serie son estimados por mínimos cuadrados (Buishand 1977).

Para determinar el número de armónicos a utilizar se empleó el periodograma y la prueba de razón de verosimilitud. El periodograma es una metodología que ayuda a encontrar “periodicidades ocultas” o frecuencias importantes en una serie que no son percibidas de forma inmediata. Esta metodología consiste en graficar las amplitudes de las ondas cosenoidales en función de las frecuencias $w_i = iw$, donde w se le conoce como la frecuencia principal $w = 1/T$, $i = 1, 2, \dots, T$ (Carcedo 2000).

$$I_n(w_i) = \frac{n}{2} \left(\hat{a}_i^2 + \hat{b}_i^2 \right) \quad (3.4)$$

A la expresión 3.4 se le conoce como periodograma o espectro de amplitud que identifica los armónicos para la serie, pero para determinar si es significativo involucrar dicha frecuencia

o armónico del modelo, se plantea una metodología de selección de variables hacia adelante, basándose en la prueba de razón de verosimilitud, iniciando con un modelo nulo e ir adicionando armónicos conforme el periodograma lo sugiera, hasta llegar al punto en el que las frecuencias no sean significativas.

Una vez realizado el ajuste de la serie finita de Fourier para cada parámetro, se verifico el cumplimiento de los supuestos:

- $e_i \text{ Normal}(0, \sigma^2)$
- $E[e_i] = 0$
- $Var[e_i] = \sigma^2$ constante
- $Corr[e_i, e_j] = 0$

La validación de estos supuestos se puede realizar de manera gráfica o analítica. A continuación se enuncian las diferentes pruebas que existen, especificando el supuesto en el cual son de utilidad.

Pruebas Gráficas

- **Histograma:** Este tipo de gráfico ilustra una primera aproximación a la distribución de los datos de interés. En el caso de los residuales, si la forma de este gráfico tiende a ser simétrica alrededor del cero, se puede decir en primera instancia que la distribución es a simple vista normal con media cero.
- **Gráfico de Residuales:** La gráfica de residuales en secuencia temporal puede indicar la autocorrelación de los errores en un periodo con los de otros periodos. Una gráfica como la de la Figura 3.1 (a) indica una autocorrelación positiva, es decir que existe dicha correlación, mientras que la Figura 3.1 (b) da indicios de una autocorrelación negativa, o sea que no existe dicha correlación (Montgomery & Peck 2005).
- **Gráfico de Residuales vs Valores ajustados:** Esta gráfica es útil para detectar algunas inadecuaciones frecuentes del modelo. Si esta gráfica se parece a la de Figura 3.2 (a), esto indica que los residuales se pueden encerrar en una banda horizontal y por tanto no hay defectos obvios del modelo. Mientras que las Figuras (b) o (c) indican deficiencias del modelo. Por ultimo, una gráfica de e_i en función de $\hat{y}_i$ que sea similar a la Figura (d), es indicio de no linealidad, y por tanto es posible pensar que se necesiten otras variables regresoras en el modelo. Por ejemplo, un término cuadrático (Montgomery & Peck 2005).

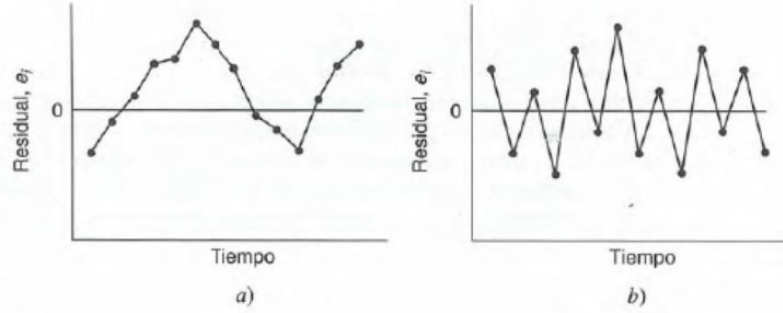


Figura 3.1: Ejemplo Gráfica e_i vs tiempo. Tomado de Montgomery & Peck (2005)

Pruebas Análíticas

- **Prueba de Anderson-Darling:** Este contraste tiene como objetivo probar si un grupo de datos está distribuido normalmente para lo cual es más potente que otras en casos en los que hay presencia de valores atípicos o outliers. El estadístico a probar en este caso viene dado por la siguiente expresión:

$$A^2 = -n - \frac{1}{n} \sum_i [(2i - 1)\ln(P_i) + (2n + 1 - 2i)\ln(1 - P_i)]$$

Donde (P_i) es el área bajo la curva normal para el intervalo $(-\infty, z_i)$, es decir que es la función de distribución normal evaluada en el i -ésimo elemento (orden ascendente) de la muestra. Para el caso de bondad de ajuste a la distribución normal, se presentan los siguientes casos:

- μ y σ^2 conocidos.
- σ^2 conocida y μ desconocida y estimada por $\hat{X}$.
- μ conocida y σ^2 desconocida y estimada por s^2 .
- Ambos parámetros desconocidos y estimados por $\hat{X}$ y s^2 .

Para cada uno de los casos, existe una tabla estadística para realizar la prueba de hipótesis H_0 : La muestra aleatoria proviene de una distribución normal. Sin embargo, el estadístico se calcula de igual manera en todos los casos para lo cual es necesario organizar la muestra ascendentemente (Marquez Dos Santos 2001).

- **Contraste Durbin-Watson:** Esta prueba se basa en la hipótesis de que los errores de un modelo se generan en un proceso autorregresivo de orden uno, que se observa a intervalos de tiempo constantes y dado que la mayor parte de los contextos de regresión donde intervienen las series temporales tienen autocorrelación positiva, la hipótesis que considera esta prueba es:

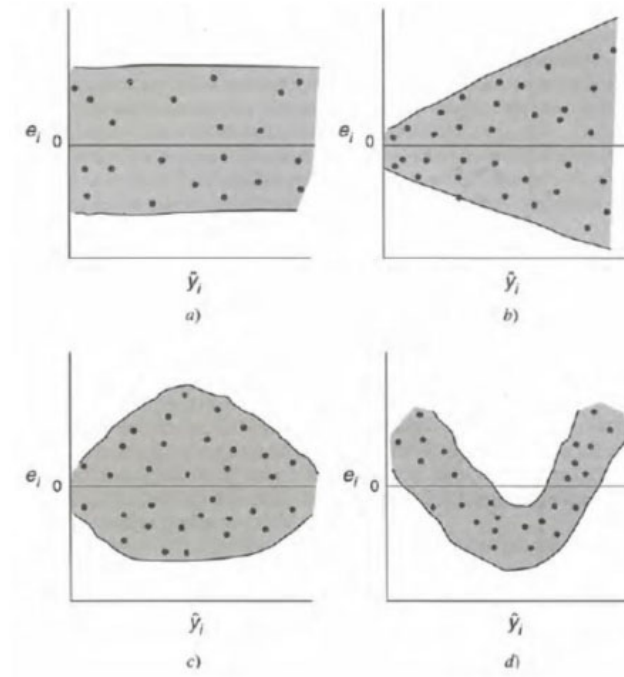


Figura 3.2: Ejemplo Gráfica e_i vs $\hat{y}_i$. Tomado de Montgomery & Peck (2005)

$$H_0 : \rho = 0 \quad \text{vs} \quad H_a : \rho \neq 0$$

La cual se resuelve calculando el siguiente estadístico:

$$d = \frac{\sum_{t=1}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n (e_t)^2}$$

El cual tiene una distribución tabulada por los mismos autores. Para que la aplicación de este contraste tenga sentido, es necesario que los datos esten ordenados; cuando se trata de una serie de tiempo, el orden más inmediato surge como la secuencia cronológica que las observaciones tengan (Murillo & González 2004).

- **Contraste Breunsh-Pagan:** Esta fue planteada por Breunsh y Pagan en 1979, el uso de esta es estrictamente válido cuando se dispone de muestras suficientemente grandes. Supone que es posible expresar la varianza de los errores como una combinación lineal de p variables explicativas, es decir:

$$V[\varepsilon] = \alpha_0 + \alpha_1 Z_1 + \dots + \alpha_p Z_p$$

El estadístico de esta prueba es:

$$BP = \frac{SCR}{2}$$

Donde SCR corresponde a la suma de cuadrados del modelo cuya variable dependiente es $e_i^2 / \sum e_i^2 / n$ donde los e son los residuales del modelo cuya homocedasticidad se quiere probar. Este estadístico bajo la hipótesis nula sigue una distribución χ^2 con p grados de libertad, es decir que si el estadístico de prueba es mayor que el valor de χ^2 tabulado o crítico, entonces se rechaza la buena especificación del modelo y deben revisarse las etapas de su construcción (Behar 2003).

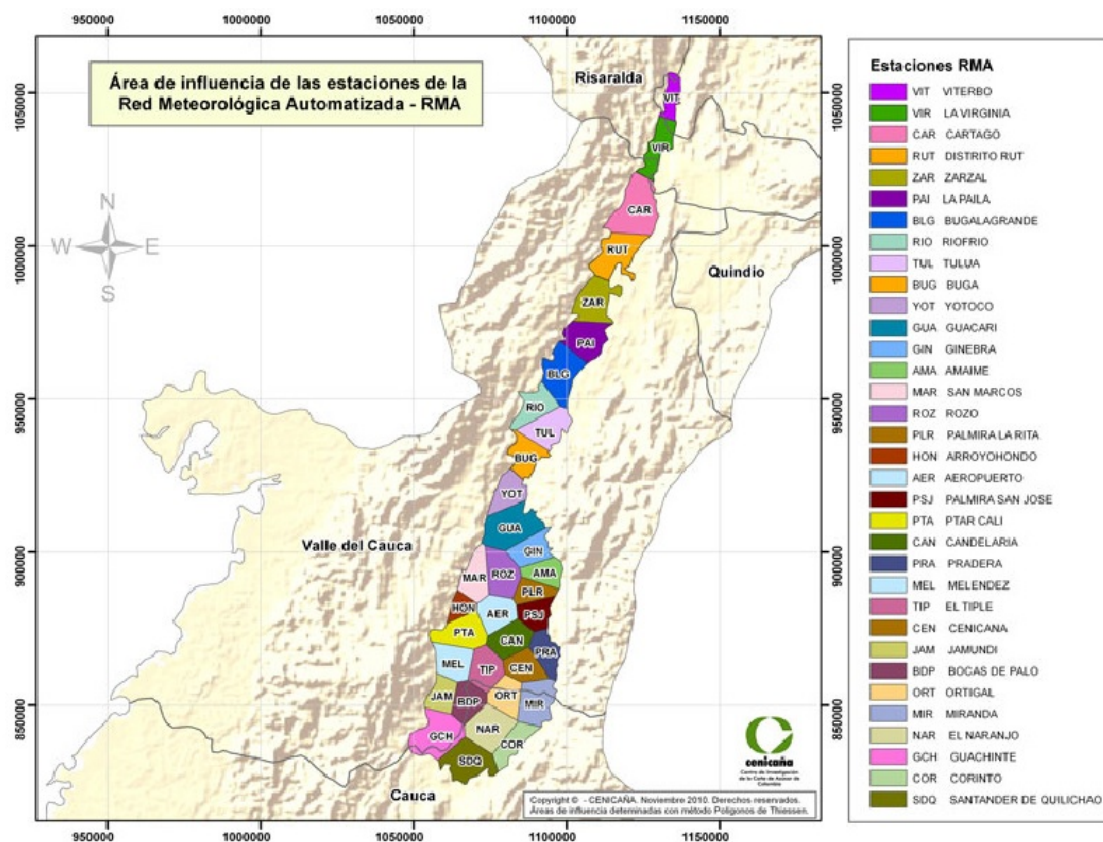
Por último, cabe aclarar que todas las pruebas realizadas para probar cada uno de los supuestos involucrados en esta investigación fueron hechas bajo un 5 % de significancia.

Capítulo 4

Resultados

La base de datos seleccionada corresponde a los registros para la estación Meteorológica Automatizada de Meléndez, la cual se encuentra ubicada en el municipio de Santiago de Cali capital del departamento del Valle del Cauca. Este municipio se encuentra a una altura de 995 msnm y cuenta aproximadamente con una extensión de 564 km².

Figura 4.1: Estaciones meteorológicas CENICAÑA



La estación perteneciente a la red meteorológica de la organización CENICAÑA, cuenta registros diarios de precipitación medidos en escala continua desde el 01 de septiembre de 1993 al 14 de octubre de 2013. Puesto que la metodología planteada requiere de un periodo de observación de por lo menos 20 años, se decide tomar los datos comprendidos entre el 01 de septiembre de 1993 al 31 de agosto de 2013. Dicho periodo presenta ausencia de 438 datos que corresponde al 0.66 % del total de los registros, por lo cual no existen problemas para realizar el ajuste de los modelos propuestos.

4.1. Análisis Exploratorio de datos

Resulta importante indicar que se utilizó la secuencia cronológica en que las observaciones fueron registradas, debido a que en la literatura el análisis de la información es realizada de dicha manera y no en orden cronológico (Enero a Diciembre).

En la Tabla 4.1 se observa que las máximas medias de precipitación mensual recaen en los meses de Marzo, Abril, Mayo, Octubre y Noviembre; mientras que las mínimas se obtuvieron en los meses de Julio y Agosto. Este comportamiento del régimen de lluvias para la zona de estudio se ve claramente en la Figura 4.2, en la cual se detalla dos temporadas de lluvias (una más intensa que la otra) y dos temporadas con menos presencia de precipitación.

Tabla 4.1: Estadísticas Descriptivas de la cantidad total de lluvia mensual. Periodo multianual 1993-2013

Mes	Prec. Mín.	Prec. Máx	Media	Desv. Est.	CV	C. Asimetría	Media días lluviosos
Sep	15.2	170.2	81.4	48.06	0.59	0.33	10.85
Oct	35.2	202.8	125.5	48.26	0.38	-0.33	16.6
Nov	62.4	269.4	174.9	69.46	0.39	0.02	17.15
Dic	3.7	252.7	113.2	63.28	0.56	0.41	13.55
Ene	6.0	230.0	100.4	66.69	0.67	0.26	11.85
Feb	7.2	231.1	89.1	66.5	0.75	0.64	11.00
Mar	33.4	258.0	149.6	56.56	0.38	0.32	14.75
Abr	52.6	334.5	195.8	64.74	0.33	-0.03	18.1
May	89.4	313.4	165.7	61.00	0.37	0.87	15.75
Jun	20.1	196.9	85.7	48.83	0.57	0.48	11.6
Jul	6.5	173.5	50.5	44.04	0.87	1.20	9.15
Ago	1.6	132.9	54.2	37.74	0.69	0.61	8.65

De igual forma se puede ver que las máximas desviaciones estándar mensuales están dadas por los meses de Noviembre, Enero y Febrero, mientras que la menor desviación registrada corresponde al mes de Agosto. Sin embargo, se observa que en general la desviación

estándar para cada uno de los meses tiende a mantenerse constante. Este hecho se refleja en la Figura 4.2, en donde se detalla que las “cajas” para los meses tienen más o menos el mismo ancho y se corrobora con el coeficiente de variación mensual, el cual oscila alrededor de un valor medio de 0.546. Por otra parte, se tiene que para la mayoría de los meses la asimetría se considera positiva o la derecha, excepto para los meses de Octubre y Abril donde registraron una asimetría negativa o a la izquierda, posiblemente debido a que en estos meses es más frecuente la presencia de una mayor cantidad de lluvia.

Figura 4.2: Cantidad de precipitación total mensual

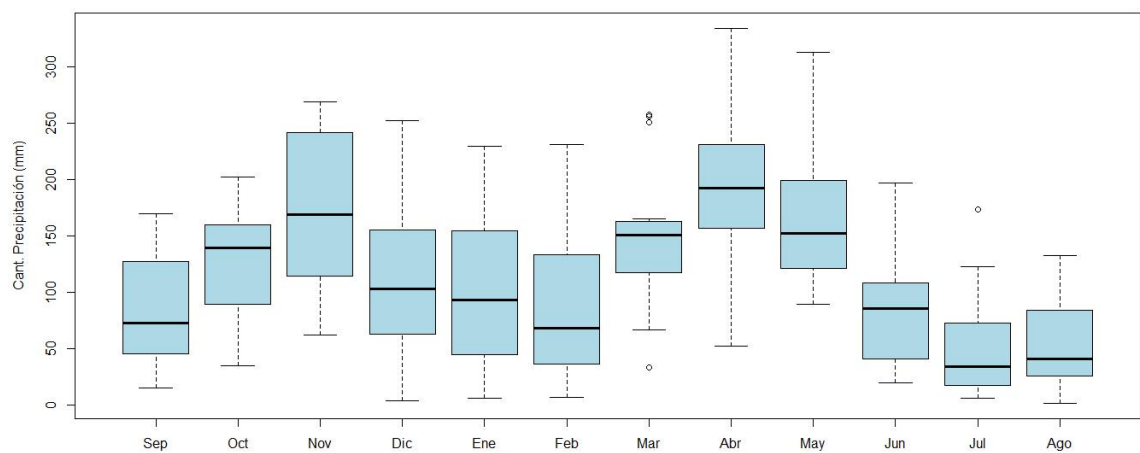
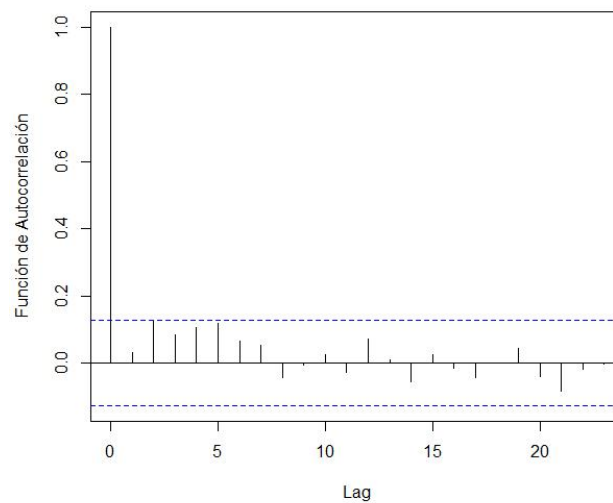


Figura 4.3: Función de Autocorrelación para la secuencia de precipitación mensual



Como ya se ha mencionado, es usual suponer que las cantidades de lluvia son mutuamente

independientes y así, ajustar distribuciones con la característica de asimetría positiva a dichas cantidades. Debido a esto, se realizó la verificación de dicho supuesto mediante la función de autocorrelación de las observaciones (fac) así como la prueba de Von-Neumann Rankeada (RVN) también conocida como el test de Bartels.

Tabla 4.2: Prueba de Independencia Von-Neumann Rankeada

Prueba de Bartels	
RVN	Valor p
1.926	0.5668

El resultado de dicha prueba se muestra en la Tabla 4.2 en la cual se observa un valor-p igual a 0.5668 que es mayor al nivel de significancia 5 %, y por ende, se concluye que no existe influencia de la precipitación en un mes sobre la lluvia en el siguiente mes. Esta afirmación es corroborada con la función de autocorrelación (Figura 4.3) donde se detalla un solo rezago significativo, que es el comportamiento característico de un ruido blanco.

4.2. Ajuste de las distribuciones propuestas

Debido al cumplimiento de la independencia entre las cantidades de precipitaciones mensuales, se procedió a realizar las estimaciones de los parámetros para las distribuciones propuestas. En la Tabla 4.3 se presentan las estimaciones de máxima verosimilitud para cada uno de los parámetros de dichos modelos.

Tabla 4.3: Estimación de parámetros por máxima verosimilitud

Mes	LDF		Gamma		Lognormal		Weibull	
	θ	ρ	k	λ	μ	σ	α	β
Sep	5.034	0.062	2.546	0.031	4.191	0.701	1.803	91.856
Oct	9.536	0.076	5.461	0.044	4.738	0.473	3.037	140.791
Nov	9.918	0.057	5.993	0.034	5.079	0.431	2.896	197.218
Dic	5.847	0.052	2.235	0.020	4.489	0.879	1.801	126.286
Ene	3.638	0.036	1.608	0.016	4.267	0.982	1.452	110.288
Feb	3.855	0.043	1.641	0.018	4.155	0.905	1.371	97.749
Mar	8.144	0.054	6.131	0.041	4.924	0.448	2.894	167.590
Abr	11.331	0.058	7.813	0.040	5.212	0.393	3.416	217.816
May	9.305	0.056	8.545	0.052	5.047	0.341	2.922	185.355
Jun	5.438	0.063	2.874	0.034	4.267	0.646	1.893	96.968
Jul	3.296	0.065	1.445	0.029	5.537	0.931	1.220	54.165
Ago	3.877	0.071	1.668	0.031	3.665	0.984	1.436	59.551

De igual forma, se presentan en la Tabla 4.4 las medias y desviaciones estándar mensuales

estimadas mediante cada una de las distribuciones en cuestión. En general se tiene que:

- La distribución loi des fuites y la Gamma presentan buenas estimaciones para la media pero tienden a arrojar valores más grandes de la desviación estándar mensual en comparación a las obtenidas por la distribución empírica de las observaciones.
- La distribución Lognormal presenta valores superiores tanto en media como en desviación estándar para cada uno de los meses, excepto en el mes de Mayo, donde produjo estimaciones muy similares a la observada en los datos.
- La distribución Weibull fue la que tuvo mejores estimaciones de la media y desviación estándar mensual, debido a que la distribución empírica de los datos muestra resultados similares con relación a los obtenidos por esta distribución teórica.

Tabla 4.4: Media ($\hat{\mu}$) y desviación estandar ($\hat{\sigma}$) estimadas

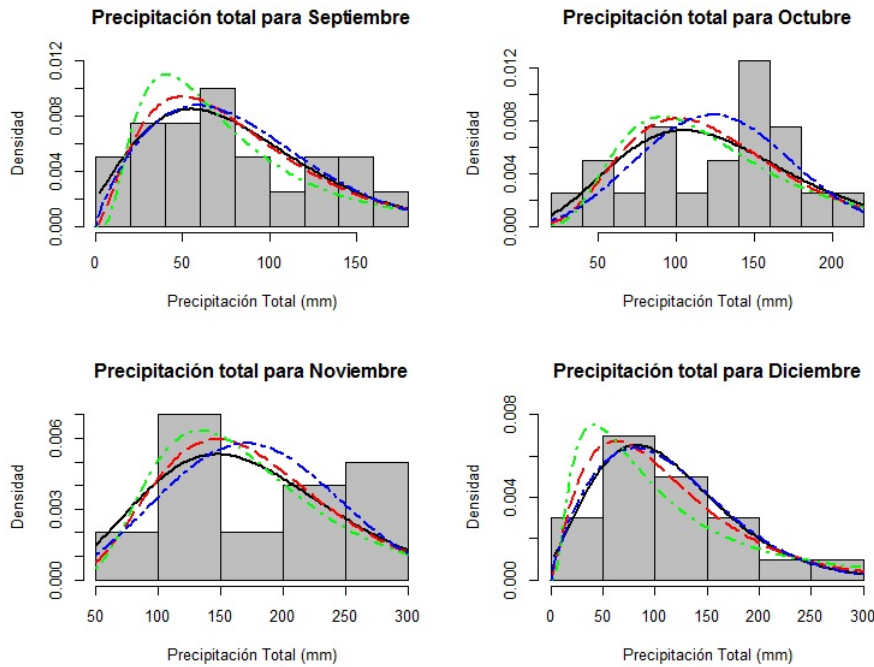
Mes	LDF		Gamma		Lognormal		Weibull	
	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$
Sep	81.194	51.178	82.129	51.472	84.496	67.311	81.680	46.885
Oct	125.474	57.463	124.114	53.111	127.723	63.955	125.792	45.219
Nov	174.00	78.136	176.265	72.002	176.246	79.630	175.847	65.964
Dic	112.442	65.762	111.750	74.750	131.016	141.441	112.301	64.527
Ene	101.056	74.928	100.500	79.254	115.487	147.128	99.981	69.954
Feb	89.651	64.574	91.167	71.168	96.016	108.131	89.395	65.974
Mar	150.815	74.738	149.537	60.392	152.072	71.694	149.426	56.088
Abr	195.362	82.077	195.325	69.879	198.190	80.955	195.730	63.309
May	166.161	77.034	164.327	56.215	164.867	57.894	165.331	61.525
Jun	86.317	52.347	84.529	49.861	87.852	63.222	86.056	47.269
Jul	50.708	39.500	49.828	41.451	53.005	62.248	50.739	41.799
Ago	54.606	39.220	53.806	41.662	63.379	81.000	54.070	38.221

En las Figuras 4.4, 4.5 y 4.6 se detalla el ajuste de las distribuciones en cuestión solapadas en el histograma para las cantidades totales en cada uno de los meses. Note que en la mayoría de los meses el modelo Weibull y Gamma son los que aparentemente mejor ajustan a los datos, mientras que la distribución loi des fuites tiene un pobre desempeño en casi todos los meses excepto para Diciembre y Enero. Finalmente, se corrobora la sobrestimación mostrada por el modelo Lognormal.

4.3. Pruebas de bondad de ajuste

A continuación se muestra el análisis de bondad de ajuste realizado, con el fin de seleccionar la distribución que mejor se ajuste a los datos de cantidad de precipitación total mensual.

Figura 4.4: Histograma precipitación Sep, Oct, Nov y Dic vs ajuste modelos. Línea negra: Loi des fuites; Línea roja: Distribución Gamma; Línea verde: Distribución Lognormal; Línea azul: Distribución Weibull



En la Tabla 4.5 se presentan el valor-p para la prueba de Kolmogorv-Smirnov y el Criterio de Información Bayesiano (BIC). Recuerde que el criterio para seleccionar el modelo que mejor desempeño tenga, está basado en el modelo con el mayor valor-p de la prueba K-S y un menor valor del BIC.

De forma General, se tiene que:

- Para los meses de Octubre, Enero, Abril y Junio la distribución seleccionada corresponde a la Weibull.
- Para los Meses de Diciembre, Mayo y Julio las distribuciones seleccionadas corresponden a la loi des fuites, Lognormal y Gamma respectivamente.
- Para el mes de Septiembre hay un empate entre la distribución Lognormal y la Weibull, mientras que para los meses Noviembre, Febrero, Marzo y Agosto los modelos empatados corresponden a la Gamma y Weibull. Estos empates se presentaron debido a que, por ejemplo, para el mes de Septiembre la distribución Lognormal tuvo un valor-p de la prueba K-S igual a 0.958 y un valor BIC de 216.15, mientras que para la Weibull el valor-p fue de 0.9016 y el BIC de 213.7. Como se puede notar, no existe una clara evidencia de que modelo podría ser el más adecuado. Este mismo hecho ocurrió en el resto de meses empatados.

Figura 4.5: Histograma precipitación Ene, Feb, Mar y Abr vs ajuste modelos. Línea negra: Loi des fuites; Línea roja: Distribución Gamma; Línea verde: Distribución Lognormal; Línea azul: Distribución Weibull

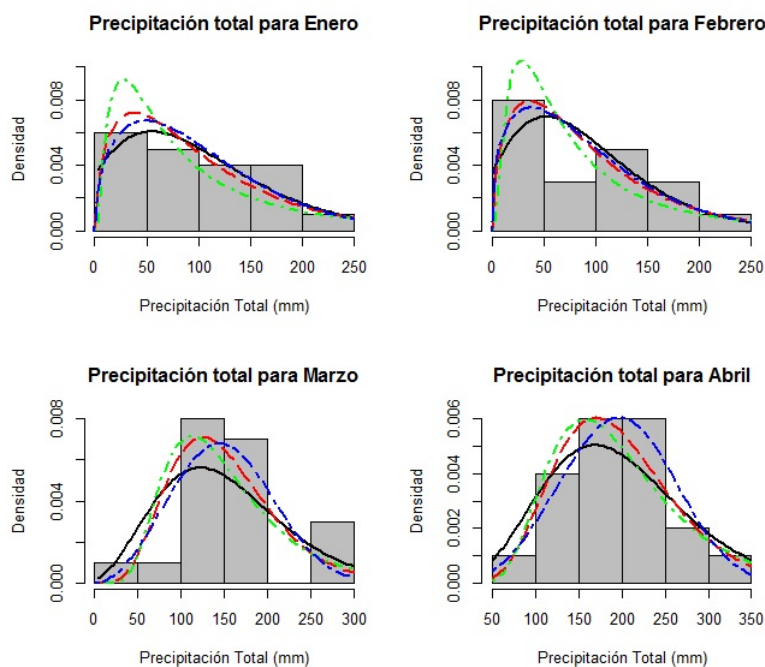


Figura 4.6: Histograma precipitación May, Jun, Jul y Ago vs ajuste modelos. Línea negra: Loi des fuites; Línea roja: Distribución Gamma; Línea verde: Distribución Lognormal; Línea azul: Distribución Weibull

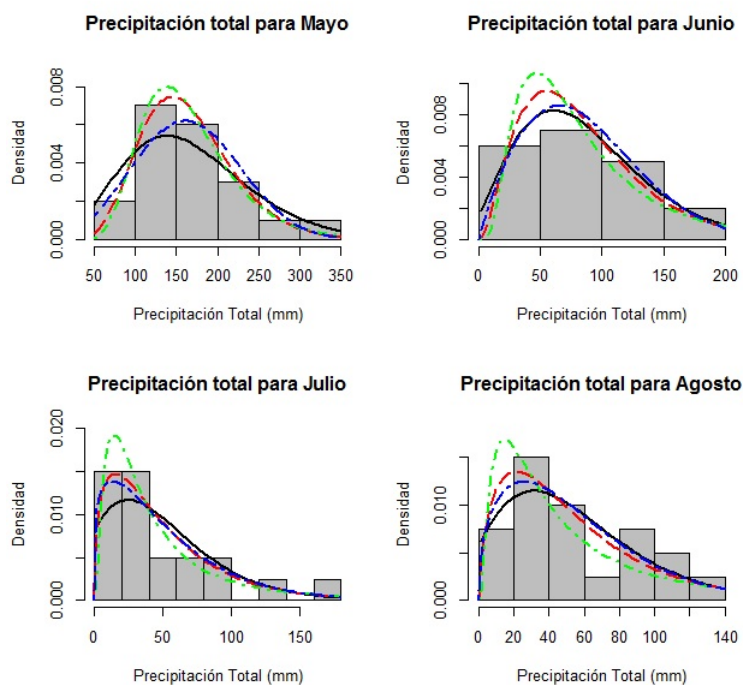
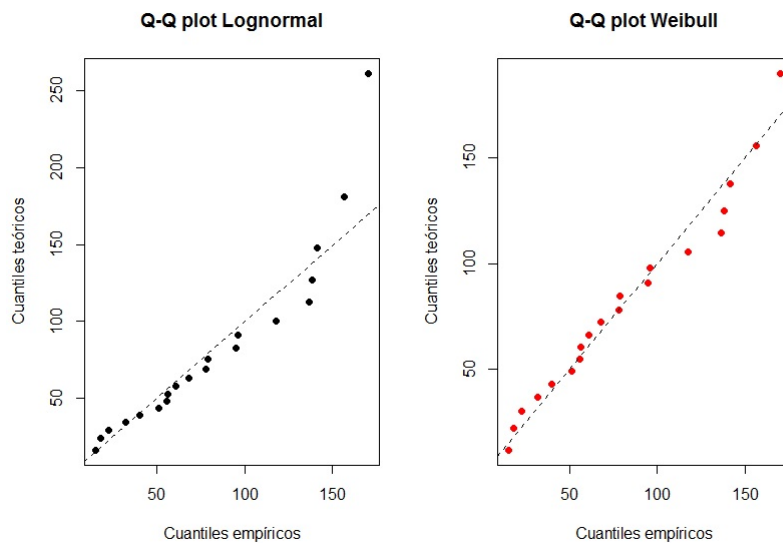


Tabla 4.5: Valor p de la prueba de Kolmogorov-Smirnov y Criterio de Información Bayesiano (BIC)

Mes	Loi des fuites		Gamma		Lognormal		Weibull	
	Valor p	BIC	Valor p	BIC	Valor p	BIC	Valor p	BIC
Sep	0.000322	215.32	0.6128	214.28	0.958	216.15	0.9016	213.7
Oct	0.0003914	219.71	0.3448	219.54	0.3312	222.31	0.6488	216.41
Nov	0.4337	232.16	0.7163	231.19	0.6427	232.18	0.6801	230.32
Dic	0.9858	226.59	0.8609	229.16	0.4616	237.16	0.9822	227.06
Ene	0.1753	229.5	0.6527	227.97	0.3211	232.68	0.8647	226.93
Feb	0.8475	224.51	0.9607	223.03	0.7516	224.96	0.9384	222.89
Mar	0.005961	227.04	0.4137	224.54	0.4489	227.61	0.238	223.18
Abr	0.4841	232.08	0.8799	230.97	0.7313	233.88	0.9915	228.67
May	0.05136	227.98	0.5848	222.51	0.7183	221.55	0.3661	225.45
Jun	0.2278	215.83	0.7974	214.63	0.6073	215.97	0.8111	214.32
Jul	3,59E-07	204.73	0.961	201.35	4,48E-12	201.36	0.9118	201.63
Ago	0.3657	203.63	0.9759	203.02	0.7492	208.7	0.9086	202.3

Figura 4.7: Q-Q plot Lognormal vs Weibull para el mes de Septiembre



Sin embargo, la cuestión de los meses empatados puede solucionarse mediante el Q-Q plot para las distribuciones involucradas (Figuras 4.7, 4.8, 4.9, 4.10 y 4.11) y mediante el análisis de las estimaciones de las medias y desviaciones estándar mensuales producidas por dichos modelos.

En el Q-Q plot de la distribución Lognormal vs Weibull para el mes de septiembre (Figura 4.7) se observa un mejor ajuste por parte del modelo Weibull, específicamente en la cola superior; además la media y desviación estándar estimada por parte de dicho modelo son más similares que las producidas por el modelo Lognormal. Por tanto, la distribución que más se adecua a los datos de precipitación total para Septiembre es la Weibull.

Por otra parte, para el resto de meses, se observa un comportamiento muy similar entre las distribuciones Gamma y Weibull. Sin embargo, se puede detallar un mejor ajuste en la cola superior por parte del modelo Weibull en comparación con la Gamma. Este hecho puede deberse a las características deseables que tiene la distribución Weibull, tales como, invertibilidad, integrabilidad y estimaciones cercanas en todos los momentos probabilísticos. Además como se mencionó anteriormente, la Weibull presentó estimaciones de la media y desviación estándar mensual más cercanas a las empíricas.

Finalmente, la cantidad de precipitación total para cada uno de los meses del año es modelada por las siguientes distribuciones, los cuales fueron seleccionados de acuerdo a los criterios de bondad de ajuste:

Tabla 4.6: Modelos seleccionados para cada uno de los meses

Mes	Distribución	Parámetros	
Sep	Weibull	$\hat{\alpha} = 1,80$	$\hat{\beta} = 91,86$
Oct	Weibull	$\hat{\alpha} = 3,04$	$\hat{\beta} = 140,79$
Nov	Weibull	$\hat{\alpha} = 2,89$	$\hat{\beta} = 197,22$
Dic	LDF	$\hat{\theta} = 5,85$	$\hat{\rho} = 0,05$
Ene	Weibull	$\hat{\alpha} = 1,45$	$\hat{\beta} = 110,29$
Feb	Weibull	$\hat{\alpha} = 1,37$	$\hat{\beta} = 97,75$
Mar	Weibull	$\hat{\alpha} = 2,89$	$\hat{\beta} = 167,59$
Abr	Weibull	$\hat{\alpha} = 3,42$	$\hat{\beta} = 217,82$
May	Lognormal	$\hat{\mu} = 5,05$	$\hat{\sigma} = 0,34$
Jun	Weibull	$\hat{\alpha} = 1,89$	$\hat{\beta} = 96,97$
Jul	Gamma	$\hat{k} = 1,45$	$\hat{\lambda} = 0,03$
Ago	Weibull	$\hat{\alpha} = 1,44$	$\hat{\beta} = 59,55$

Figura 4.8: Q-Q plot Gamma vs Weibull para el mes de Noviembre

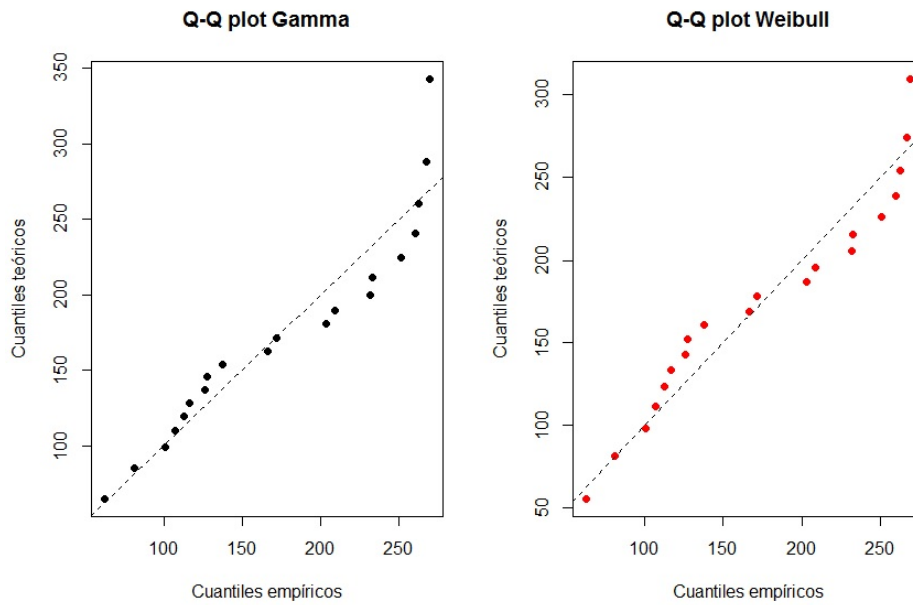


Figura 4.9: Q-Q plot Gamma vs Weibull para el mes de Febrero

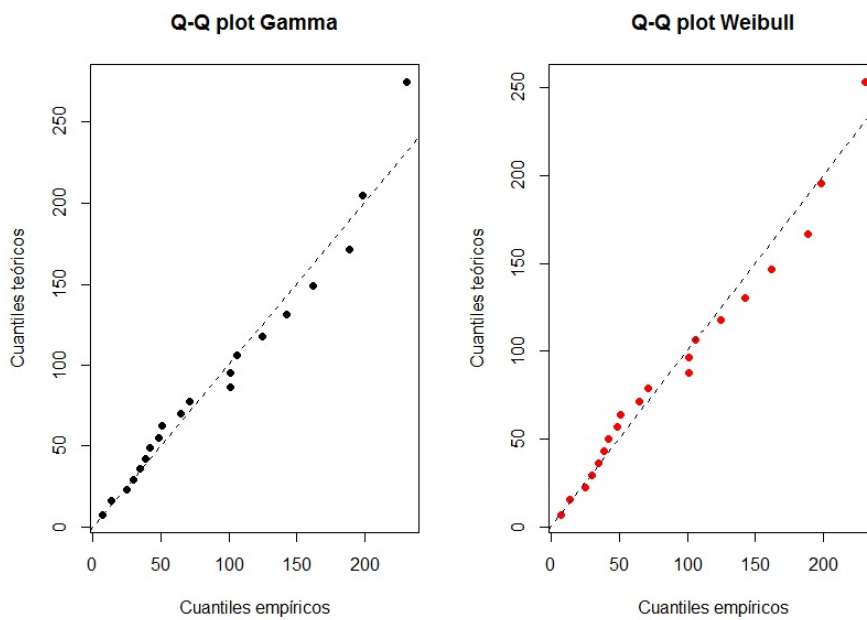


Figura 4.10: Q-Q plot Gamma vs Weibull para el mes de Marzo

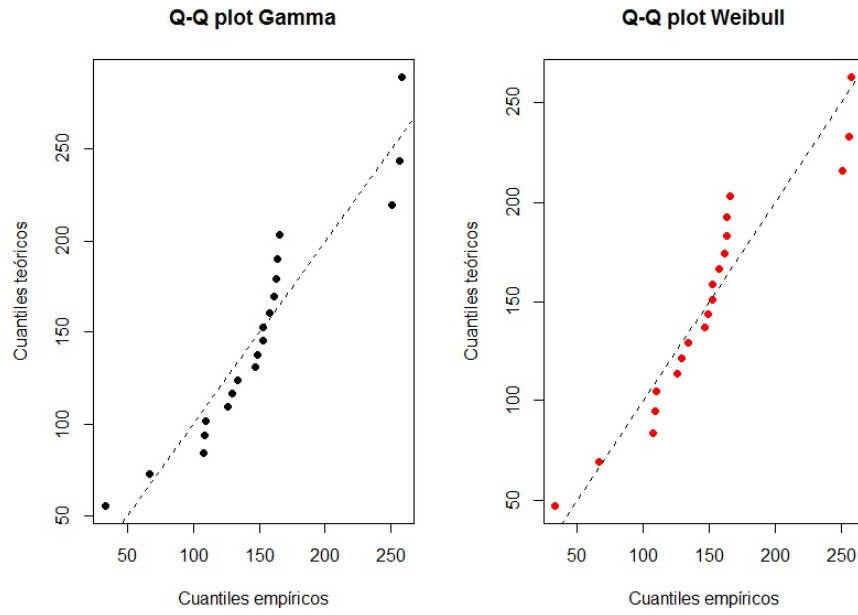
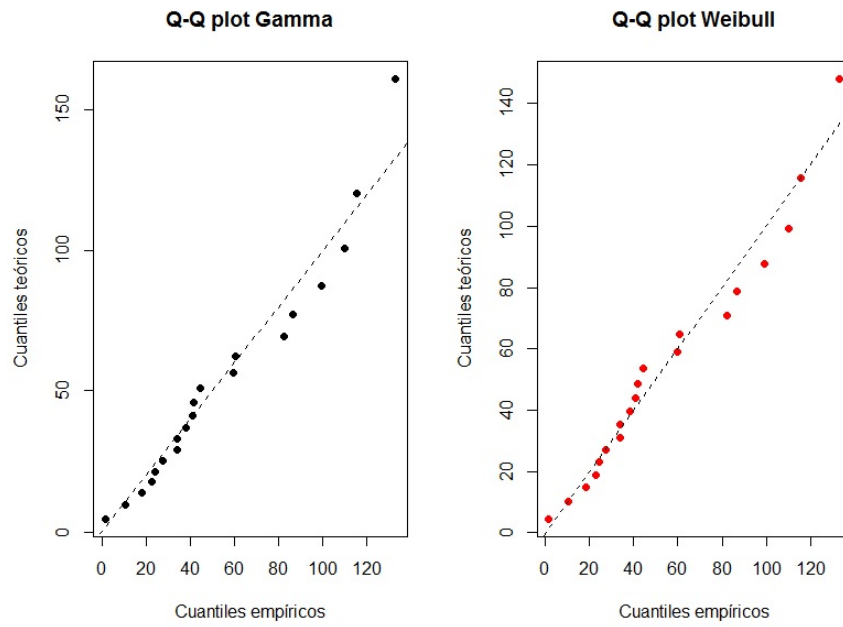


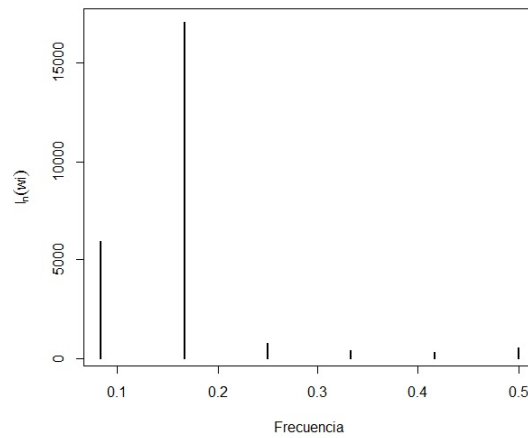
Figura 4.11: Q-Q plot Gamma vs Weibull para el mes de Agosto



4.4. Ajuste de las series de Fourier

Una vez seleccionado los modelos para cada uno de los meses, se procede a la descripción de la variación estacional de la media y desviación estándar mensual estimada. Para la selección del número de armónicos se presentan las Figuras 4.12 y 4.13, las cuales son los periodogramas para la media y desviación estándar mensual estimada, respectivamente.

Figura 4.12: Periodograma para la media mensual estimada



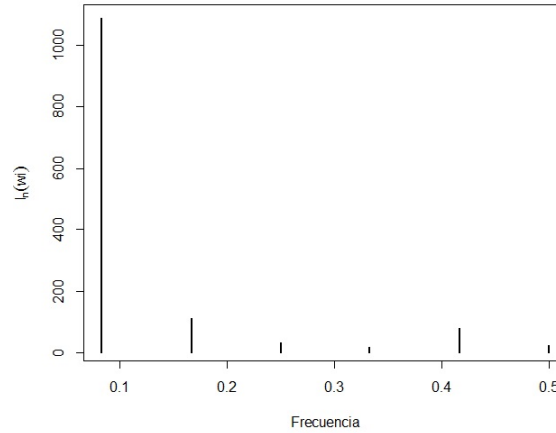
Para la media mensual estimada se tiene un pico significativo correspondiente a la segunda frecuencia, lo cual es consecuencia del comportamiento bimodal de la precipitación en el año, tal como se observó en la Figura 4.2. Por otra parte, para la desviación estándar mensual estimada, se tiene que la primera frecuencia es la más significativa, lo que es consistente con la poca variación observada en la varianza mensual (ver Figura 4.2).

Con el fin de determinar el número de armónicos a usar para cada modelo de una manera más analítica, se empleó el test de razón de verosimilitud, el cual es frecuentemente usado para tal fin. Esta prueba se contrasta con un valor crítico dado por las tablas de la distribución χ^2 con dos grados de libertad, correspondientes al número de parámetros que se estiman al incluirse un nuevo armónico.

Tabla 4.7: Prueba de razón de verosimilitud para el No. de Armónicos

Modelos	No. Armónicos	Estadístico	gl	$\chi_{0,95;2}$
Media mensual	1	3.2924	2	5.99
	2	27.05	2	5.99
	3	8.0465	2	5.99
Desv. Est. Mensual	1	17.542	2	5.99
	2	9.0894	2	5.99

Figura 4.13: Periodograma para la desviación estandar mensual estimada



En la Tabla 4.7 se observa que para el modelo para la media mensual, la segunda armónica es la más significativa, mientras que para el modelo de desviación estándar mensual la primera armónica es la más significativa. Este resultado es concorde a lo observado en los periodogramas mostrados anteriormente. Sin embargo, se tendrán presente los modelos con tres y dos armónicos para la media y desviación estándar respectivamente, puesto que la prueba de razón de verosimilitud muestra que también son significativos.

Luego de seleccionar el número de armónicos para cada uno de los modelos, se procedió a realizar las estimaciones de los coeficientes de Fourier por mínimos cuadrados. Los ajustes se realizaron mediante el uso de la herramienta estadística R Project. Cabe aclarar que, inicialmente se ajustaron los modelos con tres y dos armónicas para la media y desviación estándar respectivamente, sin embargo los coeficientes para la tercera armónica en el modelo de media y la segunda armónica para el de desviación estándar, no fueron significativos, por tanto fueron excluidos del modelo. También se excluyó el coeficiente que acompaña a la función seno en la primera armónica del modelo de desviación estándar, debido a que tampoco fue significativo.

Tabla 4.8: Ajuste del modelo de Fourier para la media mensual estimada

	Estimación	Error Est.	Valor t	$\Pr(> t)$
(Intercept)	115.4262	4.5398	25.43	0.0000
$\text{Cos}(\frac{2\pi}{12})$	-30.9318	6.4203	-4.82	0.0013
$\text{Cos}(\frac{2(2\pi)}{12})$	-44.4669	6.4203	-6.93	0.0001
$\text{Sen}(\frac{2(2\pi)}{12})$	29.4912	6.4203	4.59	0.0018
			$\sigma = 15,73$	$R^2 = 0,9202$

En las Tablas 4.8 y 4.9 se presentan el ajuste final de los modelos de Fourier para la

media y desviación estándar mensual estimada. Note que todos los coeficientes son significativos, con un valor de R^2 igual a 0.902 y 0.7682 para el modelo de media y desviación estándar respectivamente. Además los valores de la prueba conjunta de significancia fueron de 9.644e-05 y 0.0001835, en el mismo orden; lo que conlleva a concluir que ambos modelos están bien especificados, según la teoría que rige a este tipo de análisis (Ver Figuras 4.14 y 4.15).

Tabla 4.9: Ajuste del modelo de Fourier para la desviación estandar mensual estimada

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	55.3325	1.6112	34.34	0.0000
$Cos(\frac{2\pi}{12})$	-13.1165	2.2785	-5.76	0.0002
				$\sigma = 5,58 \quad R^2 = 0,7682$

Finalmente, mediante el ajuste de los modelos de Fourier se puede determinar el valor medio y el rango de oscilación para la media y desviación estándar mensual estimadas, describiendo de esta manera la variación a través del año para estos parámetros (ver Tabla 4.10). Dado que para el modelo de media, la segunda frecuencia es la más significativa, se puede tomar como referencia para la interpretación, la amplitud de oscilación a la que se llega mediante la formula $A^2 = \sqrt{a_2^2 + b_2^2}$.

Tabla 4.10: Oscilación de los modelos

Modelo	Valor Medio	Amplitud	Rango de Oscilación	
Media Estimada	115.43	53.36	62.07	168.79
Sd Estimada	55.333	13.12	42.22	68.45

De acuerdo al modelo de Fourier para la media mensual, se afirma que esta fluctúa alrededor de un valor medio de 115.43 mm con un rango de oscilación entre 62.07mm y 168.79 mm. Mientras que para el modelo de Fourier para la desviación estándar mensual, se encontró que tiene una fluctuación alrededor de un valor medio de 53.33mm con un rango de oscilación entre 42.22mm y 68.45mm. Este análisis muestra la gran variación del régimen de lluvias a lo largo del año en cuanto a su valor medio, sin embargo la varianza mensual tiende a ser parecida entre los meses. Además estos resultados son bastante apropiados para esta zona, si se tiene en cuenta que para el Valle del Cauca el clima está determinado por temporadas húmedas o secas, a diferencia de otras partes del mundo, donde cuenta con las diferentes estaciones climáticas.

4.4.1. Validación de los supuestos de Fourier

A continuación se presenta el resumen de los resultados de las pruebas de validación tanto graficas como analíticas descritas en el capítulo 2, con el fin de verificar el cumplimiento

Figura 4.14: Ajuste de Fourier con dos armónicos para el modelo de media mensual estimada

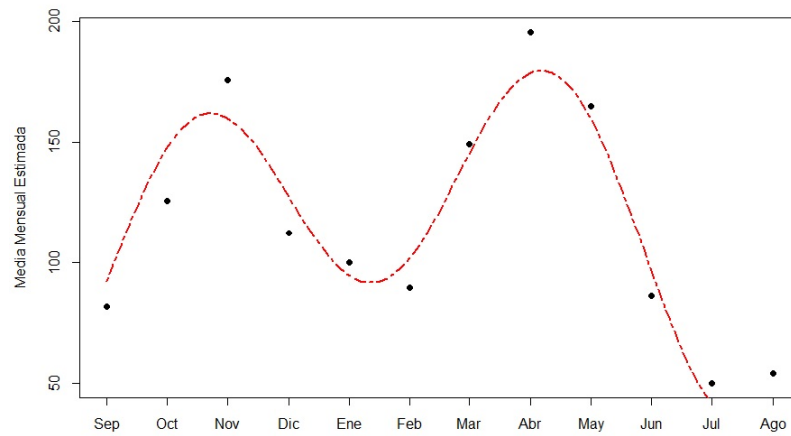


Figura 4.15: Ajuste de fourier con un armónico para la desviación estandar mensual estimada

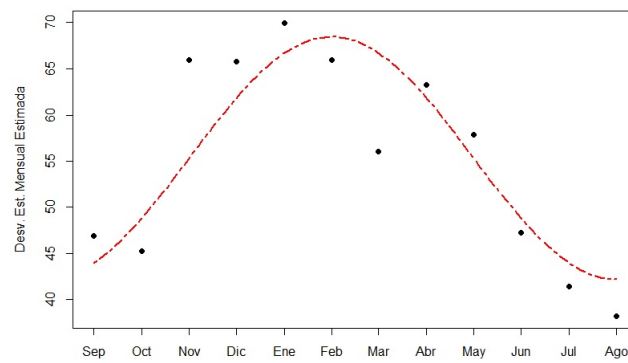


Tabla 4.11: Resumen test para la prueba de los supuestos del modelo de fourier de la media mensual estimada

Prueba de Shapiro-Wilk		Prueba de Breush-Pagan		Prueba de Durbin-Watson	
W	Valor p	BP	Valor p	DW	Valor p
0.9138	0.2386	31.526	0.3687	22.905	0.3357

Tabla 4.12: Resumen test para la prueba de los supuestos del modelo de fourier de la desviación estandar mensual estimada

Prueba de Shapiro-Wilk		Prueba de Breush-Pagan		Prueba de Durbin-Watson	
W	Valor p	BP	Valor p	DW	Valor p
0.9561	0.7268	0.7893	0.3743	17.776	0.2326

de los supuestos teóricos que se hacen alrededor de los modelos de Fourier ajustados.

De las Tablas 4.11 y 4.12 se puede concluir que los supuestos se cumplen a un nivel de confianza del 95 % para los modelos de media y desviación estándar mensual estimadas.

De igual forma las Figuras 4.16 y 4.17 verifican el cumplimiento del supuesto de normalidad; en las Figuras 4.18 y 4.19 no se observa ninguna tendencia específica de los errores vs los valores ajustados, por lo cual se verifica el cumplimiento de homogeneidad de varianza de los errores. Por último, en las Figuras 4.20 y 4.21 se detalla que ninguno de los modelos parece tener correlación entre sus errores en el tiempo, dando cumplimiento al supuesto de independencia.

Figura 4.16: Q-Q plot Normal para los errores del modelo de media mensual estimada

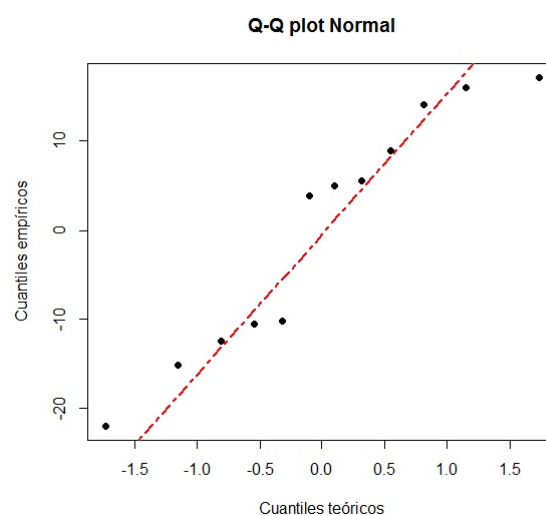


Figura 4.17: Q-Q plot Normal para los errores del modelo de desviación estandar mensual estimada

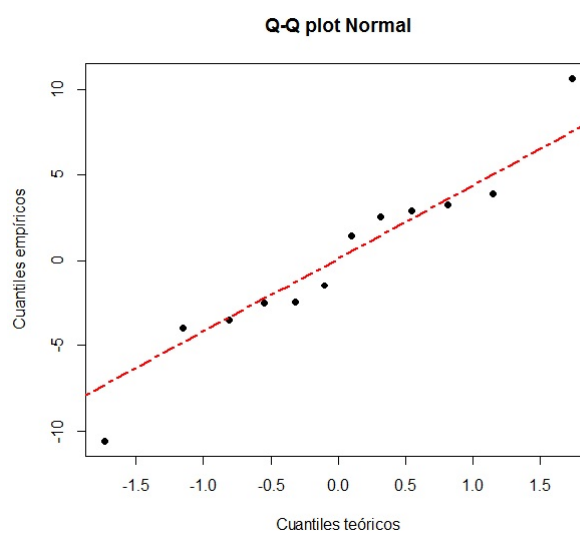


Figura 4.18: Diagrama de los errores vs valores ajustados para el modelo de media mensual estimada

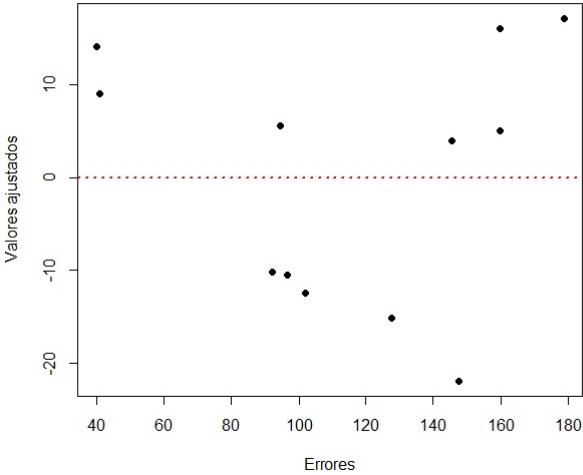


Figura 4.19: Diagrama de los errores vs valores ajustados para el modelo de desviación estandar mensual estimada

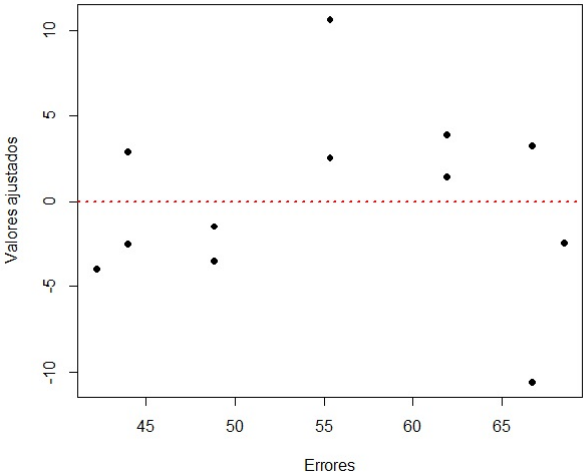


Figura 4.20: Errores a través del tiempo para el modelo de media mensual estimada

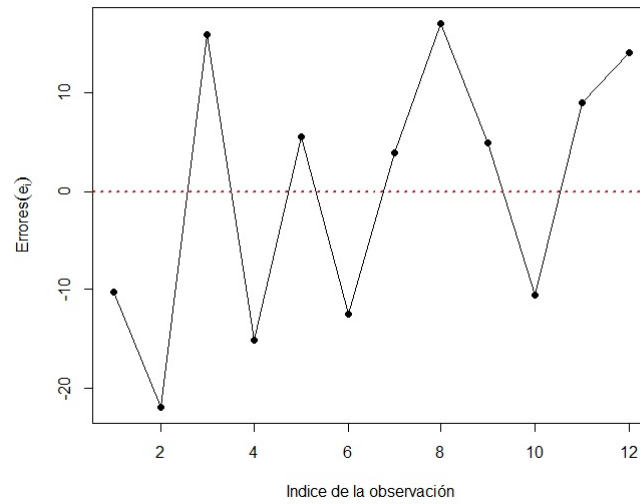
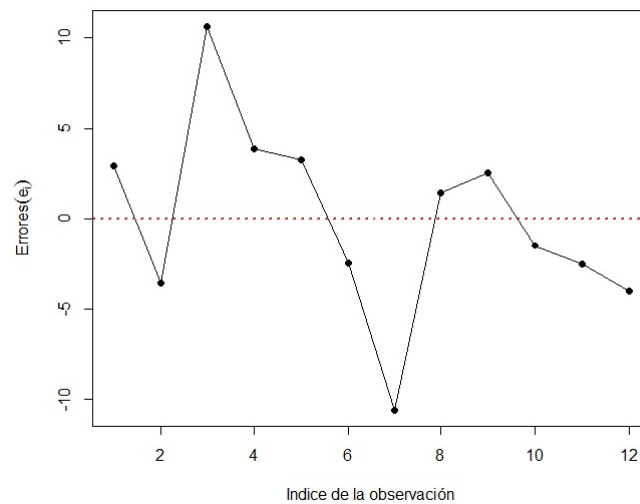


Figura 4.21: Errores a través del tiempo para el modelo de desviación estandar mensual estimada



Capítulo 5

Conclusiones

La cantidad total de precipitación para la zona de estudio está caracterizada en gran parte por una distribución Weibull, donde sus parámetros de forma y escala varían a lo largo del año debido a la variación estacional presente en este tipo de variables. Este resultado es concorde a la afirmación hecha por Gregory et al. (1993) donde probaron varios modelos para la precipitación en el reino unido y encontraron que la distribución Weibull tiene un mejor desempeño en diferentes regiones y estaciones climáticas.

Para el mes de Mayo, se encontró que la distribución Lognormal fue la que mejor ajuste mostro de acuerdo a los criterios de bondad de ajuste usados. Es posible afirmar que este hecho se debe a la especial característica que posee dicha distribución para producir datos extremos; en el contexto del problema se refiere a generar grandes cantidades de lluvia, lo que es típico de este mes, pues hace parte de una de las temporadas de lluvia (ver Figura 4.2). Además, se observó que para el resto de meses, esta distribución tuvo un pobre ajuste, que puede también deberse al hecho ya mencionado, lo que conlleva a que la precipitación mensual estimada se vea alterada por dichos valores extremos, y por ende una sobrestimación de la misma.

Por otra parte, para el mes de Julio se tiene que la distribución que mejor ajuste presentó fue la distribución Gamma. Este hecho es corroborado con el cociente entre el tercer momento central equivalente a la asimetría sobre el coeficiente de variación, el cual para este mes fue de 1.61, que es cercano a 2 y el cual es el valor teórico de dicho cociente para la distribución Gamma, sin tener en cuenta el valor de los parámetros (ver Ecuaciones 2.7 y 2.8). Esta misma justificación fue mencionada en Buishand (1977), donde empleó la distribución Gamma para modelar los datos de dos estaciones meteorológicas ubicadas en Holanda.

También se encontró que la distribución Loi des fuites tuvo un buen desempeño en el mes

de Diciembre, el cual corresponde a una de las temporadas calurosas de la zona. Para este mes se observó pocos días con registros de precipitación, donde el año 1994 fue el que registro una menor cantidad de lluvia total mensual equivalente a 3.6 mm. Este hecho fue lo que hizo que esta distribución tuviera un mejor ajuste en comparación con el resto de distribuciones. Además estos resultados son concordes a los encontrados en la literatura, donde por ejemplo, Huynh et al. (1980) afirma que las secuencias de lluvias mensuales con presencia de valores cero son mejor ajustadas a la distribución *loi des fuites* que con las distribuciones Gamma y Lognormal. Cabe aclarar que su pobre desempeño en el resto del año es debido a que ninguno de los meses presentaba totales mensuales iguales a cero. Por otra parte, de manera general, es posible afirmar que para este mes el número de días lluviosos se distribuye Poisson con una tasa media igual a 5,85 y la cantidad de lluvia recogida en uno de estos días en particular se distribuye Exponencial con media igual a 20 ($1/\rho$).

Cabe resaltar que estas distribuciones en general tienden a dar buenas estimaciones para la media mensual pero incrementan el valor de la desviación estándar mensual (exceptuando la Lognormal, la cual tiende a arrojar valores superiores tanto en la media como en la desviación estándar mensual en comparación con los obtenidos por la distribución empírica de los datos). Esto puede deberse al hecho de que los modelos teóricos son incapaces de ajustar perfectamente a los datos. Se debe tener en cuenta que la capacidad de una función analítica para ajustarse a la distribución de probabilidad empírica para cantidades de precipitación en días húmedos depende de la flexibilidad de la distribución y la naturaleza intrínseca de la forma de la distribución.

En cuanto a la variación estacional de la media y desviación estándar mensual estimadas, se encontró que los modelos de Fourier con dos y un armónico respectivamente son necesarios para describir dicha variación a lo largo del año. Esto es concorde al comportamiento mensual observado en la Figura 4.2, en donde se observa que el régimen de lluvias está caracterizado por dos temporadas, una más intensa que la otra, mientras que la varianza es más o menos constante en cada uno de los periodos. Además los resultados obtenidos en el análisis concuerdan con la afirmación realizada por Woolhiser & Roldán (1986) y Liu et al. (2011) en la cual se indica que de una a tres armónicas son suficientes para el ajuste de las series de Fourier.

Finalmente, se debe tener en cuenta que la elección de uno u otro modelo propuesto para describir o analizar el comportamiento de precipitación dependerá, en gran medida, del uso que se le vaya a dar a dicho modelo. En ese sentido, los modelos más simples, cuya aplicación es inmediata, pueden ser más útiles que los más complejos cuya utilización tal cual es poco probable y que, por lo tanto, requieren de suposiciones que los alejan de su objetivo final.

Recomendaciones

Existen diferentes metodologías que podrían mejorar los resultados obtenidos en este trabajo, estas son:

- El uso de una distribución Pearson tipo III para modelar la cantidad de precipitación total mensual, se presenta como una buena alternativa para caracterizar dicha variable en diferentes condiciones climáticas tal como lo sugieren Hanson & Vogel (2008).
- Respecto de las pruebas de bondad de ajuste empleadas para el contraste de los modelos propuestos, cabe notar que pueden ser mejoradas haciendo uso de técnicas tales como: el diagrama de razón de L-momentos, el cual proveen una conveniente forma visual para detallar las características de los datos muestrales comparados con una distribución estadística teórica (Hanson & Vogel 2008), así como el grafico de probabilidad del coeficiente de correlación (PPCC).
- La estimación de los coeficientes de Fourier puede llevarse a cabo haciendo uso de estimaciones de máxima verosimilitud a través de procesos de optimización multi-variados, empleando como valores semillas los coeficientes de Fourier hallados por mínimos cuadrados (Para más detalles ver Roldán (1994)).

La metodología utilizada en este proyecto para modelar la cantidad de precipitación mensual puede ser replicada para desarrollar el análisis diario de dicha variable. En ese contexto se puede emplear la distribución mixta exponencial la cual provee buen ajuste en diversas condiciones climáticas tal como lo sugiere Roldán (1994).

Otro enfoque utilizado para analizar la variable precipitación es por medio de un modelo de regionalización el cual mediante la modelación diaria tanto del proceso de ocurrencia como de cantidad de lluvia, además de uso de variables orográficas y climáticas permite describir el comportamiento del régimen de lluvia en una región determinada. Cabe aclarar que para ello se deben usar los registros diarios de estaciones meteorológicas ubicadas en diferentes lugares de la región o zona investigada.

Anexo A

Código en R project para hallar la función de distribución LDF

La distribución loi des fuites no se encuentra implementada en el programa estadístico R Project. Por este motivo, presentamos a continuación el código de R para hallar la función de distribución LDF, el cual fue utilizado para realizar los respectivos cálculos.

****Función de distribución loi des fuites (LDF)**
****dias** es igual al número de días lluviosos para cada mes
****theta** y **rho** corresponden a los parámetros de la distribución LDF

```
LDF = function(y, theta, rho)
{
  PLDF = 0
  for(j in 1 : 20)
  {
    p = 0
    q = matrix(0, ncol = dias[j], nrow = dias[j], byrow = F)
    A = 0
    s1 = 0
    for(k in 1 : dias[j])
    { p[k] = (theta ^ k * exp(-theta)) / (factorial(k))
      for(l in 1 : k)
      {
        q[l, k] = (((rho * y[j]) ^ (l - 1)) / (factorial(l - 1))) * (exp(-rho * y[j]))
      }
    }
  }
}
```

```

s1[k] = sum(q[, k])
A[k] = p[k] * (1 - s1[k])
}
PLDF[j] = exp(-theta) + sum(A)
}
return(round(PLDF, 4))
}

```

Anexo B

Código en R project para realizar las estimaciones de máxima verosimilitud de los parámetros de la distribución LDF

A continuación se presenta el código implementado en el programa estadístico R para realizar las estimaciones de máxima verosimilitud de los parámetros de la distribución LDF. Dicho código estaba basado en la ecuación 2.23 de la sección 2 y empleando el algoritmo de Newton-Raphson.

```
funcionLogVer = function(p,x)
{
p1 = p[1]
r = matrix(0,nrow = length(x),ncol = max(k),byrow = T)
s = 0
for(j      in      1 : length(x))
{
for(i      in      1 : k[j])
{
r[j,i] = (p1 ^ i * x[j] ^ (i - 1)) / (factorial(i) * factorial(i - 1))
}
s[j] = sum(r[j,])
}
f = -(-2 * sqrt(p1 * N * sum(x)) + sum(log(s)))
```

```
return(f)
}
```

```
** Estimación de parámetros por el método de los momentos
```

```
rho = 2 * (mean(y)/var(y))
```

```
theta = 2 * (mean(y) ^ 2/var(y))
```

```
lambda = rho * theta
```

```
lambdaVer = optim(par = c(lambda), fn = funcionLogVer,method = c("BFGS"),x = y)
```

```
thetaVer = round(sqrt(lambdaVer$par * (sum(y)/N)),3)
```

Bibliografía

- Abramowitz, M. & Stegun, I. A. (1970), 'Handbook of mathematical functions', *Dover Publications, Inc.* .
- Alcaide, M. (1986), 'Análisis estacional y regional de la precipitación en el valle del guadaluquivir', *Tesis Doctoral* .
- Asín, J. (2003), 'Modelos para la precipitación diaria en el marco de los modelos lineales generalizados', *Tesis Doctoral* .
- Bartels, R. (1982), 'The rank version of von neumann ratio test for randomness', *Journal of the American Statistical Association* .
- Behar, R. (2003), 'Validación de supuestos en el modelo de regresión', *Serie Monografias vol. 1.*
- Bernal, G. & et. al (1989), 'Estudio sobre el régimen de la precipitación en colombia', *Instituto colombiano de Hidrología, Meteorología y Adecuación de Tierras (HIMAT)* .
- Bloomfield, P. (2000), 'Fourier analysis of time series: An introduction', *2 edition* .
- Buishand, T. A. (1977), 'Stochastic modelling of daily rainfall sequences', *Communications Agricultural University Wageningen, The Netherlands* .
- Carcedo, J. M. (2000), 'Introducción al tratamiento de series temporales mediante filtros', *PhD thesis* .
- Carvajal, Y. E., Jiménez, H. & Materón, H. M. (1998), 'Incidencia del fenómeno enso en la hidroclimatología del valle del río cauca-colombia', *Bull. Inst. fr. études andines vol. 27.*
- Coe, R. & Stern, R. D. (1982), 'Fitting models to daily rainfall data', *Journal of Applied Meteorology* .
- Cox, D. R. (1962), 'Renewal theory', *Methuen & Co.* .

- Cromwell, J. B., Labys, W. C. & Terraza, M. (1994), 'Univariate test for time series models', *Sage publications, Inc.* .
- CVC (2013), 'Boletín hidroclimático', *Fecha de consulta: 28 de agosto de 2013* .
- Detzel, D. & Mine, M. (2011), 'Generation of daily synthetic precipitation series: Analyses and application in la plata river basin', *The Open Hydrology Journal* **vol. 5**.
- Dorado, J. & et al. (2006), 'Ajuste de modelos probabilísticos para el estudio de la variabilidad espacio-temporal de la precipitación: Caso de estudio sara-brut', *Revista Meteorología Colombiana* **Vol. 10**.
- Feller, W. (1950), 'An introduction to probability theory and its applications', *Third Edition* **vol. 1**.
- Fisher, R. A. & Cornish, E. A. (1960), 'The percentile points of distributions having known cumulants', *Technometrics* **Vol. 2**.
- Fogel, K. (2005), *Producing Open Source Software, How to Run a Successssful Free Software Project*.
- Gómez, G. & Hernández, L. (2013), 'Estimación de la ocurrencia de precipitación diaria mediante un proceso markoviano de segundo orden', *Tesis de Grado (Universidad del Valle)* .
- Guenni, L. & et al. (2008), 'Análisis de la tendencia y la estacionalidad de la precipitación mensual en venezuela', *Revista Colombiana de Estadística* **Vol. 31**.
- Hanson, L. S. & Vogel, R. (2008), 'The probability distribution of daily rainfall in the united states', *American Society of Civil Engineers* .
- Huxley, A. (1954), *The Doors of Perception*, Edhasa.
- Huynh, N. P., Anat, A. & Apichai, S. (1980), 'Distribution of monthly rainfall in northeast thailand', *Southeast Asian Studies* **Vol. 18**.
- Ibáñez, V. (1982), 'Elaboración de una metodología para estimar la persistencia de las precipitaciones. su aplicación al cultivo del arroz en levante', *Tesis Doctoral* .
- IDEAM (2012a), 'Atlas climatologico (cuarta parte). temas relacionados', *Fecha de consulta: 28 de agosto de 2013* .
- IDEAM (2012b), 'Atlas climatologico (primera parte). aspectos nacionales', *Fecha de consulta: 28 de agosto de 2013* .
- IDEAM (2012c), 'Atlas climatologico (segunda parte). distribución espacio-temporal de las variables del clima', *Fecha de consulta: 28 de agosto de 2013* .

- IPCC (2008), ‘Cambio climático 2007 informe de síntesis’, *Grupo Intergubernamental de Expertos sobre el Cambio Climático* .
- Joanest, D. N. & Gill, C. A. (1998), ‘Comparing measures of sample skewness and kurtosis’, *The statistician* **47**.
- Katz, R. W. (1977), ‘An application of chain-dependent processes to meteorology’, *J. Appl. Prob.* .
- Khuri, A. I. (2003), ‘Advanced calculus with applications in statistic’, **2 edn**.
- León, G. (2012), ‘Aspectos de la circulación atmosférica de gran escala sobre el noroccidente de suramérica asociada al ciclo enos 2009-2010 y sus consecuencias en el régimen de precipitación en colombia’, (*Consultado en la página de IDEAM Martes 13 de Agosto de 2013*) **1**.
- LIU, Y., ZHANG, W., SHAO, Y. & ZHANG, K. (2011), ‘A comparison of four precipitation distribution models used in daily stochastic models’, *Advances in Atmospheric Sciences* **vol. 28**.
- Manríquez, J. C. & Valdez, M. (2012), ‘Test de razón de verosimilitud generalizada’, *X Congreso Latinoamericano de Sociedades de Estadística* .
- Maronna, R. A. (1995), ‘Probabilidad y estadística elementales para estudiantes de ciencias’.
- Marquez Dos Santos, M. J. (2001), ‘Estadística básica, un enfoque no paramétrico’, *UNAM* .
- Montealegre, J. E. & Pabon, J. D. (2000), ‘La variabilidad climática interanual asociada al ciclo el niño, la niña oscilación del sur y su efecto en el patrón pluviométrico de colombia’, *Revista Meteorología Colombiana* **vol. 2**.
- Montgomery, D. & Peck, E. A. (2005), ‘Introducción al análisis de regresión lineal’.
- Moreno, M. F. & Roldán, J. (2008), ‘Interpolación espacial de series de precipitación diaria’, *Ingeniería del Agua* **Vol. 15**.
- Mormeneo, I. & F., C. (2001), ‘Application of a model on generation of daily precipitations’, *Revista Brasileira de Agrometeorologia* **Vol. 9**.
- Murillo, C. & González, B. (2004), ‘Manual de econometría’, *ULPGC* .
- Norlund, N. E. (1924), ‘Vorlesungen uber differenzenrechnung’, *Springer* .
- Pabón, J. & et al. (1998), ‘El medio ambiente en colombia, la atmósfera, el tiempo y el clima’, *Reporte técnico, IDEAM* .

- Peréz, L. (1994), ‘Modelo estocástico de la precipitación diaria. aplicación a las estaciones de córdoba y sevilla’, *Trabajo Fin de Carrera* .
- Ribstein, P. (1983), ‘Loi des fuites’, *Hydrologue à l’ORSTOM* .
- Roldán, J. (1994), ‘Tendencias actuales en el modelado de la precipitación diaria’, *Ingeniería del agua* **vol. 1**.
- Ruiz, J. (2002), ‘Cambio climático en temperatura, precipitación y humedad relativa para colombia usando modelos meteorológicos de alta resolución (panorama 2011-2100)’, *Nota técnica del IDEAM* **1**.
- Sanso, B. & Guenni, L. (2000), ‘A nonstationary multisite model for rainfall’, *Journal of the American Statistical Association* **Vol. 95**.
- Selker, J., Duan, J. & Grant, G. E. (1998), ‘Evaluation of probability density functions in precipitation models for the pacific northwest’, *Journal of the american water resources association* **34**.
- SIAC (2013), ‘Fenomenos el niño y la niña, información histórica’, *Sistema de Información Ambiental de Colombia* .
- Smith, R. & Schreiber, H. A. (1974), ‘Point process of seasonal thunderstorm rainfall 2 rainfall depth probabilities’, *Water Resour. Res.* .
- Solomon, S. & et al. (2007), ‘Climate change 2007: The physical science basis’, *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* **Vol. 31**.
- Stern, R. D. & Coe, R. (1984), ‘A model fitting analysis of daily rainfall data’, *Royal Statistical Society* .
- Suhaila, J. & Jemain, A. (2007a), ‘Fitting daily rainfall amount in peninsular malaysia using several types of exponential distributions’, *Journal of Applied Sciences Research* **vol. 10**.
- Suhaila, J. & Jemain, A. A. (2007b), ‘Fitting daily rainfall amount in malaysia using the normal transform distribution’, *Journal of Applied Sciences Research* **vol. 14**.
- Thom, H. C. S. (1958), ‘A note on the gamma distribution’, *Office of Climatology, U. S. Weather Bureau* **Vol. 86**.
- Wilk, M. B. & Gnanadesikan, R. (1968), ‘Probability plotting methods for the analysis of data’, *Biometrika* **55**.

- Wilks, D. S. (1998), 'Multisite generalization of a daily stochastic precipitation generation model', *Journal of Hydrology* .
- Wong Raymond, K. W. (1977), 'Weibull distribution, iterative likelihood techniques and hydrometeorological data', *Journal of Applied Meteorology* **16**.
- Woolhiser, D. A., Hanson, C. L. & Richardson, C. W. (1988), 'Microcomputer program for daily weather simulation', *U.S.D.A, Agric. Res. Serv.* .
- Woolhiser, D. & Pegram, G. (1979), 'Maximum likelihood estimation of fourier coefficients to describe seasonal variations of parameters in stochastic daily precipitation models', *Journal of Applied Meteorology* **vol. 18**.
- Woolhiser, D. & Roldán, J. (1986), 'Seasonal and regional variability of parameters for stochastic daily precipitation models: South dakota, u.s.a.', *Water Resources Research* **vol. 22**.
- Yevjevich, V. (1972), 'Structural analysis of hydrologic time series', *Hydrology Paper* **Vol. 56**.
- Zhang, J. (2002), 'Powerful goodness of fit test based on the likelihood ratio', *Journal Royal Statistic Society* **vol. 64**.