



UNIVERSIDAD NACIONAL DE COLOMBIA

Análisis de frecuencia hidrológico multivariado para eventos extremos mediante funciones cópula arquimedianas. Casos de estudio: cuenca baja del Río Tunjuelo y región de la Mojana (Colombia)

Wilmar Jair Gómez Rios

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de Ingeniería Civil y Agrícola
Bogotá D.C., Colombia
Año 2017

Análisis de frecuencia hidrológico multivariado para eventos extremos mediante funciones cópula arquimedianas. Casos de estudio: cuenca baja del Río Tunjuelo y región de la Mojana (Colombia)

Wilmar Jair Gómez Rios

Tesis presentada como requisito parcial para optar al título de:
Magister en Ingeniería Recursos Hidráulicos

Director:
Ing. Civil. MSc. PhD. Nelson Obregón Neira

Línea de Investigación:
Modelación de eventos naturales extremos y estructuras de dependencia de datos hidrológicos

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de Ingeniería Civil y Agrícola
Bogotá D.C., Colombia
2017

A mi familia

Lluvia

Hay un bosque de páramo en la noche del sueño,
un corazón
hecho malezas
entre el fango y la niebla que encanece la montaña.
Hay una sepultura
con cántaros de oro para recoger el vaho impuro del mundo.
Lo que buscamos en el origen de la lluvia
es el alimento,
no el camino
ni el destino
que vienen rodando por el aire.

Henry Alexander Gómez Ríos

Agradecimientos

Agradezco a mi familia, mis padres Ismael Gómez y Livia Rios quienes me han acompañado a lo largo de toda mi carrera profesional, por ser mi fortaleza en momentos difíciles y por brindarme su apoyo incondicional. Doy gracias a mis hermanos, Henry, Arley y Daniel quienes han sido parte importante en mi vida, su compañía y sus enseñanzas me han mostrado luz en medio de las sombras llenando mi vida de claridad y de aprendizaje.

Quiero agradecer a la Universidad Nacional de Colombia por haberme permitido realizar mis estudios de maestría y por darme la oportunidad de ser parte de ella. También agradezco el apoyo financiero del programa de becas para estudiantes sobresalientes de posgrado de la Universidad Nacional de Colombia.

Agradezco especialmente al profesor Nelson Obregón Neira Ph.D., M.Sc., I.C. por haberme aceptado como estudiante y dirigir mi tesis, por su colaboración, sus consejos y por poner a disposición su gran conocimiento para la elaboración de este trabajo de investigación.

Quiero hacer extensiva mi gratitud a mis amigos y colegas quienes me han brindado todo su apoyo y me han motivado para seguir adelante, agradezco a Alejandro, Oswaldo y Pablo con quienes he compartido grandes momentos y grandes experiencias, agradezco al Ingeniero Eduardo Zamudio por haber compartido conmigo sus conocimientos. También quiero agradecer a Esteban, Felipe, Gladia, Juancho, Mafe y Sergio quienes han creído en mí hasta el último momento.

Resumen

Mediante el análisis de la variabilidad espacial y temporal de los eventos extremos de precipitación se puede prevenir o reducir la amenaza y el riesgo. Muchos proyectos de recursos hídricos requieren distribuciones de probabilidad conjunta de variables aleatorias como la intensidad de las precipitaciones y su duración, que no pueden ser independientes entre sí. El problema de definir un modelo de probabilidad para las observaciones de varias variables dependientes se simplifica en gran medida mediante la distribución conjunta, en términos de sus marginales, mediante la adopción de cópulas. Este documento presenta un marco general del análisis de frecuencia conjunto bivariado y multivariado mediante cópulas arquimedianas para eventos extremos de naturaleza hidroclimatológica, tales como tormentas intensas y caudales instantáneos máximos. Este análisis se llevó a cabo en la cuenca baja del río Tunjuelo para eventos de precipitación y en la región de la Mojana para eventos de crecidas. Los resultados obtenidos muestran que para un estudio conjunto de la intensidad-duración-frecuencia se pueden obtener curvas IDF mediante cópulas y de esta forma establecer información más precisa y confiable de las tormentas de diseño y los riesgos asociados. Se muestra cómo el uso de cópulas simplifica en gran medida el estudio de las distribuciones multivariadas que introducen el concepto de periodo de retorno conjunto empleado para representar las necesidades de diseños hidrológicos apropiadamente en el análisis de frecuencia de crecidas.

Palabras clave: Funciones cópula arquimedianas, análisis de frecuencia de crecidas, periodo de retorno conjunto, curvas IDF, distribución de probabilidad conjunta.

Abstract

By analyzing the spatial and temporal variability of extreme precipitation events we can prevent or reduce the threat and risk. Many water resources projects require joint probability distributions of random variables such as precipitation intensity and duration, which can not be independent with each other. The problem of defining a probability model for observations of several dependent variables is greatly simplified by the joint distribution in terms of their marginal by taking copulas. This thesis presents a general framework set frequency analysis bivariate and multivariate using Archimedean copulas for extreme events of hydroclimatological nature such as severe storms and maximum instantaneous flows. This analysis was conducted in the lower basin Tunjuelo River for precipitation events and in the Mojana region for flood events. The results obtained show that for a joint study of the intensity-duration-frequency, IDF curves can be obtained through copulas and thus establish more accurate and reliable information from design storms and associated risks. It shows how the use of copulas greatly simplifies the study of multivariate distributions that introduce the concept of joint return period used to represent the needs of hydrological designs properly in flood frequency analysis.

Keywords: Archimedean copula functions, flood frequency analysis, Multivariate return period, IDF Curves, joint probability distribution.

Contenido

| | |
|--|-------------|
| Agradecimientos | VII |
| Resumen | IX |
| Lista de figuras | XIII |
| Lista de tablas | XVI |
| 1. Introducción | 2 |
| 1.1. Generalidades | 4 |
| 1.2. Justificación | 6 |
| 1.3. Objetivos | 8 |
| 1.3.1. Objetivo general | 8 |
| 1.3.2. Objetivos específicos | 8 |
| 1.4. Resumen del contenido | 8 |
| 2. Funciones de distribución de probabilidad y análisis de frecuencia | 10 |
| 2.1. Periodo de retorno univariado y posiciones de graficación | 10 |
| 2.2. Pruebas de bondad de ajuste | 12 |
| 2.2.1. Prueba Kolmogorov-Smirnov | 13 |
| 2.2.2. Prueba χ^2 | 13 |
| 2.3. Tormentas de diseño | 14 |
| 2.3.1. Antecedentes de las curvas IDF | 14 |
| 2.3.2. Relaciones intensidad-duración-frecuencia | 15 |
| 2.4. Representación de las curvas IDF en el área de estudio | 17 |
| 2.4.1. Localización general del área de estudio | 17 |
| 2.4.2. Recopilación de información y estaciones seleccionadas | 19 |
| 3. Cópulas Bivariadas | 28 |
| 3.1. Introducción a las cópulas | 28 |
| 3.2. Definición de Cópula | 29 |
| 3.3. Teorema de Sklar | 30 |
| 3.3.1. Densidad de la cópula | 31 |

| | |
|---|-----------|
| 3.4. Medidas de correlación de rango | 31 |
| 3.4.1. Tau de Kendall τ | 32 |
| 3.4.2. Rho de Spearman ρ | 32 |
| 3.5. Familias de Cópulas | 33 |
| 3.6. Cópulas Arquimedianas | 33 |
| 3.7. Familias de cópulas arquimedianas | 36 |
| 3.7.1. Cópula Gumbel Hougaard | 36 |
| 3.7.2. Cópula Ali-Mikhail-Haq | 37 |
| 3.7.3. Cópula Frank | 37 |
| 3.7.4. Cópula Cook-Johnson (Clayton) | 38 |
| 3.7.5. Cópula Joe | 39 |
| 3.8. Cópula Empírica | 40 |
| 3.9. Estimación del parámetro de dependencia θ | 41 |
| 3.10. Selección de la Cópula | 42 |
| 3.11. Periodo de retorno condicional y posiciones de graficación para análisis de frecuencia bivariados | 46 |
| 3.12. Distribuciones condicionales bivariadas a través de cópulas y correspondientes periodos de retorno | 49 |
| 3.13. Aplicación de las cópulas bivariadas en el área de estudio | 50 |
| 3.13.1. Construcción de curvas IDF por medio del método de la cópula | 50 |
| 3.13.2. Análisis de resultados de la construcción de curvas IDF | 63 |
| 3.14. Aplicación de las cópulas bivariadas Multisitio | 64 |
| 3.14.1. Selección de la cópula en el área de estudio | 64 |
| 3.14.2. Periodo de retorno conjunto multisitio | 71 |
| 3.14.3. Análisis de resultados de la aplicación de cópulas bivariadas multisitio | 74 |
| 3.15. Aplicación de cópulas bivariadas en el análisis de series históricas | 75 |
| 3.15.1. Análisis de resultados de la aplicación de cópulas bivariadas en el análisis de series históricas | 82 |
| 3.16. Conclusiones del capítulo | 83 |
| 4. Cópulas multivariadas | 85 |
| 4.1. Preliminares | 85 |
| 4.2. Definición de cópula N-dimensional | 86 |
| 4.3. Teorema de Sklar N-dimensional | 86 |
| 4.4. Cópulas arquimedianas N-dimensionales | 87 |
| 4.5. Familias de cópulas arquimedianas N-dimensionales | 88 |
| 4.6. Estimación del parámetro θ para cópulas N-dimensionales | 89 |

| | |
|--|------------|
| 4.7. Periodo de retorno conjunto y posiciones de graficación para análisis de frecuencia multivariado | 91 |
| 4.8. Distribuciones condicionales trivariadas a través de cópulas y correspondientes periodos de retorno | 94 |
| 4.9. Aplicación de las cópulas Multivariadas en el área de estudio | 96 |
| 4.9.1. Construcción de curvas IDF conjuntas multisitio mediante el método de la cópula | 96 |
| 4.9.2. Análisis de resultados de la construcción de curvas IDF conjuntas multisitio | 115 |
| 4.9.3. Análisis de frecuencia multivariado mediante el método de la cópula . | 116 |
| 4.9.4. Periodo de retorno conjunto multivariado multisitio | 119 |
| 4.9.5. Análisis de resultados del periodo de retorno conjunto multivariado multisitio | 125 |
| 4.10. Conclusiones del capítulo | 127 |
| 5. Conclusiones y recomendaciones | 129 |
| 5.1. Conclusiones Generales | 129 |
| 5.2. Recomendaciones | 130 |
| A. Anexo: Capítulo 2. Distribuciones de Probabilidad y Análisis de frecuencia | 132 |
| B. Anexo: Capítulo 3. Cópulas Bivariadas | 151 |
| C. Anexo: Capítulo 4. Cópulas Multivariadas | 158 |
| Bibliografía | 169 |

Lista de Figuras

| | | |
|-------|---|----|
| 2-1. | Cuenca del río Tunjuelo en el área de jurisdicción de la CAR. | 18 |
| 2-2. | Ubicación de las estaciones hidrometeorológicas seleccionadas. | 20 |
| 2-3. | Intensidad para diferentes eventos de tormenta máxima seleccionada de una serie histórica para una duración dada. Estación El Bosque. | 21 |
| 2-4. | Periodo de retorno asociado a cada uno de los registros tomados en la estación El Bosque. | 22 |
| 2-5. | Superficie de regresión generada a partir de la ecuación 2-18 para la estación El Bosque. | 23 |
| 2-6. | Curvas IDF generadas a partir de los registros históricos de la estación El Bosque. | 24 |
| 2-7. | Diferentes distribuciones de probabilidad asociadas a los registros de intensidad (mm/h) para la estación El Bosque. | 25 |
| 2-8. | Ajuste mediante el estadístico Anderson-Darling. (Minitab, 2016). | 26 |
| 2-9. | Histograma de frecuencia ajustado para la duración de la tormenta en minutos. (Estación El Bosque). | 27 |
| 3-1. | Esquema de la propiedad de la cópula. | 30 |
| 3-2. | Relación entre el τ de Kendall y el parámetro θ para la cópula Frank. | 38 |
| 3-3. | Relación entre el τ de Kendall y el parámetro θ para la cópula Joe. | 40 |
| 3-4. | Relación entre el τ de Kendall y el parámetro θ para la cópula Frank. | 52 |
| 3-5. | Curvas IDF por medio de la aplicación del método de la cópula para la estación El Bosque (Distribución asumida: Exponencial). | 55 |
| 3-6. | Comparación de las curvas IDF entre el método de la cópula y el método empírico para la estación El Bosque. | 55 |
| 3-7. | Curvas IDF por medio de la aplicación del método de la cópula para la estación El Bosque (Distribución ajustada: LogNormal). | 57 |
| 3-8. | Comparación de las curvas IDF aplicando método de la cópula y el método empírico para la estación El Bosque. | 57 |
| 3-9. | Curvas IDF por medio de la aplicación del método de la cópula para la estación El Delirio. | 59 |
| 3-10. | Comparación de las curvas IDF aplicando el método de la cópula y el método empírico para la estación El Delirio. | 59 |

| | |
|--|-----|
| 3-11. Curvas IDF por medio de la aplicación del método de la cópula para la estación Santa Lucía. | 60 |
| 3-12. Comparación de las curvas IDF aplicando el método de la cópula y el método empírico para la estación Santa Lucía. | 60 |
| 3-13. Curvas IDF por medio de la aplicación del método de la cópula para la estación Bosa Barreno. | 61 |
| 3-14. Comparación de las curvas IDF aplicando el método de la cópula y el método empírico para la estación Bosa Barreno. | 61 |
| 3-15. Método gráfico para la selección de la cópula. Estimación No paramétrica contra la estimación paramétrica (método de la cópula). | 66 |
| 3-16. Pseudo-observaciones construidas a partir de los datos de las estaciones El Delirio y Santa Lucía. | 68 |
| 3-17. Comparación entre el Periodo de retorno empírico y el periodo de retorno conjunto multisitio de las estaciones El Delirio y Santa Lucía para diferentes duraciones. | 72 |
| 3-18. Comparación entre el Periodo de retorno empírico y el periodo de retorno conjunto multisitio de las estaciones Juan Rey y Quiba para diferentes duraciones. | 73 |
| 3-19. Comparación de las curvas IDF aplicando el método de la cópula y el método empírico para diferentes series históricas para la estación Santa Lucía. | 77 |
| 3-20. Comparación de las curvas IDF aplicando el método de la cópula y el método empírico para diferentes series históricas para la estación El Bosque. | 80 |
| 4-1. Periodo de retorno conjunto para una duración de 15 minutos. Estaciones Juan Rey y Quiba. | 99 |
| 4-2. Periodo de retorno conjunto para diferentes duraciones. Estaciones El Delirio y santa lucía. | 100 |
| 4-3. Distancias al centroide de cada estación en la cuenca baja del río Tunjuelo | 102 |
| 4-4. Dependograma que resume los resultados de la prueba de independencia de la serie hidrológica para las estaciones El Delirio y Santa Lucía con $lag.max = 4$ | 104 |
| 4-5. Dependograma que resume los resultados de la prueba de independencia mutua de los datos trivariados para las estaciones El Delirio y Santa Lucía. | 105 |
| 4-6. Curvas IDF conjuntas por medio de la aplicación del método de la cópula para las estaciones El delirio y Santa Lucía. | 112 |
| 4-7. Comparación entre las curvas IDF conjuntas aplicando el método de la cópula y el método empírico para las estaciones El delirio y Santa Lucía. | 112 |
| 4-8. Curvas IDF conjuntas por medio de la aplicación del método de la cópula para las estaciones Juan Rey y Quiba. | 113 |

| | |
|---|-----|
| 4-9. Comparación entre las curvas IDF conjuntas aplicando el método de la cópula y el método empírico para las estaciones Juan Rey y Quiba. | 113 |
| 4-10. Localización general de la Mojana (Adaptación, 2016). | 116 |
| 4-11. Ubicación de las estaciones pertenecientes a la región de la Mojana. (Adaptación, 2016). | 118 |
| 4-12. Histograma de frecuencia ajustado a diferentes distribuciones de probabilidad mediante la prueba Kolmogorov-Smirnov para las estaciones pertenecientes a la región de la Mojana. | 120 |
| 4-13. Periodo de retorno multivariado multisitio para la estación Montelíbano - San Jorge. | 122 |
| 4-14. Periodo de retorno multivariado multisitio para la estación La Coquera - Cauca. | 122 |
| 4-15. Periodo de retorno multivariado multisitio para la estación La Esperanza - Nechi. | 123 |
| 4-16. Periodo de retorno multivariado multisitio para la estación La Raya - Caño Caribona. | 123 |
| 4-17. Periodo de retorno multivariado multisitio para la estación Armenia - Brazo de Loba. | 124 |
| A-1. Distribución Exponencial. | 133 |
| A-2. Distribución Normal. | 134 |
| A-3. Distribución Lognormal. | 135 |
| A-4. Distribución Gamma. | 138 |
| A-5. Distribución Pearson Tipo III. | 139 |
| A-6. Distribución Log Pearson Tipo III. | 141 |
| A-7. Distribución Gumbel. | 142 |
| A-8. Series de tiempo de intensidad para diferentes duraciones en cada una de las estaciones analizadas. | 143 |
| C-1. Series de tiempo de caudal instantáneo (m^3/s) para las estaciones pertenecientes a la región de la Mojana. | 168 |

Lista de Tablas

| | |
|--|-----|
| 2-1. Posiciones de graficación más usadas en hidrología | 11 |
| 2-2. Distribuciones de probabilidad usadas en hidrología | 11 |
| 2-3. Estaciones pluviográficas seleccionadas para la construcción de curvas IDF . | 19 |
| 2-4. Distribuciones de probabilidad marginales asociadas a cada una de las estaciones seleccionadas | 27 |
| 3-1. Coeficiente de correlación τ de Kendall y parámetro de la cópula Frank para las estaciones seleccionadas | 56 |
| 3-2. Comparación entre curvas IDF por medio de cópulas y Curvas IDF Empíricas | 62 |
| 3-3. Parámetro θ para las estaciones El Delirio y Santa Lucía | 65 |
| 3-4. Resumen de la prueba de bondad de ajuste para las familias candidatas. Parámetro θ para las estaciones El Delirio y Santa Lucía | 70 |
| 3-5. Parámetro θ para diferentes series históricas de la estación Santa Lucía . . . | 76 |
| 3-6. Comparación entre curvas IDF por medio de cópulas y Curvas IDF Empíricas para la estación Santa Lucía | 78 |
| 3-7. Parámetro θ para diferentes series históricas de la estación El Bosque | 79 |
| 3-8. Comparación entre curvas IDF por medio de cópulas y Curvas IDF Empíricas para la estación El Bosque | 81 |
| 4-1. Resumen de la prueba de bondad de ajuste para las familias candidatas. Parámetro θ trivariado para las estaciones El delirio, Santa Lucía y el inverso de la distancia | 109 |
| 4-2. Comparación entre curvas IDF conjuntas multisitio por medio de cópulas y aplicando el método empírico | 114 |
| 4-3. Estaciones hidrológicas seleccionadas pertenecientes a la región de la Mojana | 117 |
| 4-4. Distribuciones de probabilidad marginales asociadas a cada una de las estaciones pertenecientes a la región de la Mojana | 119 |
| 4-5. Parámetro θ bivariado para las estaciones pertenecientes a la región de la Mojana | 121 |
| 4-6. Parámetro θ multivariado para las estaciones pertenecientes a la región de la Mojana | 121 |

| | |
|--|-----|
| A-1. Datos nativos de intensidad (mm/h) para la estación El Bosque | 144 |
| A-2. Datos nativos de intensidad (mm/h) para la estación El Delirio | 145 |
| A-3. Datos nativos de intensidad (mm/h) para la estación Santa Lucía | 146 |
| A-4. Datos nativos de intensidad (mm/h) para la estación Bosa Barreno | 147 |

1. Introducción

El análisis estadístico de eventos extremos es importante en varias disciplinas, incluyendo la ingeniería y la hidrología. Muchos de los problemas hidrológicos involucran variables aleatorias que juegan un papel importante en el análisis y diseño de estructuras. Generalmente estas variables no son independientes. Por ejemplo, las diferentes combinaciones de intensidad de lluvia y duración de la tormenta, pueden generar tormentas que muestran características muy diferentes; el análisis de los ríos puede depender fuertemente de las características del pico de inundación y el volumen de inundación, el análisis de frecuencia requiere el estudio en conjunto de la intensidad–duración–frecuencia de un evento de lluvia (curvas IDF), etc. Por lo tanto, es de importancia fundamental vincular estas variables aleatorias a través de distribuciones marginales con el fin de obtener una distribución conjunta para describir las características principales de los eventos hidrológicos. En los últimos años se han introducido algunos métodos multivariados en aplicaciones hidrológicas y ambientales. Al principio, la función de distribución conjunta más utilizada fue la de Gauss. Esta distribución es ampliamente estudiada en la literatura y fácil de aplicar, pero tiene el limitante de que las distribuciones marginales deben ser normales. Goel et al. (1998) logran la condición de que los datos se ajusten a una distribución normal por medio de la transformación de Box-Cox, sin embargo, esta transformación no siempre garantiza que las series sigan una distribución de Gauss y en ocasiones proporcionan distorsiones significativas de las propiedades estadísticas de la muestra, Yue (2000c).

Para obtener distribuciones de probabilidad conjunta, Bacchi et al. (1994) aplican una distribución Gumbel bivariada con marginales exponenciales, Yue and Wang (2004), sugieren una distribución gamma bivariada en el análisis de frecuencia de crecidas, Johnson et al. (2002) presentan una descripción detallada de los modelos anteriores y además de otros como la distribución exponencial multivariada, la Dirichlet, Liouville, Logistic and Pareto. Sin embargo, todos estos modelos muestran algunos limitantes: (i) todas las distribuciones marginales univariadas tienen que pertenecer a la misma familia, mientras que las variables analizadas podrían mostrar diferentes marginales; (ii) la formulación matemática se complica cuando se aumenta el número de variables; (iii) no es posible distinguir el comportamiento marginal y conjunto de las variables estudiadas. (Grimaldi and Serinaldi, 2006b).

Para evaluar el desempeño de los modelos estocásticos asociados a los análisis de frecuencia, tradicionalmente se utilizan métodos univariados como el método máximo anual (AM) o el método pico-sobre-umbral (POT) para obtener la relación o las curvas entre Intensidad-duración-frecuencia (IDF). Para la construcción de las curvas IDF se utilizan enfoques univariados de las lluvias debido a su simplicidad matemática. (Singh and Zhang, 2007). La mayoría de las curvas IDF se construyen utilizando duraciones predeterminadas en intervalos de tiempo dados donde éstas no representan duraciones reales de los eventos de precipitación. Varios intentos se han hecho para obtener distribuciones conjuntas con las variables aleatorias intensidad, profundidad y duración. Los fenómenos de tormenta y las inundaciones son eventos multivariados y por lo tanto los análisis de frecuencia univariados sólo pueden proporcionar evaluaciones limitadas de estos eventos (Yue and Rasmussen, 2002).

Las curvas IDF son capaces de mostrar la relación matemática entre la intensidad de la precipitación (I), la duración (D) y el periodo de retorno (T, frecuencia anual de excedencia) Ariff et al. (2012). La construcción de las curvas IDF se deben realizar conjuntamente por medio de análisis de frecuencia bivariados a través del uso de la relación entre la intensidad de las tormentas y las duraciones utilizando las funciones cópula.

El enfoque de las cópulas es un método flexible que permite más opciones de distribuciones marginales y estructuras de dependencia que se utilizan en problemas multivariados. (Kao and Govindaraju, 2008). Las funciones cópula se basan en el teorema de Sklar (1959) que establece que para la distribución conjunta, el análisis de las marginales y la estructura de dependencia se pueden hacer por separado. Una explicación detallada de la teoría y la descripción de la cópula es dada por Nelsen (2006). Existen varias familias de cópulas y funciones disponibles para el análisis de frecuencia bivariado. Un ejemplo es la cópula arquimediana, la cual es muy utilizada debido a su maleabilidad matemática y por su simplicidad, tiene al menos 22 funciones cópula dentro de sus miembros. La elección de la familia de la cópula y su función se basa en la correlación de las variables aleatorias bajo consideración. (Ariff et al., 2012). Varios tipos de cópula se han utilizado en hidrología en la última década. Entre ellos se encuentran la cópula Farlie-Gumbel-Morgenstern (MGF) (Favre et al., 2004), las cópulas elípticas como la de Gauss (Renard and Lang, 2007), y las cópulas arquimedianas, (De Michele and Salvadori, 2003; De Michele et al., 2005; Salvadori and De Michele, 2004; Zhang and Singh, 2007a)

1.1. Generalidades

La historia de cópulas se puede decir que comienza con Fréchet y Féron en 1956, que hicieron importantes contribuciones en relación al problema de determinar la relación entre una función de distribución de probabilidad multivariada y sus distribuciones marginales de menor dimensión. (Erdely, 2009). Abe Sklar dio respuesta al problema propuesto por Fréchet para el caso de distribuciones marginales unidimensionales. Sklar (1959) obtuvo el resultado más profundo a este problema, mediante la introducción de la noción del término cópula y demostrando el teorema que lleva su nombre. Durante el periodo de 1958 a 1976, la mayoría de los resultados importantes en relación con las cópulas se obtuvieron en el curso del estudio de los espacios métricos probabilísticos. (Nelsen, 2006).

En 1990, Dall’Aglio organizó la primera conferencia dedicada a cópulas, llamada acertadamente “distribuciones de probabilidad con marginales dadas”. Este resultó ser el primero de una serie de conferencias que ayudaron en gran medida el desarrollo del campo, ya que cada una de ellas se ofreció la oportunidad de presentar los resultados y aprender de otros investigadores; estas conferencias se llevaron a cabo en Seattle en 1993, en Praga en 1996, en Barcelona en el año 2000, en Québec en 2004, en Tartu en 2007 y en Sao Paulo en 2010. (Jaworski et al., 2010).

Fisher (1997) dio dos razones principales por las cuales las cópulas son de interés para los estadísticos: “En primer lugar, como una manera de estudiar las medidas de escala libre de la dependencia; y en segundo lugar, como punto de partida para la construcción de familias de las distribuciones bivariadas.” Específicamente, las cópulas son una parte importante del estudio de la dependencia entre variables aleatorias, ya que permiten separar el efecto de la dependencia de los efectos de las distribuciones marginales.

El estudio de las funciones cópula ha ido incrementando en los últimos años con mayor frecuencia en la hidrología debido a que los procesos hidrológicos suelen ser multidimensionales. Hutchinson and Lai (1990) estuvieron entre los primeros autores que popularizaron el estudio de las cópulas. Nelsen (1999) presentó un tratamiento integral de las cópulas bivariadas, mientras que Joe (1997) dedica un capítulo de su libro “Multivariate models and dependence concepts” a las cópulas multivariadas. Las actualizaciones adicionales sobre cópulas se dan en Nelsen (2006). Los primeros trabajos en el análisis de frecuencia bivariado basados en el enfoque de las cópulas son los estudios de (Salvadori and De Michele, 2004; Favre et al., 2004). Seguido por otras contribuciones (De Michele et al., 2005; Genest et al., 2007; Kao and Govindaraju, 2008; Zhang and Singh, 2007a), además, se ha propuesto un análisis de frecuencia trivariado (Genest et al., 2007; Serinaldi and Grimaldi, 2007; Zhang and Singh,

2007b; Grimaldi and Serinaldi, 2006a). Ha habido una serie de intentos para llevar a cabo el análisis de frecuencia multivariado hidrológico (lluvias, inundaciones, sequías, la calidad del agua, etc.) teniendo en cuenta la dependencia entre variables correlacionadas (Zhang, 2005). Hasta hace poco, el análisis de frecuencia hidrológico multivariado se ha realizado utilizando el siguiente enfoque resumido por Zhang (2005):

1. La aplicación de la distribución normal multivariada, puede ser el enfoque más simple ya que es la distribución conjunta de variables correlacionadas. Si las marginales no tienen una distribución normal, las variables son convertidas por cierta transformación (es decir, la transformación de Box-Cox), que luego se convierten en una distribución normal.
2. La aplicación de la distribución Gumbel-Mixta, si marginales siguen la distribución Gumbel (Yue, 1999). Este método sólo se puede aplicar a las variables aleatorias correlacionadas positivamente.
3. La aplicación de la distribución bivariada logarítmica normal (Yue, 2000c), la distribución gamma bivariada (Yue and Rasmussen, 2002), y la distribución bivariada logística Gumbel (Yue and Rasmussen, 2002) para el análisis de frecuencia hidrológico de dos variables, donde las marginales tienen cada una el mismo tipo de distribución log-normal, gamma o la distribución de Gumbel.
4. El cambio de las variables por una cierta transformación de modo que las variables transformadas se vuelven independientes y luego se obtiene la distribución conjunta de las variables independientes (Stewardson and McMahon, 2002).
5. En la última década, el concepto cópula ha comenzado a aparecer en el análisis multivariado de frecuencia hidrológico (Favre et al., 2004).

En la literatura, se han hecho algunos intentos para hacer frente a los fenómenos hidrológicos multivariados. Ashkar (1980) considera un evento de inundación como un evento multivariado y deduce las relaciones entre pico de crecida, el volumen y la duración. Krstanovic and Singh (1987) derivan las distribuciones gaussianas y exponenciales multivariantes por el principio de máxima entropía. Sackl and Bergmann (1987), Loganathan et al. (1987), Chang et al. (1994), Goel et al. (1998) y Yue (1999, 2000c) utilizaron la distribución normal bivariada para representar las distribuciones conjuntas de inundaciones y tormentas. Los estudios anteriores no discutieron en detalle algunos conceptos clave, tales como distribuciones de probabilidad condicional, periodos de retorno condicional y períodos de retorno conjuntos, que son esenciales para la comprensión e interpretación de un evento multivariado. (Yue and Rasmussen, 2002).

1.2. Justificación

En el análisis de frecuencia univariado, los conceptos estadísticos como las probabilidades de no excedencia, la frecuencia de ocurrencia o periodo de retorno están bien definidos, sin embargo en la mayoría de los casos, los eventos hidrológicos están descritos por varias variables aleatorias correlacionadas entre sí, por ejemplo; las inundaciones se describen a través de su volumen, el pico y la duración, la intensidad de las lluvias se relaciona con la profundidad y duración de las tormentas. Eventos de dos o más variables deben ser descritos usando conceptos tales como distribuciones de probabilidad condicional, periodos de retorno condicional y periodos de retorno conjunto. (Yue and Rasmussen, 2002). Si un evento hidrológico dado es multivariado, es decir, descrito por un conjunto de variables aleatorias correlacionadas, entonces el análisis de frecuencia de una sola variable no puede proporcionar una evaluación completa de la probabilidad de ocurrencia. Una mejor comprensión de las características probabilísticas de tales eventos requiere el estudio de su distribución conjunta.

Justificaciones para la adopción del marco multivariado para el tratamiento de los fenómenos extremos se han discutido en varias referencias. En un análisis de frecuencia bivariado, Yue (1999) llegó a la conclusión de que con una sola variable hidrológica el análisis de frecuencia sólo puede proporcionar una evaluación limitada de los eventos extremos. Una mejor comprensión de las características probabilísticas de tales eventos requiere el estudio de su distribución conjunta. Esto también fue descrito por Shiau (2003) quien indica que el análisis de frecuencia multivariado requiere muchos más datos y análisis matemáticos más sofisticados. Zhang (2005) argumenta que el análisis univariado de frecuencia no puede representar las necesidades de diseños hidrológicos e hidráulicos y el riesgo correspondiente apropiadamente, el análisis multivariado es necesario con el fin de representar con mayor precisión el riesgo de fracaso. Yue and Rasmussen (2002), llegan a la conclusión general de que el análisis de frecuencia de una sola variable no puede proporcionar una evaluación probabilística suficiente de eventos hidrológicos multivariados correlacionados y pueden dar lugar a una sobreestimación de la gravedad de estos eventos. El análisis de frecuencia univariado puede ser útil cuando sólo una variable aleatoria es significativa para los propósitos del diseño o cuando dos variables aleatorias son menos dependientes. Sin embargo, un análisis separado de variables aleatorias no puede revelar la relación significativa entre ellas si la correlación es una información importante en los criterios de diseño. Por lo tanto, es de importancia considerar conjuntamente todas las variables aleatorias que caracterizan el evento hidrológico.

Las curvas IDF generalmente se obtienen a partir de enfoques univariados en donde sólo se tiene en cuenta la intensidad de la lluvia a intervalos de tiempo fijos. Las variables de la precipitación intensidad y duración están correlacionadas entre sí. Es por esto, que la

construcción de las curvas IDF se debe realizar conjuntamente por medio de análisis de frecuencia bivariados. El problema existente en estos enfoques se basa en el análisis univariado. Las curvas IDF combinan diferentes factores causales. Por ejemplo, el periodo de retorno de una estación 1 no es el mismo periodo de retorno de otra estación 2. Muchos problemas en ingeniería consideran que la recurrencia de los periodos de retorno suceden en un tiempo simultáneo, sin embargo, no hay garantía de que sucedan al mismo tiempo. ¿Cuáles son las intensidades que deben ocurrir para que se presente la tormenta con un periodo de retorno dado?. El enfoque univariado no puede representar realmente el análisis de frecuencia. Las curvas IDF deben resolverse conjuntamente. Esto se logra a través del uso de la relación entre la intensidad de las tormentas y las duraciones utilizando el método de cópula.

Por ejemplo, el análisis de frecuencia de crecidas cerca de la confluencia de varios afluentes, puede estar sujeto a inundaciones por altos flujos de alguna de las corrientes y es necesario para el diseño de muchas obras civiles. En tales casos, el enfoque de análisis de frecuencia de crecidas univariado no es aplicable. Un enfoque generalizado, basado en las funciones de probabilidad conjunta puede ser desarrollado para el análisis de frecuencia de crecidas en las confluencias.

1.3. Objetivos

1.3.1. Objetivo general

Realizar el análisis de frecuencia hidrológico multivariado de eventos extremos de naturaleza hidroclimatológica tales como tormentas intensas y caudales instantáneos máximos mediante la aplicación de funciones cópula arquimedianas para obtener curvas IDF (intensidad-duración-frecuencia) en la cuenca baja del río Tunjuelo y periodos de retorno conjunto en la región de la Mojana.

1.3.2. Objetivos específicos

- Construir curvas IDF por medio del método de la cópula para las estaciones pluviográficas pertenecientes a la cuenca baja del río Tunjuelo.
- Comparar las curvas IDF encontradas a través del método de la cópula con las curvas IDF construidas con el enfoque empírico convencional.
- Obtener la cópula arquimediana de mejor ajuste por medio del método gráfico a través de la estimación no paramétrica y mediante el estadístico Cramer-von Mises.
- Construir curvas IDF conjuntas multisitio a través de cópulas arquimedianas trivariadas y compararlas con el método empírico convencional.
- Obtener periodos de retorno conjunto para el análisis de frecuencia de crecidas por medio de cópulas arquimedianas multivariadas en la región de la Mojana.

1.4. Resumen del contenido

En el capítulo 2 se presentan las distribuciones de probabilidad más usadas en hidrología y las técnicas de selección de los modelos aplicados en el análisis de frecuencia hidrológico. En este capítulo también se presenta una revisión literaria e histórica de las curvas IDF empíricas (intensidad-duración-frecuencia) que generalmente se presentan como familias de curvas para distintos periodo de retorno. La representación de estas curvas se aplica en el área de estudio seleccionada y se construyen curvas IDF para las diferentes estaciones pluviográficas pertenecientes a la cuenca baja del río Tunjuelo.

En el capítulo 3 se presentan un marco teórico a la introducción de las cópulas bivariadas. Usando el teorema de Sklar (1959) se aclara el papel que juegan las cópulas en la relación entre las funciones de distribución multivariadas y sus marginales univariadas. En este capítulo

se presentan las cinco familias de cópulas arquimedianas estudiadas: Gumbel-Hougaard, Ali-Mikhail-Haq, Frank, Clayton y Joe, además se presenta la metodología para la selección de la cópula de mejor ajuste a los registros históricos. El periodo de retorno bivariado se define a través de la distribución de probabilidad conjunta y se presentan tres ejemplos de la aplicación de cópulas bivariadas en el área de estudio, el primero es para la construcción de curvas IDF por medio del método de la cópula, el segundo es para determinar el periodo de retorno conjunto multisitio y el tercero es para el análisis de series históricas. Se presentan las conclusiones respectivas al final del capítulo.

El capítulo 4, muestra cómo se puede extender el teorema de Sklar al caso multivariado y de esta forma obtener las cópulas arquimedianas multivariadas. Las mismas cinco familias de cópulas son estudiadas para el caso multivariado. El periodo de retorno trivariado se define a través de la distribución de probabilidad conjunta y se presentan dos ejemplos de la aplicación de cópulas trivariadas y multivariadas en el área de estudio, el primero es para obtener curvas IDF conjuntas multisitio para las estaciones pluviográficas pertenecientes a la cuenca baja del río Tunjuelo y el segundo es para determinar el periodo de retorno multivariado para análisis de frecuencia de crecidas en la región de la Mojana. Al final de este capítulo se presentan los análisis de resultados y sus respectivas conclusiones.

En el capítulo 5 se presentan las conclusiones generales y las recomendaciones obtenidas a partir de la investigación de este trabajo.

2. Funciones de distribución de probabilidad y análisis de frecuencia

El análisis de frecuencia hidrológico ha sido durante mucho tiempo, ampliamente aplicado para muchos propósitos en ingeniería y sobre todo en el diseño de obras civiles. El objetivo del análisis de frecuencia de información hidrológica es relacionar la magnitud de los eventos extremos con su frecuencia de ocurrencia mediante el uso de distribuciones de probabilidad. (Chow et al., 1994). De acuerdo con los datos seleccionados para el estudio de este trabajo (sección 2.4.2) se han empleado una serie de distribuciones de probabilidad que se mencionan en la tabla 2-2. Los parámetros de estas distribuciones por lo general han sido estimados por el método de los momentos (MOM), la estimación de máxima verosimilitud (MLE) y el método de los momentos lineal (LMM), entre otros. En este capítulo se presentan las distribuciones de probabilidad utilizadas y las técnicas de selección de los modelos aplicados en el análisis de frecuencia hidrológico. Una revisión literaria e histórica de varios estudios relacionados con análisis de frecuencia y distribuciones de probabilidad se encuentra en Zhang (2005).

2.1. Periodo de retorno univariado y posiciones de graficación

La definición tradicional del periodo de retorno de un evento con una magnitud dada, puede escribirse como el intervalo de recurrencia promedio entre un evento que iguala o excede una magnitud especificada. (Chow et al., 1994). Claramente tiene una base estadística ya que en la práctica de la ingeniería la elección del periodo de retorno (T_R) depende de la importancia de la estructura y las consecuencias de su fracaso. El periodo de retorno puede ser expresado mediante su relación con la probabilidad acumulada así:

$$T = \frac{1}{P(X > x)} = \frac{1}{1 - P(X \leq x)} = \frac{1}{1 - F(x)} \quad (2-1)$$

Posiciones de graficación

Usualmente, cuando se tienen datos de un cierto periodo, y se desea aplicar algún método estadístico para extrapolar dichos datos a periodos de retorno mayores al de las mediciones, es necesario asignar un valor del periodo de retorno a cada dato registrado (Mijares, 1992). Una posición de graficación se refiere al valor de la probabilidad asignada a cada uno de los datos que van a graficarse. Las fórmulas más usadas en análisis de frecuencia en hidrología para posiciones de graficación se presentan en la tabla 2-1 (Rao and Hamed, 2000).

Tabla 2-1.: Posiciones de graficación más usadas en hidrología

| Posición de graficación | Periodo de Retorno | Probabilidad acumulada ¹ |
|-------------------------|--------------------------|-------------------------------------|
| Weibull | $(n + 1)/m$ | $i/(n + 1)$ |
| Gringorten | $(n + 0,12)/(m - 0,44)$ | $(i - 0,44)/(n + 0,12)$ |
| Hazen | $n/(m - 0,5)$ | $(i - 0,5)/n$ |
| Blom | $(n + 0,25)/(m - 0,375)$ | $(i - 0,375)/(n + 0,25)$ |
| Cunnane | $(n + 0,2)/(m - 0,4)$ | $(i - 0,4)/(n + 0,2)$ |
| California | n/m | $(i - 1)/n$ |
| Chegodayev | $(n + 0,4)/(m - 0,3)$ | $(i - 0,3)/(n + 0,4)$ |
| Adamowski | $(n + 0,5)/(m - 0,24)$ | $(i - 0,26)/(n + 0,5)$ |

¹ Nota: i es el rango en orden ascendente como $i = n - m + 1$ donde m es el rango en orden descendente como $m = n - i + 1$

Tabla 2-2.: Distribuciones de probabilidad usadas en hidrología

| Distribución | Función de densidad de probabilidad (PDF) | Función de distribución de probabilidad acumulada (CDF) ¹ |
|------------------|--|--|
| Exponencial | $f(x) = \lambda e^{-\lambda x}$ | $F(x) = 1 - e^{-\lambda x}$ |
| Normal | $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ | $F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$ |
| LogNormal | $f(x) = \frac{1}{x\sigma_y\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu_y)^2}{2\sigma_y^2}\right)$ | $F(x) = \frac{1}{\sigma_y\sqrt{2\pi}} \int_0^x \frac{1}{x} \exp\left(-\frac{(\ln x - \mu_y)^2}{2\sigma_y^2}\right) dx$ |
| Gamma | $f(x) = \frac{1}{\alpha^\beta \Gamma(\beta)} x^{\beta-1} e^{-x/\alpha}$ | $F(x) = \frac{1}{\alpha^\beta \Gamma(\beta)} \int_0^x x^{\beta-1} e^{-x/\alpha} dx$ |
| Pearson tipo III | $f(x) = \frac{1}{\alpha \Gamma(\beta)} \left(\frac{x-\gamma}{\alpha}\right)^{\beta-1} \exp\left(-\frac{x-\gamma}{\alpha}\right)$ | $F(x) = \frac{1}{\alpha \Gamma(\beta)} \int_\gamma^x \left(\frac{x-\gamma}{\alpha}\right)^{\beta-1} \exp\left(-\frac{x-\gamma}{\alpha}\right) dx$ |
| Log-Pearson III | $f(x) = \frac{\lambda^\beta (y-\gamma)^{\beta-1} e^{-\lambda(y-\gamma)}}{x \Gamma(\beta)}$ | $F(x) = \frac{1}{\alpha \Gamma(\beta)} \int_0^x \frac{1}{x} \left(\frac{\ln x - \gamma}{\alpha}\right)^{\beta-1} \exp\left(-\frac{\ln x - \gamma}{\alpha}\right) dx$ |
| Gumbel | $f(x) = \frac{1}{\alpha} \exp\left[-\left(\frac{x-\beta}{\alpha}\right) - \exp\left(-\frac{x-\beta}{\alpha}\right)\right]$ | $F(x) = \exp\left[-\exp\left(-\frac{x-\beta}{\alpha}\right)\right]$ |

¹ En el anexo A se encuentran los rangos y las ecuaciones de los parámetros en términos de los momentos de la muestra

2.2. Pruebas de bondad de ajuste

Las pruebas de bondad de ajuste son usadas para la selección del modelo estadístico. En el análisis hidrológico, se suele encontrar que los criterios para determinar la bondad de las predicciones se basan en el examen de los valores del coeficiente de correlación y principalmente, del coeficiente de determinación R^2 (también denominado coeficiente de correlación múltiple al cuadrado o coeficiente de determinación múltiple). Las pruebas más usadas para la selección del modelo son: el coeficiente de determinación múltiple R^2 , el R^2 ajustado, el error cuadrático medio (MSE), la raíz del error cuadrático medio (RMSE) y criterios de información Bayesiana y Akaike (BIC y AIC) (Zhang, 2005).

En el caso de R^2 y R^2 ajustado, cuanto más grande son, mejor se ajusta el modelo dado. El R^2 ajustado tiene en cuenta el tamaño del conjunto de datos, y su valor es ligeramente inferior al de su correspondiente R^2 (Norusis, 1993). En el caso de MSE, RMSE, AIC y BIC, cuanto más pequeños son, mejor se ajusta al modelo. Para este estudio se utilizó la raíz del del error cuadrático medio (RMSE), el sesgo y la máxima verosimilitud (ML). Se aplicaron tanto a los modelos univariados como multivariados. El mejor modelo es el que tiene la RMSE más pequeña.

El error cuadrático medio (MSE) puede expresarse como:

$$MSE = E(x_c - x_0)^2 = \frac{1}{n-1} \sum (x_c(i) - x_0(i))^2 \quad (2-2)$$

Luego,

$$RMSE = \sqrt{MSE} \quad (2-3)$$

donde E es la esperanza, x_c indica el valor calculado, n es el tamaño de la muestra y x_0 es el valor observado.

La desviación o sesgo es una medida de la desviación de la cantidad estimada a partir de la observada (o el valor verdadero). El mejor modelo es el que tiene menor sesgo (Zhang, 2005). La desviación se expresa mediante:

$$Sesgo = \sum_{i=1}^n \frac{x_0(i) - x_c(i)}{x_0(i)} \quad (2-4)$$

donde $x_0(i)$ indica el valor observado de orden i y $x_c(i)$ indica el valor calculado de orden i . ML es el criterio que indica la falta de ajuste del modelo a los datos. La máxima verosimilitud de la distribución de probabilidad es obtenida, usando los parámetros estimados. Usando el

criterio ML, el mejor modelo es el que tiene el mayor valor de probabilidad.

El criterio AIC, desarrollado por Akaike (1974), consta de dos partes: la falta de ajuste de el modelo y la falta de fiabilidad del modelo debido al número de parámetros del modelo. El criterio AIC se puede expresar a través de dos enfoques: verosimilitud maximizada y MSE del modelo como:

$$AIC = -2 \log(\text{verosimilitud maximizada para el modelo}) + 2 (\text{No. de parámetros ajustados})$$

o también:

$$AIC = N \log(MSE) + 2 (\text{No. de parámetros ajustados})$$

Por lo tanto, el mejor modelo es el que tiene el valor mínimo AIC.

2.2.1. Prueba Kolmogorov-Smirnov

Esta prueba consiste en comparar el máximo valor absoluto de la diferencia D entre la función de distribución de probabilidad observada $F_0(x_m)$ y la estimada $F(x_m)$ (Mijares, 1992).

$$D = \text{máx} |F_0(x_m) - F(x_m)| \quad (2-5)$$

Con un valor crítico d que depende del número de datos y el nivel de significancia seleccionado. si $D < d$, se acepta la hipótesis nula. Esta prueba tiene la ventaja sobre la Chi-cuadrado χ^2 de que compara los datos con el modelo estadístico sin necesidad de agruparlos. La función de distribución de probabilidad observada se calcula como:

$$F_0(x_m) = 1 - \frac{m}{n+1} \quad (2-6)$$

donde m es el número de orden del dato x_m en una lista de mayor a menor y n es el número total de datos.

2.2.2. Prueba χ^2

La prueba χ^2 (Chi-cuadrado). Fue propuesta por Karl Pearson en 1900. Para aplicar la prueba, el primer paso es dividir los datos en un número k de intervalos de clase. Posteriormente se calcula el parámetro estadístico:

$$D = \sum_{i=1}^k (\theta_i - \epsilon_i)^2 / \epsilon_i \quad (2-7)$$

donde θ_i es el número observado de eventos en el intervalo i y ϵ_i es el número esperado de eventos en el mismo intervalo (Mijares, 1992).

ϵ_i se calcula como:

$$\epsilon_i = n [F(S_i) - F(I_i)] \quad i = 1, 2, \dots, k \quad (2-8)$$

donde $F(S_i)$ es la función de distribución de probabilidad en el límite superior del intervalo i , $F(I_i)$ es la misma función en el límite inferior y n es el número de eventos. Una vez calculado el parámetro D para cada función de distribución considerada, se determina el valor de una variable aleatoria con distribución χ^2 para $\nu = k - 1 - m$ grados de libertad y un nivel de significancia α , donde m es el número de parámetros estimados a partir de los datos. Para aceptar una función de distribución dada, se debe cumplir:

$$D \leq \chi_{1-\alpha, k-1-m}^2 \quad (2-9)$$

El valor de $\chi_{1-\alpha, k-1-m}^2$ se obtiene a partir de tablas de la función de distribución χ^2 ver Mijares (1992).

2.3. Tormentas de diseño

2.3.1. Antecedentes de las curvas IDF

La construcción de las curvas IDF tiene una larga historia, que se remonta al menos al artículo de Bernard (1932). Numerosos modelos de uso común de las curvas IDF han sido publicados en la literatura hidrológica; una visión histórica y una discusión se han incluido en Garcia-Bartual and Schneider (2001). Koutsoyiannis et al. (1998) han proporcionado una fórmula general para la relación IDF consistente con el fundamento teórico probabilístico para los eventos máximos de precipitación. Los modelos de fenómenos naturales extremos son de creciente interés y esta teoría ha sido ampliamente utilizada en la hidrología y otras ciencias ambientales. (Katz et al., 2002; Schliep et al., 2010; Van de Vyver, 2015b). En las últimas décadas, muchos autores que trabajan en las curvas IDF han modelado eventos anuales de precipitación máxima con la distribución de valores extremos generalizada (GEV) o la distribución de Gumbel, ver por ejemplo Demarée (1985), Mohymont et al. (2004), Muller et al. (2008) y Overeem et al. (2008) por mencionar algunos.

Las curvas IDF generalmente se presentan como familias de curvas para los diferentes periodos de retorno y son una de las herramientas más utilizadas en ingeniería de recursos hídricos. Un evento de tormenta se define mediante un valor de profundidad de precipitación

en un punto, mediante un hietograma de diseño que especifique la distribución temporal de la precipitación durante la tormenta, o mediante el denominado patrón IDF (Intensidad-Duración-Frecuencia de las lluvias extremas) (Pérez and Carballo, 2008). La razón por la cual las curvas IDF se obtienen a partir de análisis de frecuencia univariados de las lluvias, se debe a que es difícil derivar distribuciones conjuntas de características de la precipitación (intensidad o profundidad y duración) utilizando métodos estadísticos estándar. Varios intentos se han hecho para llevar a cabo el análisis de frecuencia de dos variables, teniendo en cuenta la dependencia entre las variables de precipitación (por ejemplo, la intensidad, cantidad y duración). Sin embargo, éstos han sido poco realistas bajo supuestos restrictivos (Singh and Zhang, 2007).

2.3.2. Relaciones intensidad-duración-frecuencia

- Una de las primeras relaciones empíricas de las curvas IDF se expresó por Bernard (1932) como:

$$R_D^T = \frac{a_0 T^{a_1}}{D^{a_2}} \quad (2-10)$$

donde R_D^T = intensidad de la precipitación (en mm/h) de duración D (en horas) y periodo de retorno T (en años) con a_0 , a_1 , y a_2 constantes. Para tormentas de corta duración (menores a 24 h) a_1 varia entre 0,18 a 0,26, el exponente a_2 de 0,7 a 0,85 (Aron et al., 1987).

- Una expresión empírica ampliamente utilizada entre la precipitación, la intensidad y la duración fue propuesta por Sherman (1931), que puede expresarse como:

$$R = \frac{a}{(D + b)^c} \quad (2-11)$$

donde R = intensidad de la precipitación media (en pulgadas por hora) para la duración D en minutos y un periodo de retorno dado. a , b y c son constantes. El parámetro a varia con el periodo de retorno y la localización geográfica, y los parámetros b y c varían únicamente con la localización geográfica.

- Por ejemplo, Wenzel (1982) dedujo, para algunas ciudades de los Estados Unidos, coeficientes para utilizarse en una ecuación de la forma:

$$i = \frac{c}{T_d^e + f} \quad (2-12)$$

donde i es la intensidad de la lluvia de diseño, T_d es la duración y c , e y f son coeficientes que varían con el lugar y el periodo de retorno. (Chow et al., 1994)

- Mijares (1992) menciona básicamente dos métodos con los que se puede determinar la relación entre las variables i , d y T para un sitio dado.

El primero, llamado *intensidad-periodo de retorno* relaciona estas dos variables para cada duración por separado mediante alguna de las funciones de distribución de probabilidad mencionadas al inicio de este capítulo. El segundo método relaciona simultáneamente las tres variables en una familia de curvas cuya ecuación es:

$$i = \frac{kT^m}{(d+c)^n} \quad (2-13)$$

donde k , m , n y c son constantes que se calculan mediante un análisis de correlación lineal múltiple (Mijares, 1992). Cuando los datos se agrupan lo suficiente en torno a líneas rectas, el valor de c puede tomarse como cero quedando así la ecuación:

$$i = \frac{kT^m}{d^n} \quad (2-14)$$

- El modelo propuesto por Chow et al. (1994) extiende la ecuación 2-12 incluyendo el periodo de retorno T utilizando la ecuación:

$$i = \frac{cT^m}{T_d + f} \quad (2-15)$$

o también:

$$i = \frac{cT^m}{D^e + f} \quad (2-16)$$

- El modelo propuesto por Koutsoyiannis et al. (1998) fue ajustado para Atenas Grecia, y considera que la variable intensidad i tiene una distribución doble exponencial o Gumbel, con el parámetro adimensional ψ constante e independiente de la duración, y el parámetro de escala varía con la duración d .

$$i = \lambda \left\{ \frac{\psi - \ln \left[-\ln \left(1 - \frac{1}{T} \right) \right]}{(d + \theta)^\eta} \right\} \quad (2-17)$$

donde i es la intensidad de la lluvia en mm/h, T es el periodo de retorno en años, d es la duración de la tormenta en horas y λ , ψ , η y θ son parámetros que varían con el lugar y el periodo de retorno, cuyos valores se estiman por aproximaciones sucesivas cuando se ajusta cada modelo.

Hay dos preocupaciones principales en cuanto a por qué el enfoque estándar conduce a la sobreestimación de la probabilidad de eventos desastrosos. La primera preocupación está relacionada con la falta de disponibilidad de la cantidad y calidad de los datos históricos en el desarrollo de un modelo adecuado. Los métodos clásicos se basan en la distribución de muestreo, que requiere datos suficientemente “grandes” para determinar la exactitud del modelo. La segunda preocupación se relaciona con la incertidumbre de los parámetros, que a menudo se pasa por alto o es ignorado la mayor parte de las veces. Es importante cuantificar estas incertidumbres y utilizarlas en las decisiones de diseño y análisis de riesgos. El enfoque bayesiano proporciona un marco coherente para la incorporación de estas incertidumbres (Chandra et al., 2015).

2.4. Representación de las curvas IDF en el área de estudio

2.4.1. Localización general del área de estudio

Para la selección del área de estudio se tuvieron en cuenta los siguientes aspectos: (1) una cuenca que contara con más de tres estaciones pluviográficas, (2) datos representativos con series históricas mayores de 20 años y (3) una cuenca de uso urbano en la cual se pudieran evaluar las temáticas como la oferta, la demanda y el riesgo. Es por esto que se escogió la cuenca del río Tunjuelo ya que es de gran importancia para el abastecimiento de agua para la zona sur de Bogotá.

La cuenca del río Tunjuelo se encuentra ubicada en el departamento de Cundinamarca y forma parte de la cuenca alta y media del río Bogotá (ver figura 2-1). El cauce principal del río nace en el páramo de Sumapaz, a una altura aproximada de 3450 msnm. Lo forman tres cauces principales, los ríos Mugroso, Chisacá y Curubital, que, al unirse, forman el cauce principal del río Tunjuelo, que entrega sus aguas al río Bogotá a una altura de 2543 msnm (EAAB-ESP, 2014).

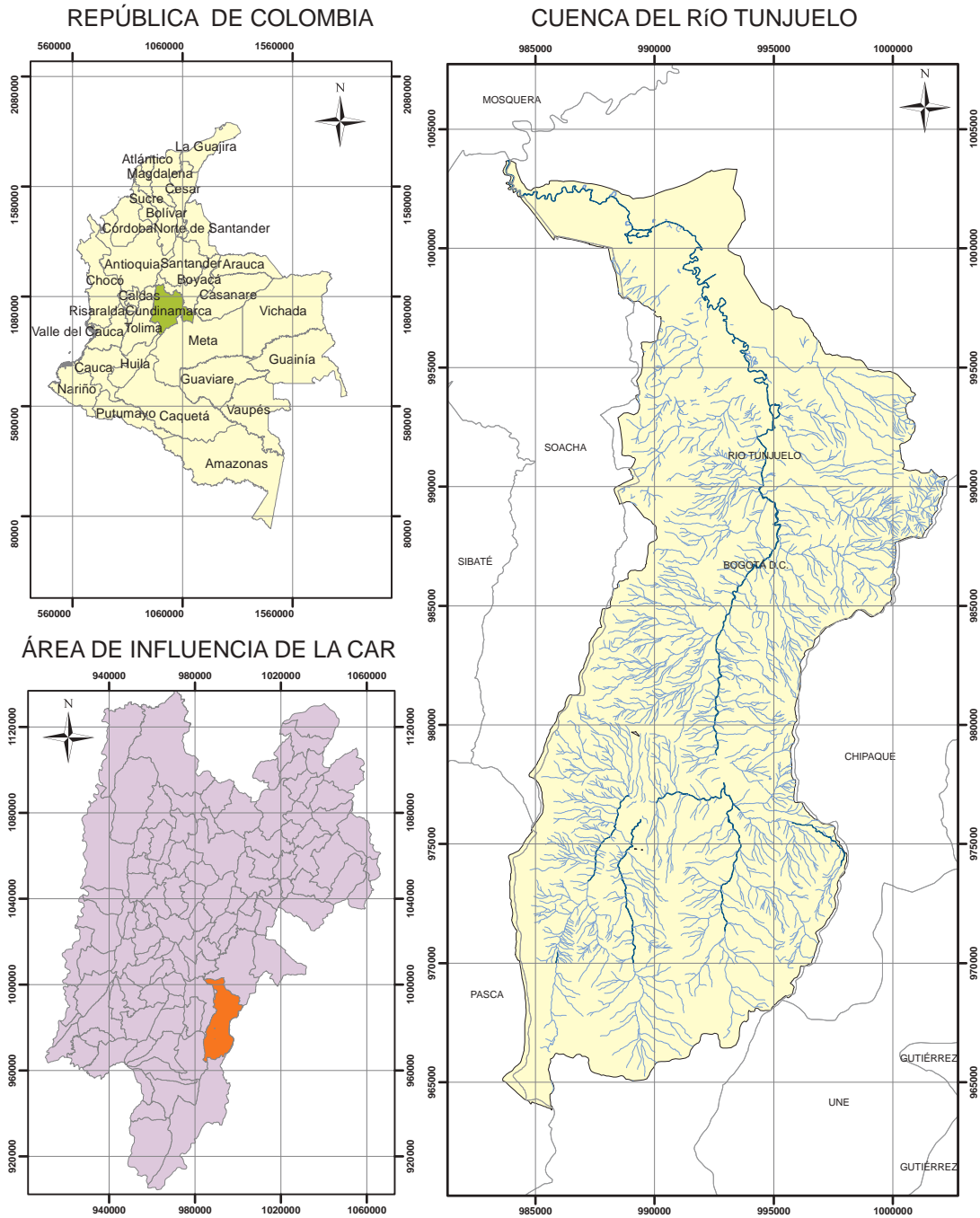


Figura 2-1.: Cuenca del río Tunjuelo en el área de jurisdicción de la CAR.

2.4.2. Recopilación de información y estaciones seleccionadas

La serie a la cual se le realiza el análisis de frecuencias puede ser seleccionada de tres formas (Acevedo Aristizábal et al., 2009):

- Series máximas anuales, es aquella serie que está conformada por cada uno de los valores máximos de precipitación observados en cada uno de los años de registro.
- Series de duraciones parciales, en la cual la serie de datos está conformada por aquellos datos que sobrepasan un valor base predefinido.
- Serie de excedencias anuales, en la cual el valor base, de la serie anterior, es seleccionado de tal forma que el número de datos que conforman la serie sea igual al número de años de registros.

La información recopilada de las estaciones seleccionadas en el área objeto de estudio corresponde a series históricas de precipitación sub-horaria extraídas de EAAB-ESP (2014). El tipo de serie adoptada es la de series máximas anuales ya que se cuentan con los registros del evento máximo de la tormenta. Las estaciones seleccionadas se encuentran ubicadas en la cuenca baja del río Tunjuelo cuyas características se encuentran en la tabla **2-3**:

Tabla 2-3.: Estaciones pluviográficas seleccionadas para la construcción de curvas IDF

| Nombre | Código | Tipo | Norte | Este | Elev. | Entidad | Periodo |
|----------------|---------|------|---------|---------|-------|---------|-------------|
| El Bosque | 2120085 | PG | 986500 | 1000000 | 2880 | CAR | 1967 - 2012 |
| La Picota | 2120156 | PG | 994720 | 995620 | 2580 | CAR | 1980 - 2003 |
| El Delirio | P-35 | PG | 994730 | 1002120 | 3000 | EAAB | 1961 - 2012 |
| Santa Lucía | P-42 | PG | 997550 | 995080 | 2630 | EAAB | 1956 - 2012 |
| Bosa Barreno 2 | P-51 | PG | 1001915 | 988072 | 2550 | EAAB | 1986 - 2012 |
| Juan Rey | P-81 | PG | 991780 | 999260 | 2985 | EAAB | 1990 - 2012 |
| Quiba | P-90 | PG | 992305 | 989998 | 3000 | EAAB | 1990 - 2012 |

La información disponible, cuenta con eventos de precipitación con una resolución mínima de cinco minutos. Para la generación de las curvas IDF se definieron intervalos de [15, 30, 45, 60, 90, 120, 150, 180, 240, 300, 360] minutos.

Las curvas IDF resultan de unir puntos representativos de la intensidad media en intervalos de diferente duración y correspondientes todos ellos a una misma frecuencia o periodo de retorno. Su obtención directa sólo es posible en las estaciones dotadas de pluviógrafo (Peláez, 1978).

La metodología tradicional usada para el cálculo de las curvas IDF consiste básicamente en realizar un análisis de frecuencia a cada una de las series de valores máximos de precipitación obtenidas para cada duración.

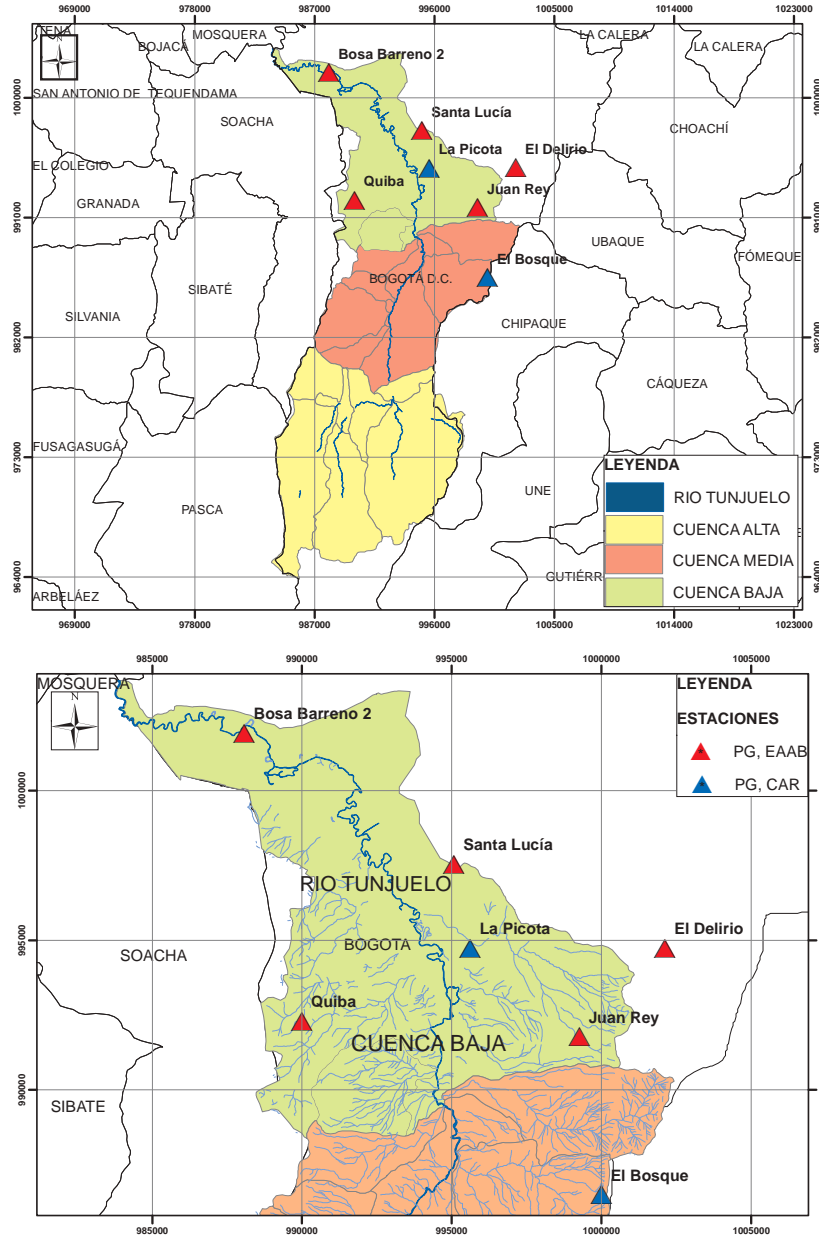


Figura 2-2.: Ubicación de las estaciones hidrometeorológicas seleccionadas.

Para ilustrar la construcción de estas curvas, se presenta a continuación un ejemplo para la estación El Bosque, el cual se comparará más adelante con las curvas obtenidas mediante el

método de la cópula.

Inicialmente, se grafican los datos construidos de una serie histórica (1967 - 2012) de precipitación anual máxima para cada una de las duraciones establecidas: [15, 30, 45, 60, 90, 120, 150, 180, 240, 300, 360] minutos.

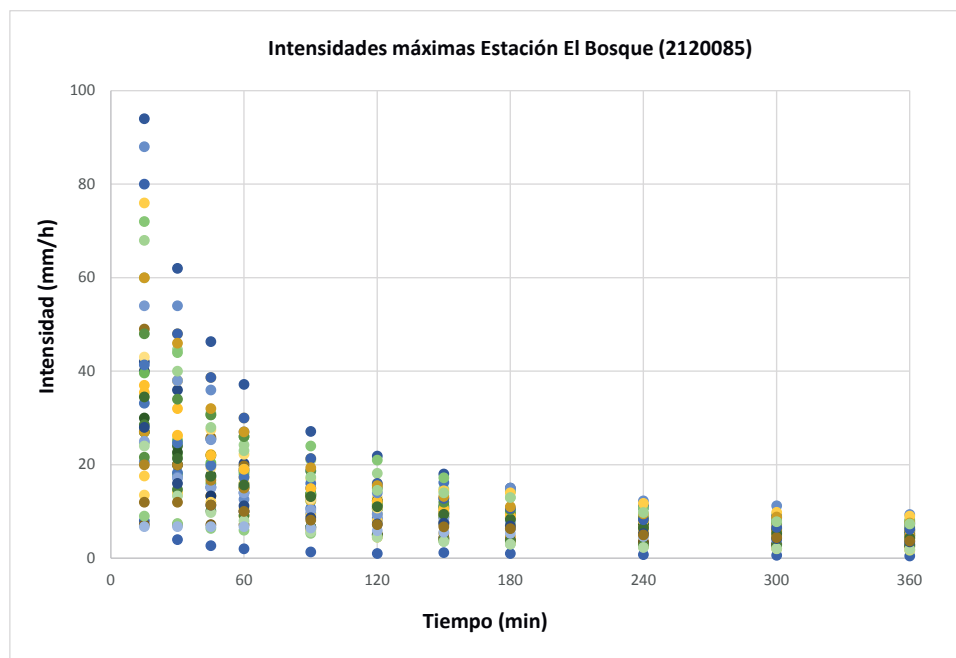


Figura 2-3.: Intensidad para diferentes eventos de tormenta máxima seleccionada de una serie histórica para una duración dada. Estación El Bosque.

Asociando un periodo de retorno a cada uno de los datos de intensidad y duración, se puede construir una representación gráfica en tres dimensiones. Mediante las posiciones de graficación mencionadas en la tabla **2-1** se asigna un periodo de retorno a cada uno de los datos. Realizando una prueba de bondad de ajuste (sección 2.2) a cada una de las posiciones de graficación de la tabla **2-1** se encontró que la distribución empírica de mejor ajuste es la indicada por Gringorten:

$$T = \frac{n + 0,12}{m - 0,44}$$

donde n es el número de datos que van a ser graficados, y m es el orden o posición de un valor en una lista ordenada por magnitud descendente. Aplicando la ecuación de Gringorten se obtiene la siguiente gráfica:

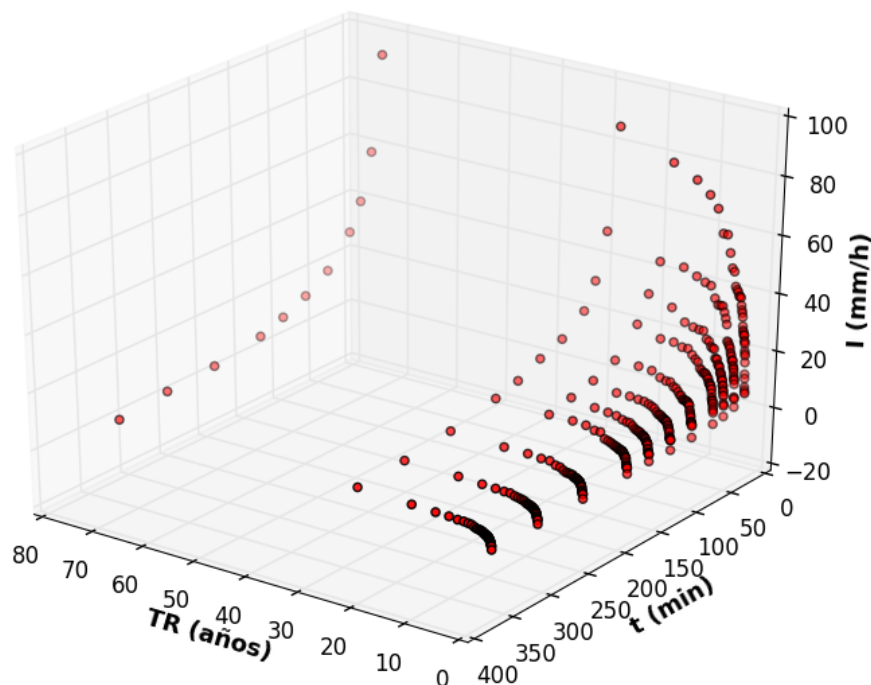


Figura 2-4.: Periodo de retorno asociado a cada uno de los registros tomados en la estación El Bosque.

Los periodos de retorno asociados a cada registro, llegan hasta un valor cercano a 80 años. En el Anexo A se encuentran los registros para la estación el Bosque, donde se presenta el código para la obtención de los puntos graficados en la figura 2-4.

Una vez obtenidos los periodos de retorno, es necesario ajustar un método estadístico o de regresión para obtener las familias de curvas que mejor se ajusten a los datos. Como se vio en la sección 2.3.2, las relaciones intensidad-duración-frecuencia se presentan por medio de una serie de ecuaciones empíricas, las cuales intentan representar estas familias de curvas. Por medio del lenguaje de programación Python (2015) se realizó una regresión no lineal a los datos representados en la figura 2-4. La ecuación de mejor ajuste y que presenta menor error es la correspondiente al modelo propuesto por Chow et al. (1994). (ver ecuación 2-16).

$$i = \frac{cT^m}{D^e + f} = \frac{285,37 T^{0,34}}{D^{0,755} + 3,024} \quad (2-18)$$

La superficie resultante de esta ecuación es la que se presenta en la figura 2-5.

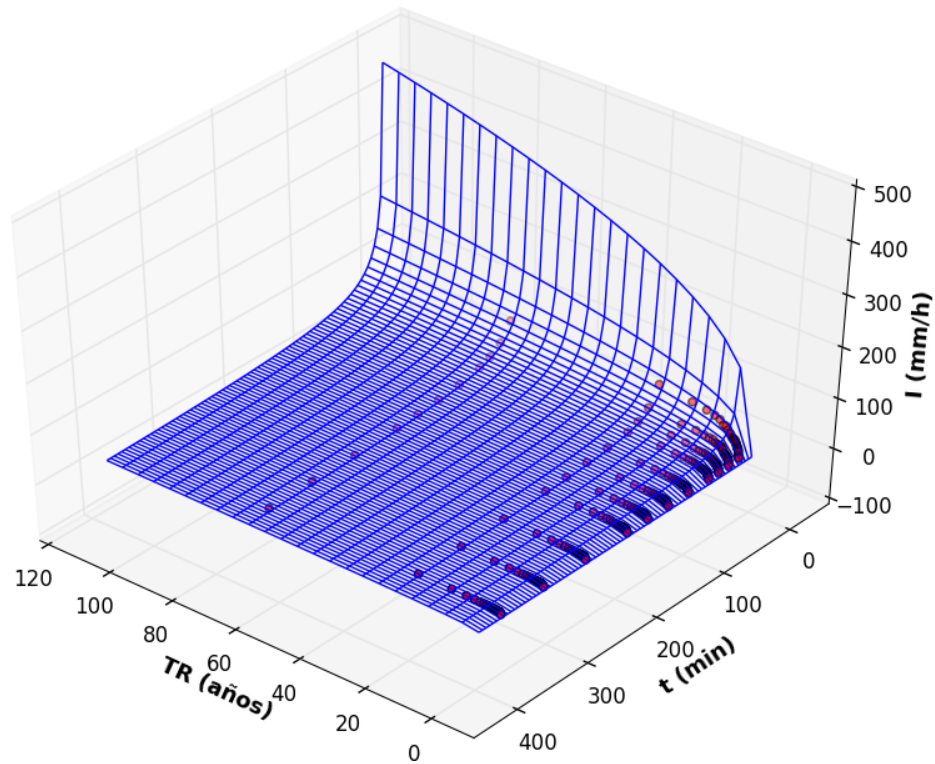


Figura 2-5.: Superficie de regresión generada a partir de la ecuación 2-18 para la estación El Bosque.

Si se dibujan planos que pasen por el eje del periodo de retorno, éstos se intersectan con la superficie creada generando familias de curvas IDF, es decir; las trazas de la superficie son las curvas IDF. Las familias de curvas IDF se presentan en la figura 2-6.

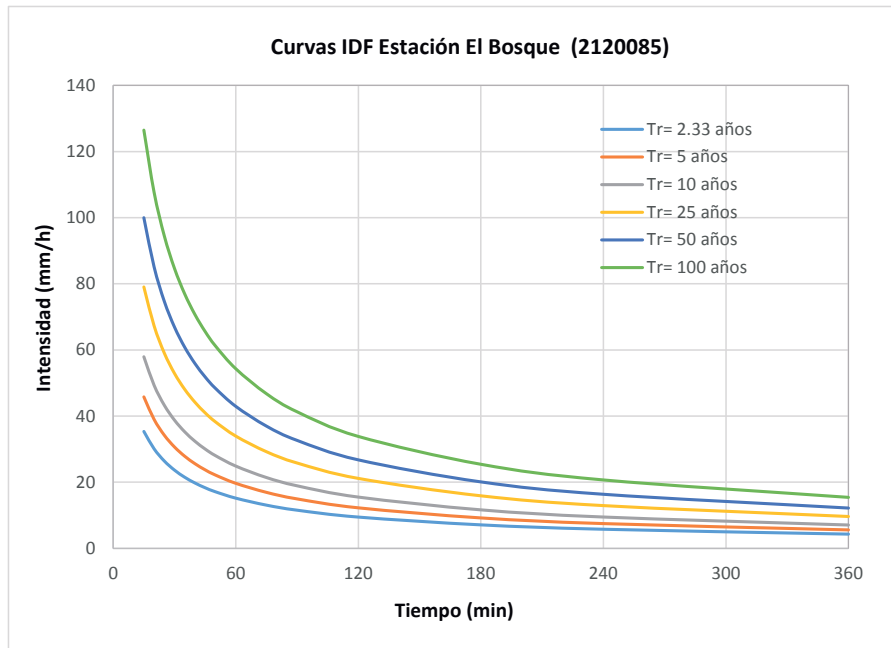


Figura 2-6.: Curvas IDF generadas a partir de los registros históricos de la estación El Bosque.

Distribuciones de probabilidad seleccionadas

Las pruebas de bondad de ajuste son necesarias para saber qué tan bien se ajusta una distribución de probabilidad a los registros de las estaciones seleccionadas. A continuación se presenta un resumen de las distribuciones seleccionadas para este trabajo.

Inicialmente se graficó el histograma de frecuencia con las diferentes distribuciones de probabilidad asociadas. (ver figura 2-7).

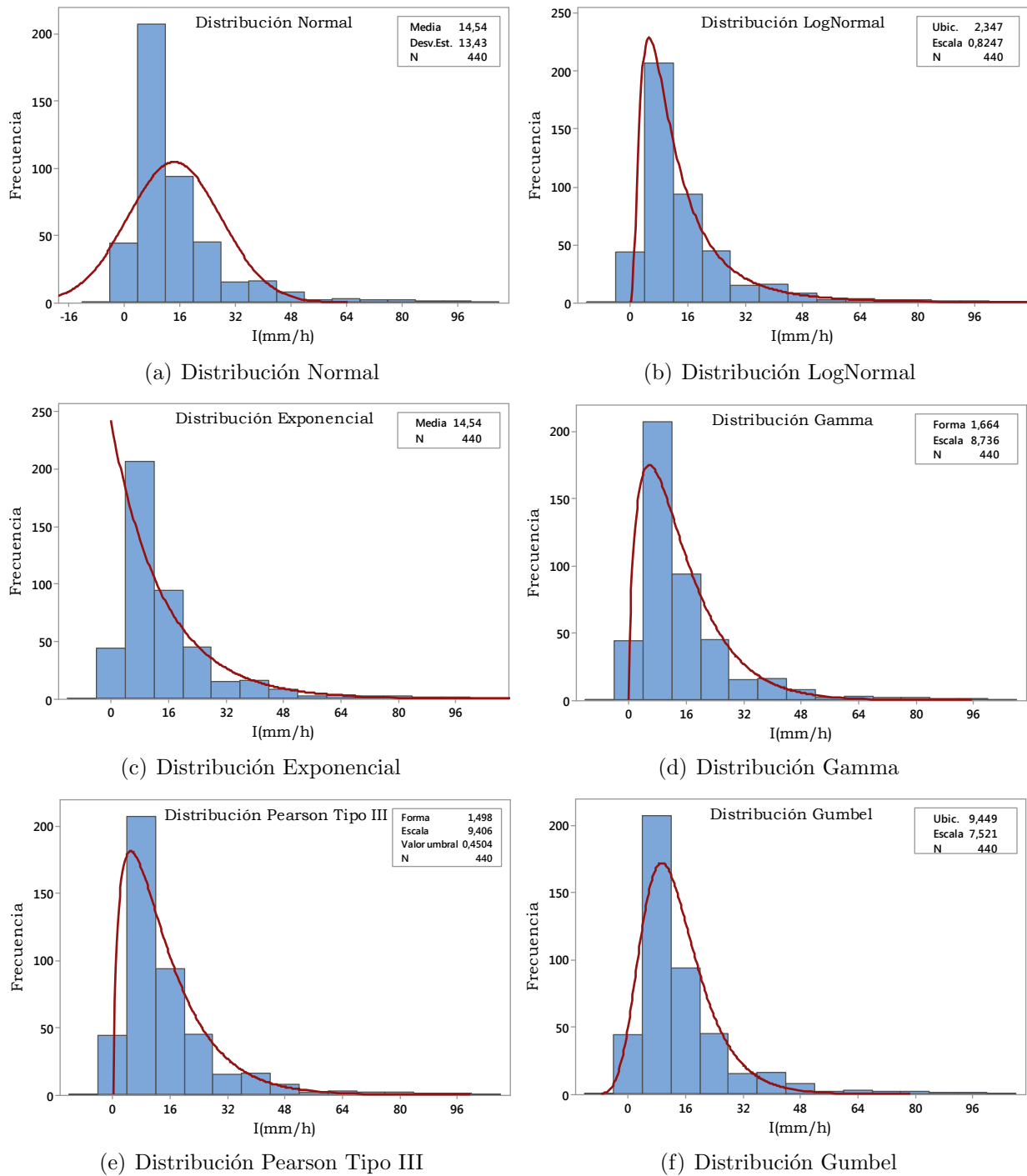


Figura 2-7.: Diferentes distribuciones de probabilidad asociadas a los registros de intensidad (mm/h) para la estación El Bosque.

Como se puede observar en la figura 2-7, la distribución que más se ajusta a los datos es la distribución LogNormal. La prueba no paramétrica Anderson-Darling propuesta por el software estadístico Minitab versión académica, (Minitab, 2016) muestra que la distribución de mejor ajuste es la LogNormal. Mientras mejor se ajuste la distribución a los datos, menor será este estadístico (ver figura 2-8).

Gráfica de probabilidad para I(mm/h)

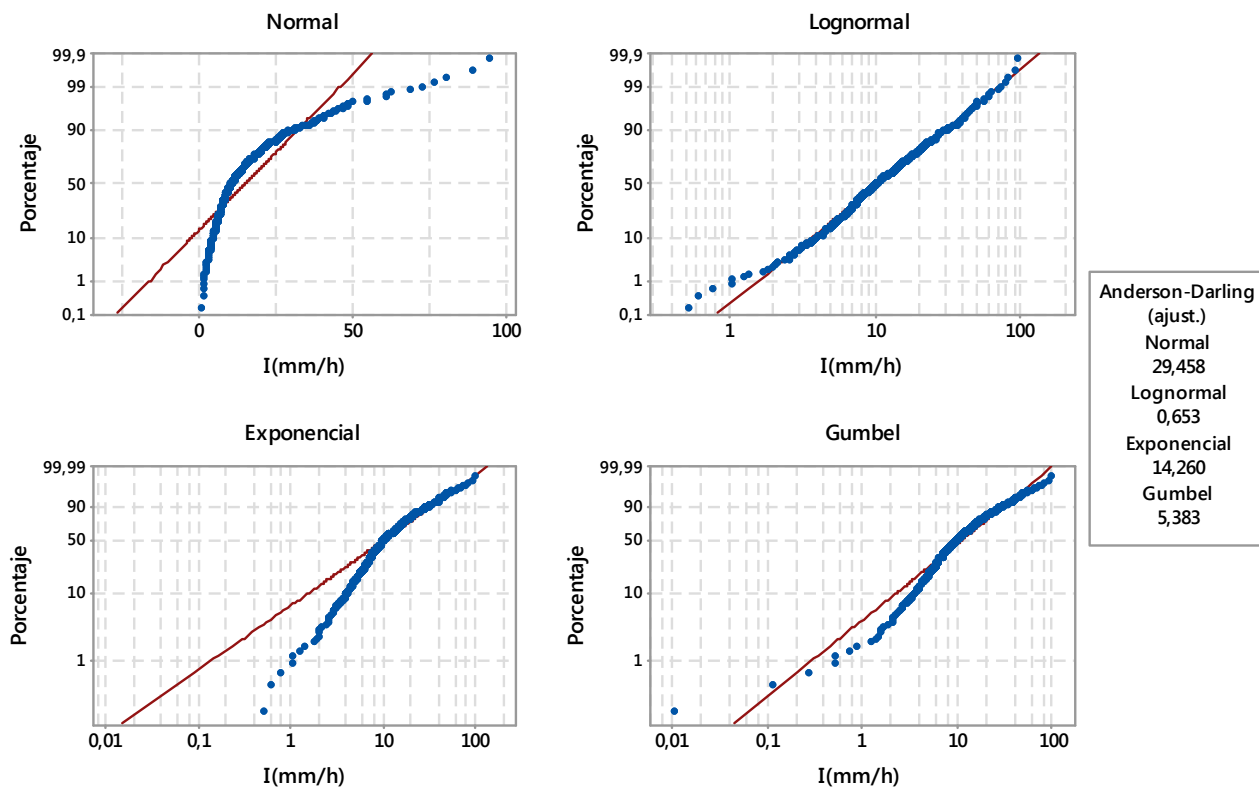


Figura 2-8.: Ajuste mediante el estadístico Anderson-Darling. (Minitab, 2016).

Sin embargo, para tener más seguridad en el ajuste, se realizaron las pruebas de bondad de ajuste Kolmogorov-Smirnov y χ^2 mencionadas en la sección 2.2.

Igualmente, para los datos de duración se realizó el mismo procedimiento obteniendo que la distribución LogNormal es la que mejor se ajusta al conjunto de datos (ver figura 2-9).

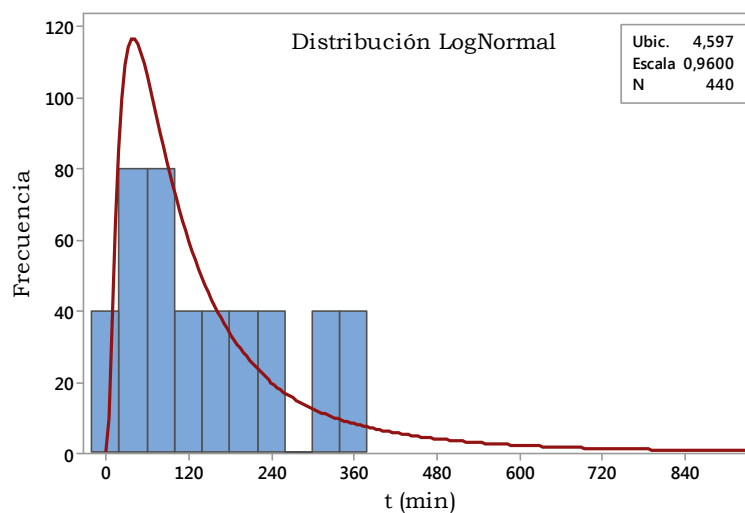


Figura 2-9.: Histograma de frecuencia ajustado para la duración de la tormenta en minutos. (Estación El Bosque).

Las distribuciones de probabilidad seleccionadas mediante las pruebas Kolmogorov-Smirnov y χ^2 se presentan en la tabla 2-4.

Tabla 2-4.: Distribuciones de probabilidad marginales asociadas a cada una de las estaciones seleccionadas

| Estación | I(mm/h) | P(mm) | t(min) |
|----------------|------------------|------------------|-----------|
| El Bosque | LogNormal | LogNormal | Lognormal |
| La Picota | Pearson Tipo III | Pearson Tipo III | Lognormal |
| El Delirio | Pearson Tipo III | Pearson Tipo III | Lognormal |
| Santa Lucía | LogNormal | LogNormal | Lognormal |
| Bosa Barreno 2 | Pearson Tipo III | Pearson Tipo III | Lognormal |
| Juan Rey | LogNormal | LogNormal | Lognormal |
| Quiba | LogNormal | LogNormal | Lognormal |

3. Cópulas Bivariadas

Una cópula es una función que une o enlaza (su nombre proviene del latín “enlazar”) una función de distribución multivariada a sus funciones de distribución marginales unidimensionales. La palabra cópula fue utilizada por primera vez en un sentido matemático por Sklar (1959) en el teorema que lleva su nombre. Las funciones cópula permiten representar una distribución multivariada basada en la distribución de probabilidad univariada (o simplemente llamadas marginales), independientes de su forma o tipo. A continuación se discute la definición y algunas propiedades de las cópulas.

3.1. Introducción a las cópulas

Para comenzar con la introducción a las cópulas, es necesario presentar algunas notaciones. Un rectángulo en $\overline{\mathbb{R}}^2$, ($\overline{\mathbb{R}} = [-\infty, \infty]$) es el producto cartesiano de dos intervalos cerrados: $B = [x_1, x_2] \times [y_1, y_2]$. Los vértices del rectángulo B son los puntos (x_1, y_1) , (x_1, y_2) , (x_2, y_1) y (x_2, y_2) . Una función real bivariada H es una función cuyo dominio, $DomH$ es un subconjunto de $\overline{\mathbb{R}}^2$ y cuyo rango $RanH$ es un subconjunto de $\overline{\mathbb{R}}$ (Moreno Chavarro, 2012).

Definición 1. Sean S_1 y S_2 subconjuntos no vacíos de $\overline{\mathbb{R}} = [-\infty, \infty]$, sea H una función real tal que $DomH = S_1 \times S_2$, $B = [x_1, x_2] \times [y_1, y_2]$ un rectángulo cuyos vertices están en $DomH$, entonces el H-volumen de B está definido por:

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1) \quad (3-1)$$

Definición 2. La función H dada en la definición 1 es creciente si $V_H(B) \geq 0$ para todo B , cuyos vértices están en el dominio de H .

El siguiente lema es de gran utilidad para establecer la continuidad de las cópulas y es consecuencia directa de las definiciones anteriores (Nelsen, 2006).

Lema 1. Sean S_1 y S_2 subconjuntos no vacíos de $\overline{\mathbb{R}}$ y sea H una función creciente bivariada con dominio $S_1 \times S_2$. Sean x_1, x_2 en S_1 con $x_1 \leq x_2$, y sean y_1, y_2 en S_2 con $y_1 \leq y_2$. Entonces la función $t \mapsto H(t, y_2) - H(t, y_1)$ es no decreciente sobre S_1 y la función

$t \mapsto H(x_2, t) - H(x_1, t)$ es no decreciente sobre S_2 .

Como una aplicación inmediata a este lema se pueden mostrar las siguientes propiedades:

Suponga que S_1 tiene un mínimo a_1 y que S_2 tiene un mínimo a_2 . Decimos que una función H de $S_1 \times S_2$ en \mathbb{R} es *fundamentada* si $H(x, a_2) = 0 = H(a_1, y)$ para todo (x, y) en $S_1 \times S_2$.

Ahora suponga que S_1 tiene un máximo b_1 y que S_2 tiene un máximo b_2 . Decimos que una función H de $S_1 \times S_2$ en \mathbb{R} *tiene marginales* y estas marginales son las funciones F y G dadas por:

$$\begin{aligned} \text{Dom}F &= S_1 \text{ y } F(x) = H(x, b_2) \text{ para todo } x \in S_1; \\ \text{Dom}G &= S_2 \text{ y } G(y) = H(b_1, y) \text{ para todo } y \in S_2. \end{aligned}$$

Ejemplo 1. (Nelsen, 2006). Sea H una función con dominio $[-1, 1] \times [0, \infty]$, dada por:

$$H(x, y) = \frac{(x+1)(e^y - 1)}{x + 2e^y - 1}$$

Entonces H es fundamentada ya que:

$$H(x, 0) = \frac{(x+1)(e^0 - 1)}{x + 2e^0 - 1} = 0, \quad H(-1, y) = \frac{(-1+1)(e^y - 1)}{-1 + 2e^y - 1} = 0$$

Con marginales:

$$F(x) = H(x, \infty) = \frac{x+1}{2}$$

$$G(y) = H(1, y) = 1 - e^{-y}$$

3.2. Definición de Cópula

Una cópula es una función C de I^2 a I , con $I = [0, 1]$ es decir; $C : [0, 1] \times [0, 1] \rightarrow [0, 1]$ que satisface las siguiente condiciones:

- Para todo $u, v \in [0, 1]$,
 $C(u, 0) = 0 = C(0, v)$
 y
 $C(u, 1) = u$ y $C(1, v) = v$;

- Para todo $u_1, u_2, v_1, v_2 \in [0, 1]$ tal que $u_1 \leq u_2$ y $v_1 \leq v_2$,

$$V_C([u_1, u_2] \times [v_1, v_2]) = C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$$

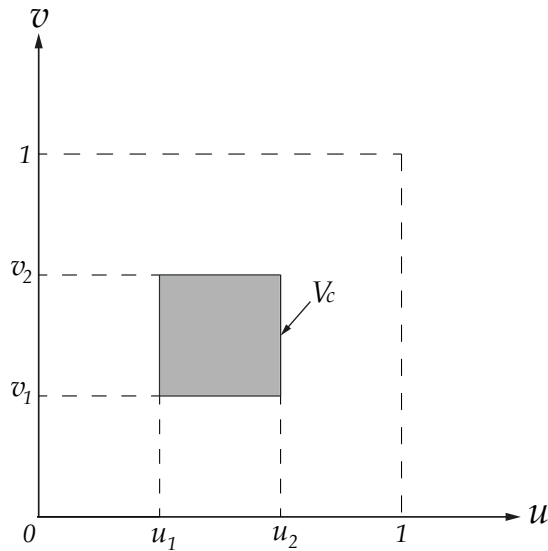


Figura 3-1.: Esquema de la propiedad de la cópula.

3.3. Teorema de Sklar

El teorema de Sklar aclara el papel que juegan las cópulas en la relación entre las funciones de distribución multivariadas y sus marginales univariadas.

Teorema 1. *Sea $H(x, y)$ una función de distribución bivariada con marginales $F(x)$ y $G(y)$ entonces existe una cópula C tal que para todo $x, y \in \overline{\mathbb{R}}$:*

$$H(x, y) = C(F(x), G(y)) \tag{3-2}$$

Si $F(x)$ y $G(y)$ son continuas, la cópula $C(u, v)$ es única. Dicho de otro modo, $C(u, v)$ queda determinada de forma única en $\text{Ran}F \times \text{Ran}G$, con $\text{Ran}F$ y $\text{Ran}G$ en rango de F y G respectivamente. Inversamente, si C es una cópula y F, G son funciones de distribución, entonces H es una función de distribución conjunta con marginales F y G . (Nelsen, 1999)

3.3.1. Densidad de la cópula

Si $F(x)$, $G(y)$ y la cópula $C(u, v)$ son diferenciables, la densidad conjunta de (x, y) correspondiente a la función de distribución conjunta de la ecuación 3-2 es igual a:

$$h(x, y) = f(x)g(y)[C(F(x), G(y))] \quad (3-3)$$

donde $f(x)$, $g(y)$ son las funciones de densidad marginales de $F(x)$ y $G(y)$ y

$$c(u, v) = \frac{\partial C(u, v)}{\partial u \partial v} \quad (u, v)^T \in [0, 1]^2 \quad (3-4)$$

Esta última función se conoce como *densidad de la cópula* (Moreno Chavarro, 2012).

3.4. Medidas de correlación de rango

Hay varias formas para medir la dependencia de los datos o registros históricos de eventos hidrológicos. Como veremos, muchas de estas propiedades y medidas son, en palabras de Hoeffding (1940, 1941), “de escala invariante”, es decir, que no se modifican bajo transformaciones estrictamente crecientes. Es precisamente la cópula la que capta las propiedades de la distribución conjunta que son invariantes bajo estas transformaciones (Schweizer and Wolff, 1981). Los coeficientes de correlación de rango miden el grado de dependencia monótona entre dos variables aleatorias, a continuación se presentan dos importantes medidas de dependencia; el Tau de Kendall (τ) y ρ -Spearman, pero antes es necesario realizar algunas definiciones:

Definición 3. Concordancia: Sean (x_i, y_i) y (x_j, y_j) dos observaciones de un vector de variables aleatorias continuas (X, Y) se dice que (x_i, y_i) y (x_j, y_j) son **concordantes** si $x_i < x_j$ y $y_i < y_j$ o $x_i > x_j$ y $y_i > y_j$. Una forma alterna de presentar la anterior definición es: (x_i, y_i) y (x_j, y_j) son concordantes si $(x_i - x_j)(y_i - y_j) > 0$ (Moreno Chavarro, 2012).

Definición 4. Discordancia: Sean (x_i, y_i) y (x_j, y_j) dos observaciones de un vector de variables aleatorias continuas (X, Y) se dice que (x_i, y_i) y (x_j, y_j) son **discordantes** si $x_i < x_j$ y $y_i > y_j$ o $x_i > x_j$ y $y_i < y_j$. Una forma alterna de presentar la anterior definición es: (x_i, y_i) y (x_j, y_j) son discordantes si $(x_i - x_j)(y_i - y_j) < 0$.

3.4.1. Tau de Kendall τ

El τ de Kendall se define en términos de concordancia de la siguiente manera (Kruskal, 1958; Hollander and Wolfe, 1973; Lehmann and D'abrera, 1975):

Sea $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ una muestra aleatoria de n observaciones de un vector (X, Y) de variables aleatorias continuas. Hay $\binom{n}{2}$ parejas distintas (x_i, y_i) y (x_j, y_j) de observaciones en la muestra. Algunas de ellas serán concordantes, denotadas por c y otras serán discordantes, denotadas por d . Entonces, el Tau de Kendall τ para la muestra se define como:

$$t = \frac{c - d}{c + d} = (c - d) / \binom{n}{2} \quad (3-5)$$

De manera equivalente, t es la probabilidad de concordancia menos la probabilidad de discordancia para un par de observaciones (x_i, y_i) y (x_j, y_j) escogidas al azar de la muestra. La versión poblacional del τ de Kendall para un vector (X, Y) de variables aleatorias continuas con función de distribución conjunta H se define similarmente.

Sean (X_1, Y_1) y (X_2, Y_2) vectores aleatorios independientes e igualmente distribuidos en \mathbb{R}^2 con función de distribución conjunta H , entonces el coeficiente de correlación de rango o τ de Kendall está definido como la probabilidad de concordancia menos la probabilidad de discordancia (Nelsen, 2006):

$$\tau = \tau_{X,Y} = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] \quad (3-6)$$

Como τ de Kendall es invariante a transformaciones estrictamente crecientes, el siguiente teorema ofrece una expresión de este coeficiente en términos de cópulas.

Teorema 2. *Sean X e Y variables aleatorias continuas cuya cópula es C . Entonces el coeficiente τ de Kendall para X e Y está dado por:*

$$\tau_{X,Y} = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1 \quad (3-7)$$

3.4.2. Rho de Spearman ρ

Al igual que el τ de Kendall, la medida de asociación conocida como ρ de Spearman se basa en concordancia y discordancia y se obtiene de la siguiente manera (Kruskal, 1958; Lehmann and D'abrera, 1975):

Sean (X_1, Y_1) , (X_2, Y_2) y (X_3, Y_3) tres vectores aleatorios independientes con una función de distribución conjunta H (cuyas marginales son otra vez F y G) y Cópula C . Se define el ρ de Spearman como proporcional a la probabilidad de concordancia menos la probabilidad de discordancia para los dos vectores (X_1, Y_1) y (X_2, Y_2) es decir, un par de vectores con

las mismas marginales, pero un vector tiene función de distribución H mientras que los componentes del otro son independientes.

$$\rho = \rho_{X,Y} = 3(P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0]) \quad (3-8)$$

El siguiente teorema ofrece una expresión del coeficiente de correlación de Spearman en términos de cópulas.

Lema 2. Si (X, Y) tiene función de distribución conjunta H y marginales $F(x)$ y $G(y)$, entonces la covarianza entre X e Y $cov(X, Y)$, cuando es finita está dada por:

$$cov(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (H(x, y) - F(x)G(y)) dx dy \quad (3-9)$$

Teorema 3. Sean X e Y variables aleatorias absolutamente continuas con marginales $F(x)$ y $G(y)$ respectivamente, y cópula asociada C entonces el coeficiente de correlación Spearman para X e Y $\rho_{X,Y}$, está dado por:

$$\rho_{X,Y} = 12 \int_0^1 \int_0^1 C(u, v) du dv - 3 \quad (3-10)$$

3.5. Familias de Cópulas

Hay una multitud de cópulas. En términos generales, las cópulas pueden agruparse como cópulas con forma cuadrática, cópulas de forma cúbica, y cópulas Arquimedianas. Las cópulas también pueden ser agrupadas por paramétricas y no paramétricas. En las primeras el parámetro, o parámetros cuantifican la relación de dependencia entre las variables que se asocian, un ejemplo de ellas son las cópulas arquimedianas, en las segundas no participa un parámetro ya que por su estructura empírica se ajustan de forma local a los datos, el ejemplo más común es la cópula empírica.

3.6. Cópulas Arquimedianas

Para este trabajo se estudiarán las cópulas arquimedianas ya que éstas son quizás, las cópulas más importantes para el análisis hidrológico por las siguientes razones: (1) Se pueden construir fácilmente, (2) La mayoría de familias de cópulas paramétricas pertenecen a esta clase y (3) Muchas de las propiedades de esta familia pueden describir una gran variedad de

estructuras de dependencia (Nelsen, 2006).

Definición 5. Una cópula C es llamada *Arquimediana* si existe una función continua, estrictamente decreciente y convexa $\varphi: [0, 1] \rightarrow [0, \infty]$ con $\varphi(1) = 0$ y sea $\varphi^{[-1]}$ la pseudo-inversa de φ definida como:

$$\varphi(C_\theta(u, v)) = \varphi(u) + \varphi(v) \quad (3-11)$$

Entonces C es una función de $[0, 1] \times [0, 1] \rightarrow [0, 1]$ que puede escribirse de la siguiente manera:

$$C_\theta(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)) \quad (3-12)$$

donde $\varphi^{[-1]}$ es la pseudo-inversa de φ definida por:

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & 0 \leq t \leq \varphi(0) \\ 0, & \varphi(0) \leq t \leq +\infty \end{cases} \quad (3-13)$$

donde φ es llamado el generador arquimediano de la cópula C , θ es el parámetro de la cópula y C_θ denota la representación de la cópula.

Otro resultado útil para las cópulas arquimedianas es la expresión para el τ de Kendall de la ecuación 3-7 que puede ser simplificada de la siguiente manera (Genest and MacKay, 1986):

Proposición 1. Sean X e Y variables aleatorias con una cópula arquimediana C con generador φ . La versión τ de Kendall $\tau_{X,Y}$ para X e Y de la de ecuación 3-7 está dado por:

$$\tau_{X,Y} = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt \quad (3-14)$$

Propiedades de las cópulas arquimedianas

Las propiedades de las cópulas arquimedianas se encuentran en Nelsen (1999) y Genest and Rivest (1993).

Propiedad 1. La función generadora φ es estrictamente decreciente y convexa de $[0, 1] \rightarrow [0, \infty]$ con $\varphi(1) = 0$.

Esto garantiza que la función generadora disminuye, es decir, la cópula generada satisface

las propiedades de la distribución de probabilidad.

Ejemplo 2. Sea $\varphi(t) = -\ln(t)$, $t \in [0, 1]$ ya que $\varphi(0) = +\infty$, φ es una función convexa decreciente, entonces tenemos que:

$$\varphi^{-1}(t) = \exp(-t)$$

Así, la cópula arquimediana correspondiente generada mediante el uso de esta función de generación es:

$C(u, v) = \exp(-[(-\ln u) + (-\ln v)])$. Esta cópula representa la distribución conjunta de dos variables aleatorias independientes.

Propiedad 2. Sea C una cópula arquimediana con generador φ en ϕ , es decir, $\phi \in (0, 1]$. Entonces para cualquier u, v en $[0, 1]$ existe un número entero positivo m tal que $u_c^m < v$. Esta propiedad describe cómo se comporta una cópula arquimediana. Además, teniendo en cuenta el axioma de Arquímedes para los números reales positivos: si a, b son números reales positivos, entonces existe un número entero tal que $ma \geq b$.

Propiedad 3. Las curvas de nivel de las cópulas arquimedianas son convexas. Sea C una cópula arquimediana con generador φ en ϕ .

1. Para $t \in (0, 1)$ la C-medida de la curva de nivel de C $\varphi(u) + \varphi(v) = \varphi(t)$ es dada por:

$$\varphi(t) \left[\frac{1}{\varphi'(t^-)} - \frac{1}{\varphi'(t^+)} \right]$$

donde $\varphi'(t^-)$, $\varphi'(t^+)$ son derivadas por izquierda o por derecha de φ .

2. Si C no es estricta, la C-medida de la curva de nivel de cero $\varphi(u) + \varphi(v) = \varphi(0)$ es igual a:

$$-\frac{\varphi(0)}{\varphi'(0^+)}$$

Y por lo tanto es igual a cero cuando $\varphi'(0^+) = -\infty$.

Propiedad 4. Sea C una cópula arquimediana con generador φ en ϕ , si $K_C(t)$ indica la C-medida del conjunto $\{(u, v) \in [0, 1]^2 \mid C(u, v) \leq t\}$ o de manera equivalente, si el conjunto $\{(u, v) \in [0, 1]^2 \mid \varphi(u) + \varphi(v) \geq \varphi(t)\}$, esto es:

$$K_C(t) = P(C(u, v) \leq t) \tag{3-15}$$

entonces para cualquier t en $[0, 1]$:

$$K_C(t) = t - \frac{\varphi(t)}{\varphi'(t^+)} \quad (3-16)$$

Teorema 4. *Sea C una cópula arquimediana con generador φ , entonces:*

- a) C es simétrica, es decir $C(u, v) = C(v, u) \quad \forall_{u,v} \in [0, 1]$
- b) C es asociativa, es decir $C(C(u, v), w) = C(u, C(v, w)) \quad \forall_{u,v,w} \in [0, 1]$
- c) si $k > 0$ es una constante y φ el generador, entonces $k\varphi$ es también un generador de C

3.7. Familias de cópulas arquimedianas

Existe una gran variedad de familias de cópulas arquimedianas que se utilizan para la construcción de modelos que representan distribuciones multivariadas. En general existen cerca de 22 familias, pero en este trabajo sólo se discutirán los parámetros de las cópulas arquimedianas más utilizadas. Cada una de estas cópulas conforman una familia parametrizada por θ .

Las cópulas arquimedianas bidimensionales más utilizadas se presentan a continuación:

3.7.1. Cópula Gumbel Hougaard

La cópula arquimediana Gumbel-Hougaard fue introducida por primera vez por Gumbel (1960). La formulación de esta familia se expresa como sigue (Zhang, 2005):

$$C_\theta(u, v) = C_\theta(F_X(x), F_Y(y)) = H(x, y)$$

$$C_\theta(u, v) = \exp\left(-\left((-\ln u)^\theta + (-\ln v)^\theta\right)^{1/\theta}\right), \quad \theta \in [1, \infty) \quad (3-17)$$

donde θ es el parámetro de la función de generación $\varphi(t) = (-\ln t)^\theta$, con $t = u$ o v como una variable aleatoria uniformemente distribuida que varía de 0 a 1.

El coeficiente de correlación entre los vectores aleatorios X y Y es el correspondiente Tau de kendall, que se puede obtener mediante la aplicación de la ecuación 3-14:

$$\tau_\theta = 1 + 4 \int_0^1 \frac{t \ln t}{\theta} dt = 1 - \frac{1}{\theta}$$

$$\tau = 1 - \theta^{-1} \quad 0 < \tau < 1 \quad (3-18)$$

donde $\varphi(u) = (-\ln u)^\theta$ y $\varphi(v) = (-\ln v)^\theta$

3.7.2. Cópula Ali-Mikhail-Haq

La cópula arquimediana Ali-Mikhail-Haq fue desarrollada por Ali et al. (1978). Esta familia de cópula fue desarrollada basada en el concepto de distribución logística univariante. El parámetro de esta cópula es la medida de la salida de la independencia o la medida de la asociación entre dos variables. La formulación de esta familia se expresa como sigue (Zhang, 2005):

$$C_\theta(u, v) = C_\theta(F_X(x), F_Y(y)) = H(x, y) = \frac{uv}{1 - \theta(1-u)(1-v)} \quad , \quad \theta \in [-1, 1] \quad (3-19)$$

con

$$\varphi(t) = \ln \frac{1 - \theta(1-t)}{t}$$

$$\tau = \left(\frac{3\theta - 2}{\theta} \right) - \frac{2}{3} \left(1 - \frac{1}{\theta} \right)^2 \ln(1 - \theta) \quad , \quad \tau \approx (-0,1817 < \tau < 0,333) \quad (3-20)$$

3.7.3. Cópula Frank

La cópula arquimediana Frank fue desarrollada por Frank (1979). La cópula Frank satisface todas las condiciones para la construcción de distribuciones bivariadas con marginales fijas. Es absolutamente continua y tiene soporte completo en el cuadrado de la unidad. La formulación de esta familia se expresa como (Zhang, 2005):

$$C_\theta(u, v) = C_\theta(F_X(x), F_Y(y)) = H(x, y)$$

$$C_\theta(u, v) = -\frac{1}{\theta} \ln \left(1 + \frac{(\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}{\exp(-\theta) - 1} \right) \quad , \quad \theta \neq 0 \quad (3-21)$$

con

$$\varphi(t) = -\ln \left(\frac{\exp(-\theta t) - 1}{\exp(-\theta) - 1} \right)$$

$$\tau = 1 - \frac{4}{\theta}(1 - D_1(\theta)) \quad (3-22)$$

donde D_1 es la función Debye de primer orden. D_k se define como:

$$D_k(\theta) = \frac{k}{\theta^k} \int_0^\theta \frac{t^k}{\exp(t) - 1} dt \quad (3-23)$$

Resolviendo las ecuaciones 3-22 y 3-23 se obtuvo la figura 3-2.

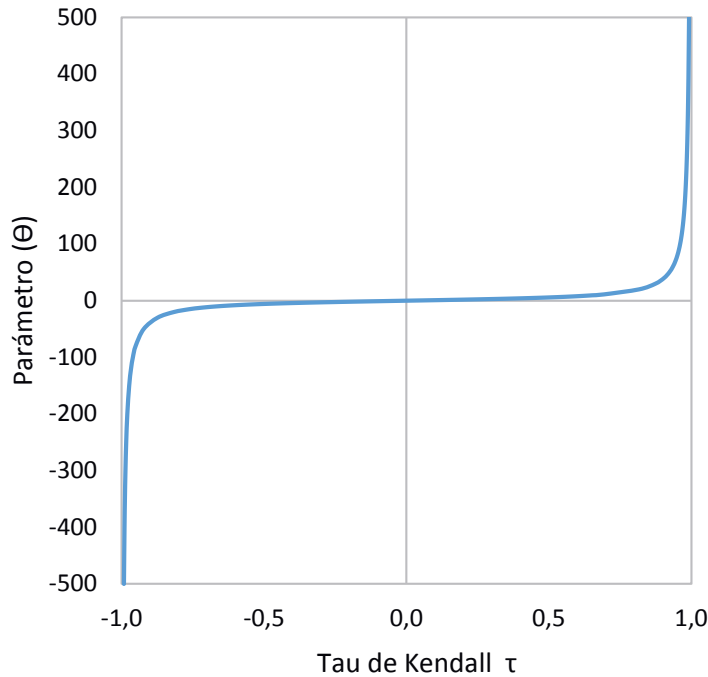


Figura 3-2.: Relación entre el τ de Kendall y el parámetro θ para la cópula Frank.

3.7.4. Cópula Cook-Johnson (Clayton)

Como se menciona en Nelsen (1999), la familia de la cópula arquimediana Clayton se obtuvo por Clayton (1978), Cox and Oakes (1984) y Cook and Johnson (1981). Como se discute en Cook and Johnson (1981), esta familia puede ser utilizada para el modelado de datos multivariados simétricos no elípticos. Cuando $\theta = 0$, esta cópula representa la distribución logística bivariada. La formulación de esta familia se expresa como sigue (Zhang, 2005):

$$C_\theta(u, v) = C_\theta(F_X(x), F_Y(y)) = H(x, y) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta} \quad \theta \geq 0 \quad (3-24)$$

con

$$\varphi(t) = \frac{1}{\theta}(t^{-\theta} - 1)$$

El Tau de Kendall se puede obtener aplicando la ecuación 3-14:

$$\tau_{\theta} = 1 + 4 \int_0^1 \frac{t^{\theta+1} - t}{\theta} dt = 1 + \frac{4}{\theta} \left(\frac{1}{\theta+2} - \frac{1}{2} \right) = \frac{\theta}{\theta+2}$$

$$\tau = \frac{\theta}{\theta+2} \quad , \quad 0 < \tau < 1 \quad (3-25)$$

La cópula Clayton no se puede considerar para la dependencia negativa. Alcanza el límite superior de Fréchet cuando $\theta \rightarrow \infty$, pero no puede alcanzar el límite inferior de Fréchet

3.7.5. Cópula Joe

La cópula arquimediana Joe fue introducida por primera vez por Joe (1993). Cuando $\theta = 1$ esta cópula representa la distribución bivariada de dos variables independientes. De nuevo, esta cópula tiene la motivación de aplicaciones en inferencias de valores extremos, y es independiente de las marginales univariadas. La formulación de esta familia se expresa como sigue (Zhang, 2005):

$$C_{\theta}(u, v) = C_{\theta}(F_X(x), F_Y(y)) = H(x, y)$$

$$C_{\theta}(u, v) = 1 - [(1-u)^{\theta} + (1-v)^{\theta} - (1-u)^{\theta}(1-v)^{\theta}]^{1/\theta} \quad (3-26)$$

con

$$\varphi(t) = -\ln(1 - (1-t)^{\theta}) \quad , \quad \theta \geq 1$$

La cópula Joe es similar a la cópula Clayton. No se puede considerar para la dependencia negativa. Se alcanza el límite superior de Fréchet cuando $\theta \rightarrow \infty$, pero no puede alcanzar el límite inferior de Fréchet. La relación entre τ y θ para la cópula de Joe no tiene una forma cerrada, pero tiene la siguiente forma (Bhat and Eluru, 2009):

$$\tau = 1 + \frac{4}{\theta} D_J(\theta) \quad , \quad 0 < \tau < 1 \quad (3-27)$$

$$D_J(\theta) = \int_{t=0}^1 \frac{[\ln(1-t^{\theta})] (1-t^{\theta})}{t^{\theta-1}} dt \quad (3-28)$$

Resolviendo las ecuaciones 3-27 y 3-28 se obtuvo la figura **3-3**.

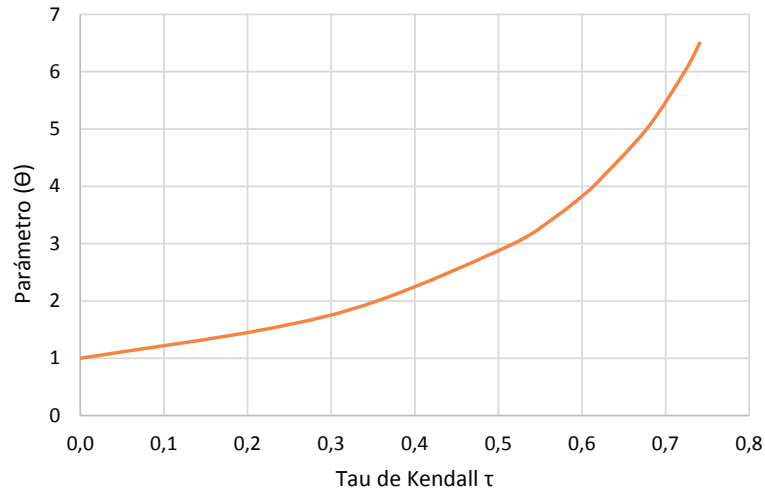


Figura 3-3.: Relación entre el τ de Kendall y el parámetro θ para la cópula Joe.

3.8. Cópula Empírica

La cópula empírica también conocida como cópula Deheuvels, fue introducida por Deheuvels (1979), y posteriormente estudiada por Doukhan et al. (2005), hace parte de la estimación no paramétrica que consiste en construir una función cópula a partir de valores muestrales recolectados de las variables aleatorias sin establecer dependencia de ningún parámetro (Bouyé et al., 2000).

Definición 6. Sea $\{(x_k, y_k)\}_{k=1}^n$ una muestra de tamaño n obtenida a partir de una distribución bivariada continua. La *cópula empírica* es la función C_n dada por:

$$C_n \left(\frac{i}{n}, \frac{j}{n} \right) = \frac{1}{n} \sum_{k=1}^n I(x_k \leq x_{(i)}, y_k \leq y_{(j)}) \quad i, j = 1, \dots, n \quad (3-29)$$

esto es lo mismo que:

$$C_n \left(\frac{i}{n}, \frac{j}{n} \right) = \frac{\text{número de parejas } (x, y) \text{ en la muestra con } x \leq x_{(i)}, y \leq y_{(j)}}{n} \quad (3-30)$$

donde $x_{(i)}$ y $y_{(j)}$, $1 \leq i, j \leq n$, son las estadísticas de orden de la muestra.

$I(A)$ es la función característica del conjunto A dada por:

$$I(A) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{en otro caso} \end{cases}$$

La frecuencia de la cópula empírica C_n está dada por:

$$C_n \left(\frac{i}{n}, \frac{j}{n} \right) = \begin{cases} \frac{1}{n}, & \text{si } (x_{(i)}, y_{(j)}) \text{ es un elemento de la muestra,} \\ 0 & \text{en otro caso} \end{cases} \quad (3-31)$$

La cópula empírica converge uniformemente a la cópula verdadera cuando el tamaño de la muestra crece (Paloma, 2006). Esto implica que cuando la cópula verdadera es desconocida, un criterio de selección como se verá más adelante, es comparar cada una de las cópulas seleccionadas con la cópula empírica.

3.9. Estimación del parámetro de dependencia θ

El parámetro θ de la cópula es un vector que mide la dependencia entre variables aleatorias, es decir; la dependencia entre las marginales. Para el caso bivariado, el parámetro de la cópula θ se determina a partir de una estimación no paramétrica por medio del método de los momentos. Para determinar este parámetro es necesario primero obtener el coeficiente de correlación τ de Kendall como se vio en la sección 3.4.1.

$$\tau_n = \binom{n}{2}^{-1} \sum_{i < j} \text{sign} [(x_i - x_j)(y_i - y_j)] \quad (3-32)$$

donde

$$\text{sign}(\Delta x, \Delta y) = \begin{cases} 1 & \text{si } x_i < x_j \text{ y } y_i < y_j \text{ ó } x_i > x_j \text{ y } y_i > y_j \\ 0 & \text{si } (x_i - x_j)(y_i - y_j) = 0 \\ -1 & \text{en otro caso } i, j = 1, 2, \dots, n \end{cases} \quad (3-33)$$

Además, n es el número de observaciones y τ_n es la estimación de τ a partir de n observaciones.

Ejemplo 3. Tomamos la función generadora de la cópula Gumbel Hougaard.

Sea $\varphi(u) = (-\ln t)^\theta$, entonces aplicando la ecuación 3-14:

$$\tau_n = 1 + 4 \int_0^1 \frac{\varphi(u)}{\varphi'(u)} du$$

$$\tau_n = \frac{\theta - 1}{\theta}$$

$$\theta = \frac{1}{1 - \tau_n}$$

3.10. Selección de la Cópula

Muchas propuestas se han realizado recientemente de pruebas de bondad y ajuste para modelos implantados con cópulas. Para evaluar y seleccionar la cópula de mejor ajuste a los datos, en este trabajo se usaron dos métodos. El primero de ellos consiste en un método gráfico utilizando la función de distribución de la cópula (únicamente para el caso bivariado). El segundo método (caso bivariado y multivariado) se realizó mediante el test desarrollado por Genest et al. (2009) que se basa en el estadístico Cramer-von Mises, el cual consiste en realizar pruebas “blanket test” (pruebas manta) cuya aplicación no requiere categorizaciones arbitrarias ya que se realizan mediante experimentos de Monte Carlo diseñados para evaluar el efecto del tamaño de la muestra con la fuerza de dependencia entre los datos. Esta prueba es un proceso que compara la cópula empírica con las cópulas estudiadas o seleccionadas bajo una hipótesis nula de independencia. Este test está implementado en el paquete “copula” de la herramienta computacional estadística R, el cual se usó para el análisis de la cópula siguiendo la metodología propuesta por Kojadinovic et al. (2010).

Aproximación utilizando la función de distribución de la cópula. (Método gráfico)

Para cópulas bivariadas, Genest and Rivest (1993) describen un procedimiento gráfico para identificar una función cópula en base a una estimación no paramétrica. A continuación se muestran los pasos para identificar la cópula apropiada:

1. Definir una variable aleatoria intermedia $Z = H(X, Y)$ con una función de distribución $K(z) = P(Z \leq z)$. Esta distribución está relacionada con el generador de la cópula a través de la propiedad 4, es decir; la ecuación 3-16.

2. Construir una estimación no paramétrica de K como:

a) Obtener:

$$z_i = \frac{\{\text{Número de } (x_j, y_j) \text{ tales que } (x_j < x_i) \text{ y } (y_j < y_i)\}}{n - 1} \quad \text{para } i = 1, \dots, n \quad (3-34)$$

b) Construir una estimación de K como $K_n(z) =$ la proporción de z_i 's $\leq z$

3. Construir una estimación paramétrica de K usando la ecuación 3-16.

$$K_C(t) = t - \frac{\varphi(t)}{\varphi'(t)}$$

4. Construir la gráfica no paramétrica $K_n(z)$ contra la estimación paramétrica K .

La gráfica $K_n(z)$ vs $K_C(t)$ también es conocida como *Q-Q plot*

Si la gráfica concuerda con una línea recta que pasa por el origen en un ángulo de 45° la función generadora es satisfactoria. De lo contrario, la función cópula debe ser re-identificada

Selección de la cópula mediante el estadístico Cramer-von Mises usando el paquete “copula” de R

Los modelos basados en el paquete “copula” de R se presentan como una sucesión de pruebas basadas en filas: una prueba multivariada de aleatoriedad seguida de una prueba de independencia mutua y una serie de pruebas de bondad de ajuste, todas ellas basadas en la cópula empírica que es un estimador basado en el rango no paramétrico de la verdadera cópula que es desconocida.

En una gran parte de la literatura, se argumenta que la estimación de C debe basarse únicamente en los vectores fila R_1, \dots, R_n donde $R_i = (R_{i1}, \dots, R_{id})$ y $R_{i,j}$ es la fila de $X_{i,j}$ entre X_{1j}, \dots, X_{nj} . (Kojadinovic et al., 2010).

Resumiendo, los pasos que se siguen son los siguientes:

a) El primer paso práctico en la construcción de un modelo de C es probar si las variables aleatorias X_1, \dots, X_n son independientes entre sí, es decir, si pueden ser consideradas como una muestra aleatoria de H :

$$H(X) = C\{F_1(x_1), \dots, F_d(x_d)\} \quad X \in \mathbb{R}^d$$

- b) Si la hipótesis es rechazada se puede intentar ajustar un modelo de series de tiempo para cada marginal y trabajar con los residuales, la idea es intentar remover la dependencia en los componentes de la serie histórica como se discutió en Grégoire et al. (2008)
- c) Si la hipótesis no se rechaza, un segundo paso razonable es realizar la prueba en contra de la presencia de la dependencia entre los componentes de X , esto equivale a la prueba:

$$H_0 : C = \prod \quad \text{contra} \quad H_1 : C \neq \prod,$$

donde $\prod(u) = \prod_{i=1}^d u_i$, $u \in [0, 1]^d$ es la *independencia de la cópula*.

- d) Si la independencia es rechazada, el siguiente paso es ajustar una familia de cópulas paramétricas apropiadas a los datos disponibles. En la práctica, esto equivale a efectuar pruebas de bondad de ajuste de la forma:

$$H_0 : C \in \mathcal{C} \quad \text{contra} \quad H_1 : C \notin \mathcal{C}$$

para varias familias paramétricas $\mathcal{C} = \{C_\theta\}$

- e) El último paso consiste en elegir una de las familias que no fueron rechazadas, en su caso, es posible, proporcione los errores estandar de las estimaciones de los parámetros.

Todos los pasos mencionados anteriormente pueden llevarse a cabo por medio del paquete “copula” de R.

Un elemento importante de los ensayos bajo consideración es la cópula empírica de los datos (Deheuvels, 1979), (Deheuvels 1979, 1981b), que es un estimador consistente de la cópula desconocida C . (ver sección 3.8). Las pruebas que se realizan en el paquete “copula” de R, se muestran a continuación:

1. **Prueba de independencia multivariada:** Inspirada en el trabajo de Blum et al. (1961); Deheuvels (1980), y más recientemente por Ghoudi et al. (2001); Genest and Rémillard (2004), sugiere una prueba de independencia mutua de los componentes de X en el estadístico (Kojadinovic et al., 2010):

$$I_n = \int_{[0,1]^d} n \left\{ C_n(u) - \prod_{i=1}^d u_i \right\}^2 du \quad (3-35)$$

Un aspecto interesante de esta prueba bajo la independencia mutua de los componentes X_1, \dots, X_d de X el proceso empírico $\sqrt{n} \{C_n - \prod\}$ puede ser descompuesto, usando la

transformación de Möbius (Rota, 1964). Una propiedad fundamental de esta descomposición, cuya forma está dada en Genest and Rémillard (2004) es que la independencia mutua entre X_1, \dots, X_d es equivalente a tener $M_A(C)(u) = 0$ para todo $u \in [0, 1]^d$ y todo $A \subseteq \{1, \dots, d\}$ tal que $|A| > 1$. Las propiedades de este estadístico se encuentran en Kojadinovic et al. (2010). Como una alternativa del estadístico I_n Genest and Rémillard (2004) estudiaron varias combinaciones para la independencia de pruebas. Dos reglas de combinación implementadas en el paquete “copula” de R son las de Fisher (1992) y Tippett (1931) que tienden a dar mejores resultados. Para visualizar los resultados de la prueba de independencia, una representación gráfica llamada *dependogram* puede ser usada. Los ensayos descritos anteriormente se implementan en las funciones *indepTestSim* y *indepTest* del paquete “copula” de R.

2. **Prueba de aleatoriedad:** La prueba de independencia multivariada de Deheuvels (1980) se puede extender para probar la aleatoriedad como sugiere Genest and Rémillard (2004). Las funciones de aplicación de estas pruebas son *serialIndepTestSim* y *serialIndepTest*. Para saber más acerca de esta prueba, ver Kojadinovic et al. (2010).
3. **Pruebas de bondad de ajuste:** Las pruebas de bondad de ajuste implementadas en el paquete “copula” de R se basan en el proceso empírico:

$$\mathbb{C}_n(u) = \sqrt{n}\{C_n(u) - C_{\theta_n}(u)\}, \quad u \in [0, 1]^d \quad (3-36)$$

donde C_n es la cópula empírica definida en la sección 3.8 y C_{θ_n} es un estimador de C bajo la hipótesis $H_0 : C \in \{C_\theta\}$. El estimador θ_n está basado únicamente en filas. Es uno de los dos estimadores del método de momentos basado respectivamente en la inversión de tau de Kendall y rho de Spearman, o el estimador de máxima pseudo-likelihood de Genest et al. (1995).

En la gran escala de experimentos de Monte Carlo realizados por Berg (2009) y Genest and Rémillard (2004) el estadístico:

$$S_n = \int_{[0,1]^d} \mathbb{C}_n(u)^2 dC_n(u) = \sum_{i=1}^n \{C_n(\hat{U}_i) - C_{\theta_n}(\hat{U}_i)\}^2 \quad (3-37)$$

dará los mejores resultados globales.

Un p -valor aproximado para S_n se puede obtener por medio de un procedimiento basado en bootstrap paramétrico y cuya validez ha sido recientemente demostrada por Genest and Rémillard (2008). El principal inconveniente de este enfoque es su alto coste computacional, ya que cada iteración requiere tanto la generación de números aleatorios de la cópula y la estimación de los parámetros de la cópula.

Con el fin de eludir este muy alto coste computacional, un procedimiento de ensayo basado en multiplicadores del teorema del límite central fue propuesto por Kojadinovic et al. (2011).

Ambos métodos se implementan en el paquete “copula” de R con una función de interfaz común que se llama *gofCopula*. Los argumentos importantes *gofCopula* son los siguientes (Kojadinovic et al., 2010).

- *copula*: un objeto que representa la cópula hipotética cuyos atributo *parameters* es ignorado.
- *X*: la matriz de datos observados, cada fila de las cuales es una observación multivariada.
- *simulation*: el método de simulación puede ser: “*pb*” (bootstrap paramétrico) o “*mult*” (multiplicador). El método de simulación por defecto es “*pb*”.
- *N*: el número de bootstrap / iteraciones múltiples.
- *method*: método de estimación para el parámetro de la cópula; puede ser “*mpl*” (máximo pseudo-likelihood), “*itau*” (inversión del tau de Kendall τ) o “*irho*” (inversión del rho de Spearman ρ).

4. **Aspectos computacionales:** En el paquete “copula” de R se han implementado las cinco familias de cópulas estudiadas: Gumbel-Hougaard, Frank, Clayton, Ali-Mikhail-Haq y Joe. Para cópulas bivariadas, los tres métodos para estimar el parámetro de la cópula θ “*mpl*”, “*itau*” y “*irho*” pueden ser utilizados. En dimensión tres o superior sólo la estimación máximo pseudo-likelihood (“*mpl*”) está disponible. Las familias Clayton y Gumbel-Hougaard no pueden ser utilizadas para dimensiones superiores a diez (10), mientras que para la familia Frank sólo se cuenta hasta un máximo de seis (6) dimensiones, ésto es debido principalmente a que las expresiones de los pdfs de estas cópulas (y sus derivadas parciales) no se han obtenido y almacenado para dimensiones superiores.

3.11. Periodo de retorno condicional y posiciones de graficación para análisis de frecuencia bivariados

Periodo de retorno bivariado

El periodo de retorno bivariado conjunto de un evento (x, y) , $T_{X,Y}(x, y)$ puede ser definido a través de una distribución de probabilidad conjunta $H_{X,Y}(x, y)$, como:

$$T_{X,Y}(x, y) = \frac{1}{1 - H_{X,Y}(x, y)} \quad (3-38)$$

donde las funciones de distribución acumuladas (CDF) de X y Y son $F_X(x)$ y $F_Y(y)$ respectivamente. Los periodos de retorno de $X = x$ y $Y = y$ son denotados respectivamente como, $T_X(x)$ y $T_Y(y)$

Los periodos de retorno condicional se presentan a continuación:

Sea $f_{X,Y}(x, y)$ la función de densidad de probabilidad conjunta (PDF) de X y Y y sean $f_X(x)$ y $f_Y(y)$ las PDF marginales de X y Y respectivamente, entonces los periodos de retorno condicional se pueden expresar mediante los siguientes dos casos (Zhang, 2005):

Caso 1.

- La función de densidad de probabilidad conjunta condicional (CPDF) de X dado $Y = y$, $f_{X|Y=y}(x | y)$ puede ser expresada como:

$$f_{X|Y=y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(Y = y)} \tag{3-39}$$

- Del mismo modo, el condicional CDF (CCDF) de X dado $Y = y$ puede ser expresado como:

$$F_{X|Y=y}(x | y) = \int_{-\infty}^x f_{X|Y=y}(u | y) du = \frac{\int_{-\infty}^x f_{X,Y}(u, y) du}{f_Y(y)} \tag{3-40}$$

- El periodo de retorno condicional asociado del evento $X > x$ dado $Y = y$ puede ser escrito como:

$$T_{X|Y=y}(x | y) = \frac{1}{1 - F_{X|Y=y}(x | y)} \tag{3-41}$$

El periodo de retorno condicional $T_{Y|X=x}(y | x)$ del evento $Y > y$ dado $X = x$ puede ser expresado de manera similar.

Caso 2.

En la práctica, uno puede estar más interesado en el período de retorno condicional del evento $X \geq x$ dado $Y \leq y$ (Yue, 1999, 2000c; Yue et al., 1999; Yue, 2000a,b).

- El CCDF de $X \geq x$ dado $Y \leq y$, $F'_{X|Y \leq y}(x | y)$, es dado por:

$$F'_{X|Y \leq y}(x | y) = \frac{H_{X,Y}(x, y)}{F_Y(y)} \quad (3-42)$$

- Así, el periodo de retorno condicional de $X \geq x$ dado $Y \leq y$ es:

$$T'_{X|Y \leq y}(x | y) = \frac{1}{1 - F'_{X|Y \leq y}(x | y)} \quad (3-43)$$

Posiciones de graficación bivariada

Sea (x_i, y_i) , $i = 1, \dots, n$, las observaciones bivariadas o de dos variables, que se clasifican en orden ascendente por los valores de la variable aleatoria X . Contando el número de parejas (x_j, y_j) , cuando $x_j \leq x_i$, $y_j \leq y_i$, $j < i = 1, \dots, n$, la probabilidad conjunta acumulada para cada pareja (x_i, y_i) por ejemplo; lluvia y duración o intensidad y duración. Si tomamos por ejemplo, la distribución Gringorten (ver tabla **2-1**) la posición de graficación bivariada se pueden calcular empíricamente como (Zhang, 2005):

$$F(x_i, y_i) = P(X \leq x_i, Y \leq y_i) = \frac{\text{No. de } (x_j \leq x_i, \text{ y } y_j \leq y_i) - 0,44}{n + 0,12} \quad (3-44)$$

donde n es el tamaño de la muestra y $1 \leq i, j \leq n$ y $j < i$
 Esto también se puede escribir como:

$$F(x_i, y_i) = P(X \leq x_i, Y \leq y_i) = \frac{\sum_{m=1}^i \sum_{l=1}^i n_{ml} - 0,44}{N + 0,12} \quad (3-45)$$

donde N es el tamaño de la muestra; y además:

$$n_{ml} = \begin{cases} 1 & \text{si } x_m \leq x_i \text{ y } y_l \leq y_j, \quad j < i = 1, \dots, n \\ 0 & \text{en otro caso} \end{cases}$$

3.12. Distribuciones condicionales bivariadas a través de cópulas y correspondientes periodos de retorno

La distribución conjunta condicional utilizando el método de la cópula se presenta a continuación:

Sean X y Y dos variables aleatorias con marginales $u = F_X(x)$, $v = F_Y(y)$. Las funciones de distribución condicional bajo diferentes condiciones pueden ser expresadas utilizando la cópula arquimediana bivariada mediante los siguientes dos casos (Zhang, 2005):

Caso 1.

- La función de distribución condicional para la variable X dado $Y = y$ puede ser escrita como: $H(X | Y = y) = C_{\theta}^2_{U|V=v}(u) = C_{\theta}^2(U \leq u | V = v)$

$$H(X | Y = y) = \lim_{\Delta v \rightarrow 0} \frac{C_{\theta}^2(u, v + \Delta v) - C_{\theta}^2(u, v)}{\Delta v} = \frac{\partial}{\partial v} C_{\theta}^2(u, v) | V = v \quad (3-46)$$

- Así, el periodo de retorno correspondiente bajo la condición dada anteriormente se puede escribir como:

$$T(X | Y = y) = \frac{1}{1 - C_{\theta}^2(U \leq u | V = v)} \quad (3-47)$$

De la misma manera, una fórmula equivalente para la función de distribución condicional para la variable Y dado $X = x$ puede ser obtenida.

Caso 2.

- La función de distribución condicional de X dado $Y \leq y$ puede ser expresada como:

$$H(X | Y \leq y) = C_{\theta}^2_{U|V \leq v} = C_{\theta}^2(U \leq u | V \leq v) = \frac{C_{\theta}^2(u, v)}{v} \quad (3-48)$$

- Así, el periodo de retorno correspondiente bajo esta condición es escrito como:

$$T(X | Y \leq y) = \frac{1}{1 - C_{\theta}^2(U \leq u | V \leq v)} \quad (3-49)$$

3.13. Aplicación de las cópulas bivariadas en el área de estudio

Como ya se mencionó, las cópulas son funciones que conectan las distribuciones de probabilidad conjunta con sus distribuciones de probabilidad marginales. Para obtener estas distribuciones marginales se realizaron las pruebas de ajuste Kolmogorov-Smirnov y χ^2 para cada una de las estaciones seleccionadas. Las distribuciones resultantes se presentaron en el capítulo 2 (ver tabla 2-4). El objetivo de este capítulo, es obtener las curvas IDF utilizando el método de la cópula.

3.13.1. Construcción de curvas IDF por medio del método de la cópula

De las cinco familias de cópulas arquimedianas mencionadas en la sección 3.7, la cópula Gumbel-Hougaard, la cópula Cook-Johnson y la cópula Joe son apropiadas, sólo para variables correlacionadas positivamente, mientras que la cópula Frank y la cópula AliAli-Mikhail-Haq son apropiadas para variables correlacionadas tanto positivamente como negativamente, sin embargo, la cópula Ali-Mikhail-Haq puede no ser apropiada para variables altamente correlacionadas positivamente o negativamente, por lo tanto, la cópula Frank fue elegida para obtener las curvas IDF. El criterio para saber si dos variables están correlacionadas negativamente o positivamente es determinado mediante el coeficiente de correlación τ de Kendall.

Estación El Bosque

Para expresar una cópula arquimediana con un parámetro de dos variables aleatorias, es necesario establecer las distribuciones marginales acumuladas $F_X(x)$ y $F_Y(y)$ mediante:

$$\begin{aligned} u &= F_X(x) = \text{Probabilidad acumulada de la intensidad de la lluvia (mm/h)} \\ v &= F_Y(y) = \text{Probabilidad acumulada de la duración de la tormenta (min)} \end{aligned}$$

Como se vio en la sección 3.7.3 la cópula Frank tiene la siguiente expresión:

$$C_\theta(u, v) = -\frac{1}{\theta} \ln \left(1 + \frac{(\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}{\exp(-\theta) - 1} \right), \quad \theta \neq 0$$

Esta expresión también puede ser escrita como:

$$C_\theta(u, v) = -\frac{1}{\theta} \ln \left(\frac{(\exp(-\theta) - 1) + (\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}{\exp(-\theta) - 1} \right), \quad \theta \neq 0 \quad (3-50)$$

En términos de sus distribuciones marginales, la función de distribución conjunta es:

$$H(x, y) = C[F_X(x), F_Y(y)] = -\frac{1}{\theta} \ln \left(\frac{[\exp(-\theta) - 1] + [\exp(-\theta F_X(x)) - 1][\exp(-\theta F_Y(y)) - 1]}{\exp(-\theta) - 1} \right) \quad (3-51)$$

con

$$\varphi(t) = -\ln \left(\frac{\exp(-\theta t) - 1}{\exp(-\theta) - 1} \right)$$

$$\tau = 1 - \frac{4}{\theta} (1 - D_k(\theta)), \quad D_k(\theta) = \frac{k}{\theta^k} \int_0^\theta \frac{t^k}{\exp(t) - 1} dt, \quad \theta > 0 \quad (3-52)$$

donde θ es el parámetro de la función cópula que puede ser obtenida resolviendo las ecuaciones 3-22 y 3-23. t es el valor de una variable aleatoria uniformemente distribuida u o v que varía de 0 a 1 y el τ de Kendall es el coeficiente de correlación entre las variables aleatorias.

Como se vio en la sección 3.9, la estimación de parámetro de dependencia θ puede ser obtenido a través del coeficiente de correlación τ de Kendall como se muestra a continuación:

$$\tau_n = \binom{n}{2}^{-1} \sum_{i < j} \text{sign} [(x_i - x_j)(y_i - y_j)]$$

donde

$$\text{sign}(\Delta x, \Delta y) = \begin{cases} 1 & \text{si } x_i < x_j \text{ y } y_i < y_j \text{ ó } x_i > x_j \text{ y } y_i > y_j \\ 0 & \text{si } (x_i - x_j)(y_i - y_j) = 0 \\ -1 & \text{en otro caso } i, j = 1, 2, \dots, n \end{cases}$$

x_i = valores para la intensidad de la lluvia (mm/h)

y_i = valores para la duración de la tormenta (min)

Así, para la estación El Bosque, el coeficiente de correlación es (ver anexo B):

$$\tau = -0,5564$$

Como el parámetro θ no puede obtenerse directamente de la ecuación 3-52, éste se determinó numéricamente a partir de la construcción de la ecuación 3-52 representada en la figura 3-2. El gráfico ampliado se presenta en la figura 3-4:

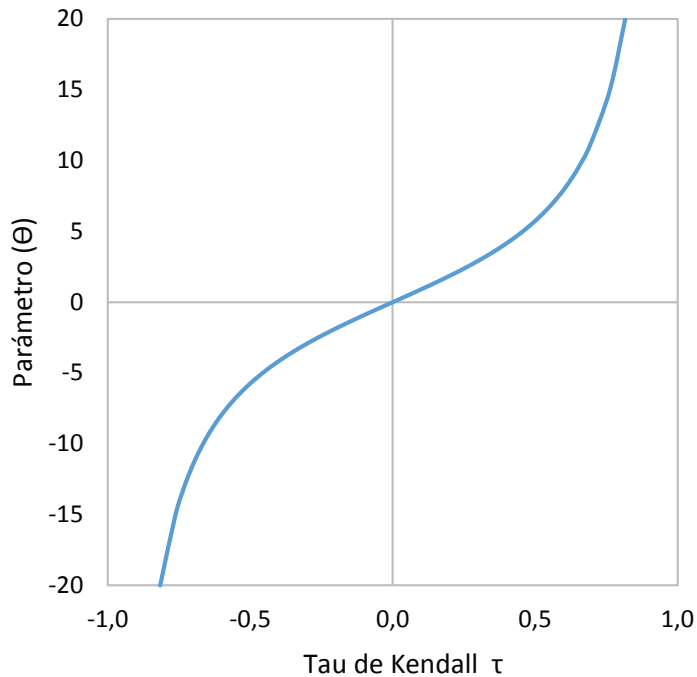


Figura 3-4.: Relación entre el τ de Kendall y el parámetro θ para la cópula Frank.

Entrando a la figura 3-4 Con $\tau = -0,5564$ se encuentra que el parámetro $\theta = -6,87$.

Para curvas IDF, la función cópula Frank condicional es necesaria (ver sección 3.12). Teniendo $u = F_X(x) = i$ y $v = F_Y(y) = d$ la función de distribución condicional de I dado $D = d$ (intensidad y duración respectivamente) puede ser expresada mediante la cópula Frank como (Joe, 1997):

$$F(X \leq x | Y = y) = C_\theta(u | V = v) = \lim_{\Delta v \rightarrow 0} \frac{C_\theta(u, v + \Delta v) - C_\theta(u, v)}{\Delta v} = \frac{\partial}{\partial v} C_\theta(u, v) | V = v \quad (3-53)$$

Esto quiere decir que es necesario fijar alguna de las dos variables aleatorias. Es recomendable fijar la duración como se mencionó en la sección 2.4.2, con intervalos definidos de [15, 30, 45, 60, 90, 120, 150, 180, 240, 300, 360] minutos.

Al derivar la expresión 3-50 con respecto a nuestra variable fija, es decir; la duración, la cópula Frank condicional es expresada como:

$$\frac{\partial}{\partial v} C_\theta(u, v) | V = v = C_{I|D=d} = \frac{[\exp(-\theta i) - 1][\exp(-\theta d)]}{[\exp(-\theta d) - 1][\exp(-\theta i) - 1] + [\exp(-\theta) - 1]} \quad (3-54)$$

El periodo de retorno condicional correspondiente, puede ser expresado como se mencionó en la sección 3.12 (caso 1):

$$T(I | D = d) = \frac{1}{1 - C_{I|D=d}} \quad (3-55)$$

$$T(I | D = d) = \frac{1}{1 - \frac{[\exp(-\theta i) - 1][\exp(-\theta d)]}{[\exp(-\theta d) - 1][\exp(-\theta i) - 1] + [\exp(-\theta) - 1]}} \quad (3-56)$$

La expresión 3-56, es la curva IDF.

Ejemplo 4. Sean $u = F_X(x) = i$ y $v = F_Y(y) = d$, las distribuciones marginales de la intensidad y duración respectivamente para la estación el bosque, en donde:

$u = F_X(x) = \text{Prob. Intensidad de la lluvia (mm/h)}$ CDF \rightarrow Distribución exponencial
 $v = F_Y(y) = \text{Prob. Duración de la tormenta (min)}$ CDF \rightarrow Distribución exponencial

donde CDF = función de distribución acumulada.

Dadas la duración y el periodo de retorno, se puede obtener la curva IDF mediante la expresión 3-56, de esta expresión se despeja la probabilidad acumulada de la intensidad.

Como esta intensidad no puede obtenerse directamente de la ecuación 3-56 ya que esta es una ecuación implícita, es necesario buscar una función objetivo para obtener este resultado.

Es importante resaltar que el valor de la intensidad en la ecuación 3-56 es un valor acumulado, es decir; que pertenece a la función de distribución acumulada. Como inicialmente se asumió una distribución marginal ajustada a una distribución exponencial para la variable aleatoria intensidad, es necesario despejar la intensidad que se busca de la ecuación A-2 que se encuentra en el Anexo A. (Ver tabla **2-2**):

$$F(x) = 1 - e^{-\lambda I} \quad , \quad I = -\frac{-\ln(1 - F_X(x))}{\lambda} \quad (3-57)$$

La curva IDF obtenida aplicando el método de la cópula se presenta en la figura **3-5**.

Una vez obtenidas las curvas IDF por medio del método de la cópula, se realiza la comparación con el método empírico como se muestra en la figura **3-6**.

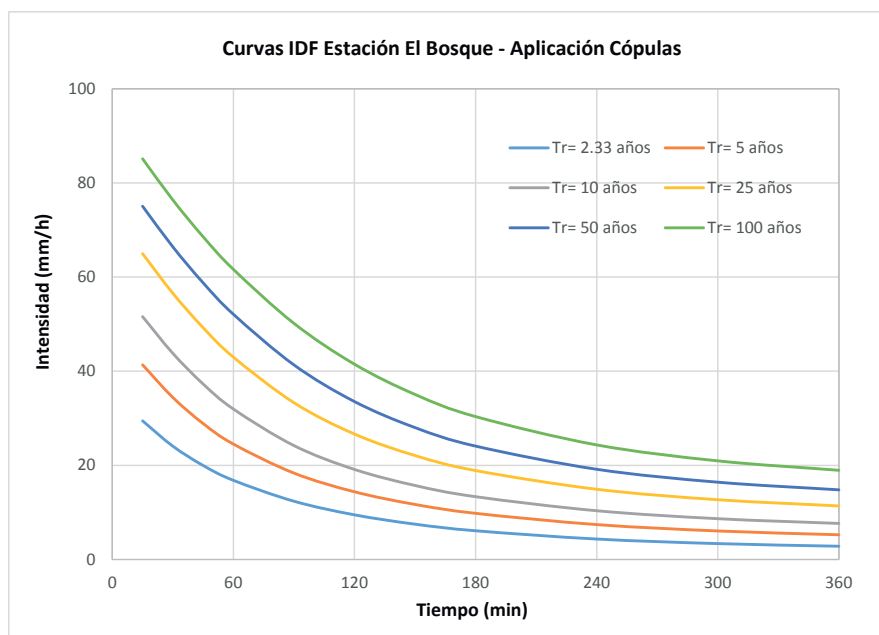


Figura 3-5.: Curvas IDF por medio de la aplicación del método de la cópula para la estación El Bosque (Distribución asumida: Exponencial).

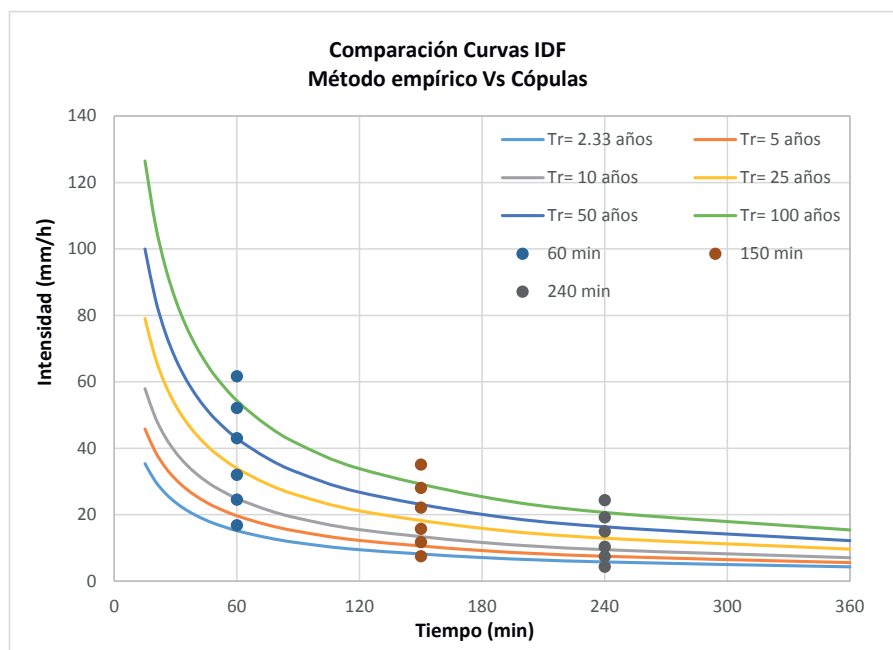


Figura 3-6.: Comparación de las curvas IDF entre el método de la cópula y el método empírico para la estación El Bosque.

Éstas dos últimas figuras representan las curvas IDF obtenidas a partir del método empírico y el método de la cópula para la función cópula arquimediana Frank. Se puede observar a simple vista que no son muy diferentes una de la otra. Sin embargo, como la distribución exponencial no es la que mejor se ajusta a los registros históricos de intensidad y duración, es necesario realizar una prueba de bondad y ajuste a estos datos con el fin de obtener las distribuciones marginales de mejor ajuste tal y como se realizó en la sección 2.4.2. Las distribuciones marginales seleccionadas se encuentran en la tabla **2-4**.

Finalmente, para la estación el Bosque, se tiene que:

$u = F_X(x) = \text{Prob. Acum. Intensidad de la lluvia (mm/h)} \text{ CDF} \rightarrow \text{Distribución LogNormal}$

$v = F_Y(y) = \text{Prob. Acum. Duración de la tormenta (min)} \text{ CDF} \rightarrow \text{Distribución LogNormal}$

Las curvas IDF obtenidas con las distribuciones marginales ajustadas a la distribución LogNormal por medio del método de la cópula se presentan en la figura **3-7**.

Las curvas IDF obtenidas con el método de la cópula se comparan con las curvas IDF empíricas por medio de la figura **3-8**. Para no saturar la figura, se establecieron tres diferentes duraciones [60, 150 y 240] minutos, que corresponden al método de la cópula. El porcentaje de diferencia se encuentra en la tabla **3-2**.

La curva IDF empírica para la estación El Bosque se construyó mediante el modelo propuesto por Chow et al. (1994):

$$i = \frac{cT^m}{D^e + f} = \frac{285,37 T^{0,34}}{D^{0,755} + 3,024} \quad (3-58)$$

Tabla 3-1.: Coeficiente de correlación τ de Kendall y parámetro de la cópula Frank para las estaciones seleccionadas

| Estación | τ de Kendall | θ |
|--------------|-------------------|----------|
| El Bosque | -0,556 | -6,87 |
| El delirio | -0,718 | -12,27 |
| Santa Lucía | -0,738 | -13,41 |
| Bosa Barreno | -0,788 | -17,1 |

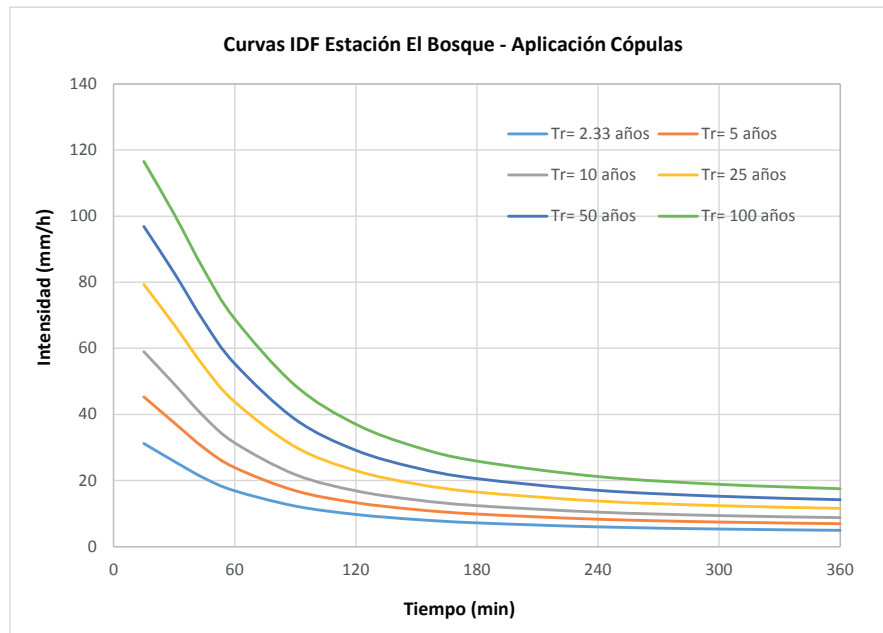


Figura 3-7.: Curvas IDF por medio de la aplicación del método de la cópula para la estación El Bosque (Distribución ajustada: LogNormal).

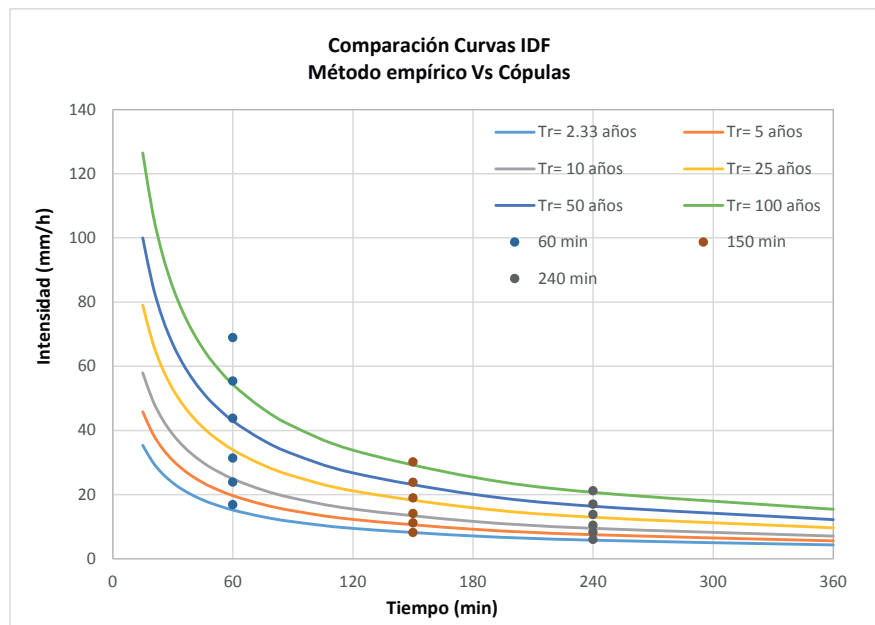


Figura 3-8.: Comparación de las curvas IDF aplicando método de la cópula y el método empírico para la estación El Bosque.

Estación El Delirio

De igual forma, se obtuvieron las gráficas de las curvas IDF mediante el método de la cópula para la estación El Delirio y se realizó la comparación con el método empírico a través de la figura 3-10. Las curvas IDF de la figura 3-9 se obtuvieron mediante la cópula Frank con distribuciones marginales Pearson Tipo III para la intensidad y Lognormal para la duración de la tormenta.

La curva IDF empírica se construyó con la ecuación propuesta por Chow et al. (1994)

$$i = \frac{cT^m}{D^e + f} = \frac{1160 T^{0,27}}{D^{0,91} + 15,91} \quad (3-59)$$

Estación Santa Lucía

Las curvas IDF mediante el método de la cópula se obtuvieron para la estación Santa Lucía con distribuciones marginales LogNormal para la intensidad y la duración. Las curvas IDF aplicando la cópula Frank, se muestra en la figura 3-11.

Para la estación Santa Lucía, se obtuvieron las gráficas de las curvas IDF por medio del método empírico, a través de la ecuación propuesta por Chow et al. (1994):

$$i = \frac{cT^m}{D^e + f} = \frac{1169,93 T^{0,25}}{D^{0,97} + 18,93} \quad (3-60)$$

Estación Bosa Barreno

Al igual que las demás estaciones, las curvas IDF mediante el método de la cópula se obtuvieron para la estación Bosa Barreno con distribuciones marginales Pearson Tipo III para la intensidad y LogNormal para la duración de la tormenta. Las curvas IDF aplicando la cópula Frank, se muestra en la figura 3-13.

Para la estación Bosa barreno, se obtuvieron las gráficas de las curvas IDF por medio del método empírico, a través de la ecuación propuesta por Chow et al. (1994):

$$i = \frac{cT^m}{D^e + f} = \frac{560,78 T^{0,24}}{D^{0,84} + 8,02} \quad (3-61)$$

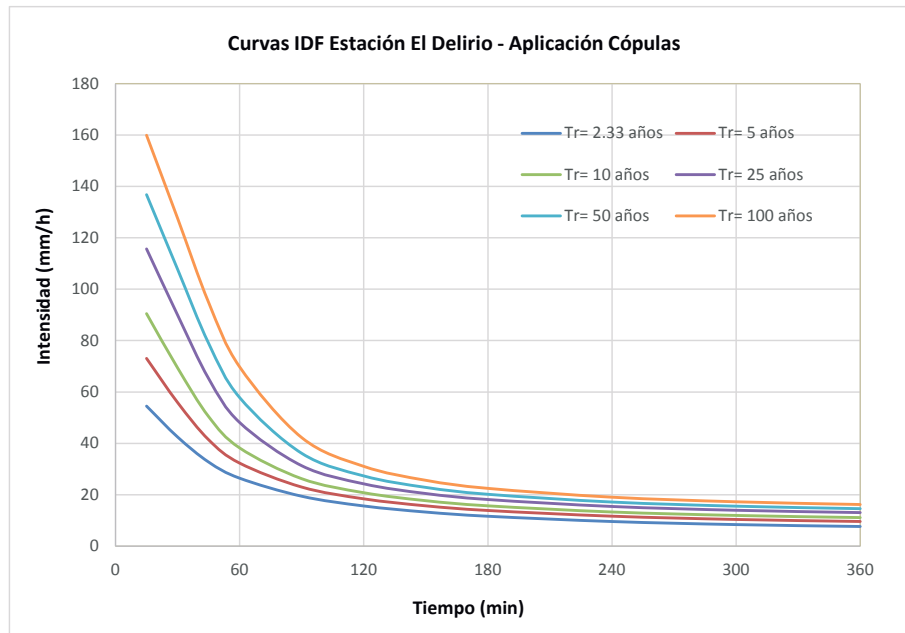


Figura 3-9.: Curvas IDF por medio de la aplicación del método de la cópula para la estación El Delirio.

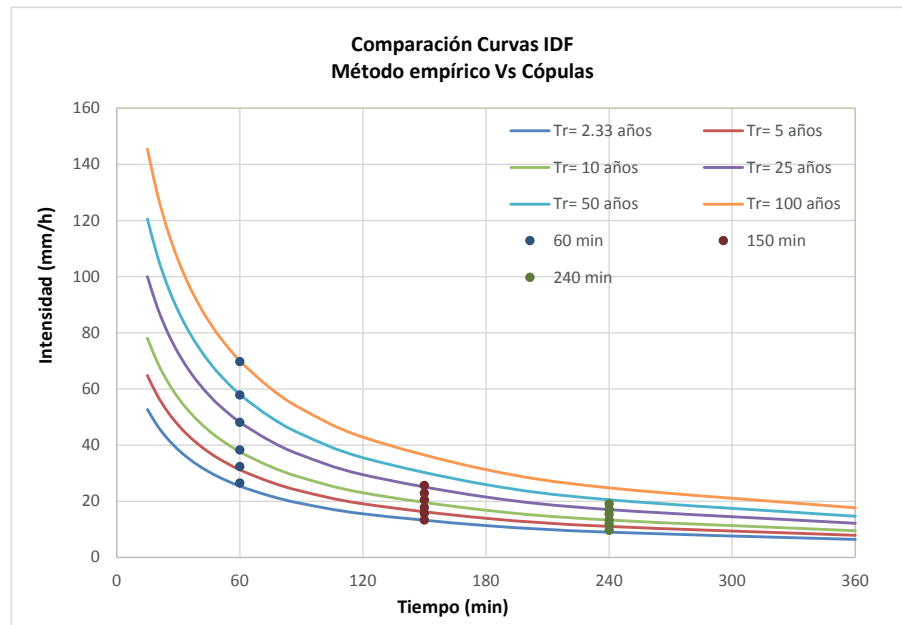


Figura 3-10.: Comparación de las curvas IDF aplicando el método de la cópula y el método empírico para la estación El Delirio.

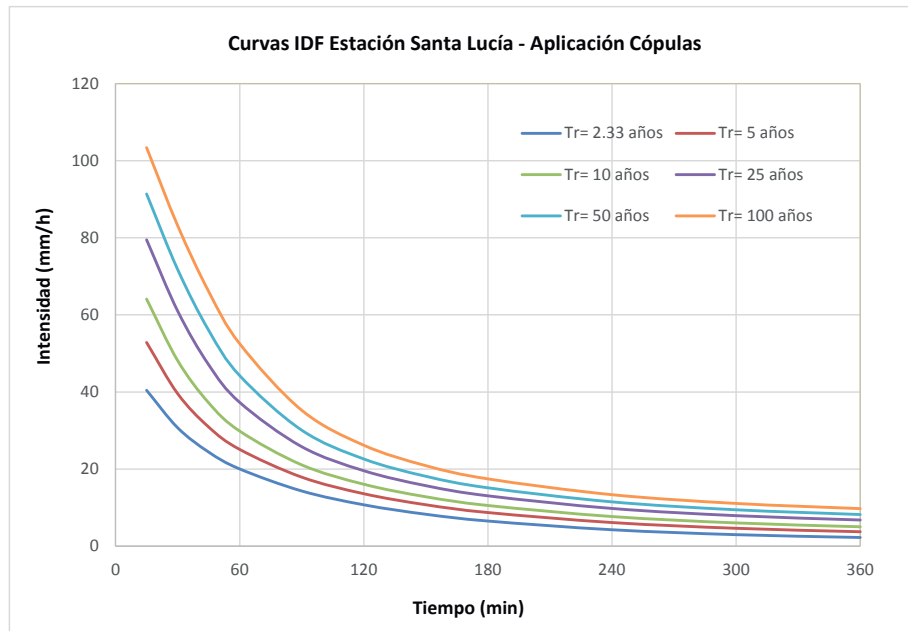


Figura 3-11.: Curvas IDF por medio de la aplicación del método de la cópula para la estación Santa Lucía.

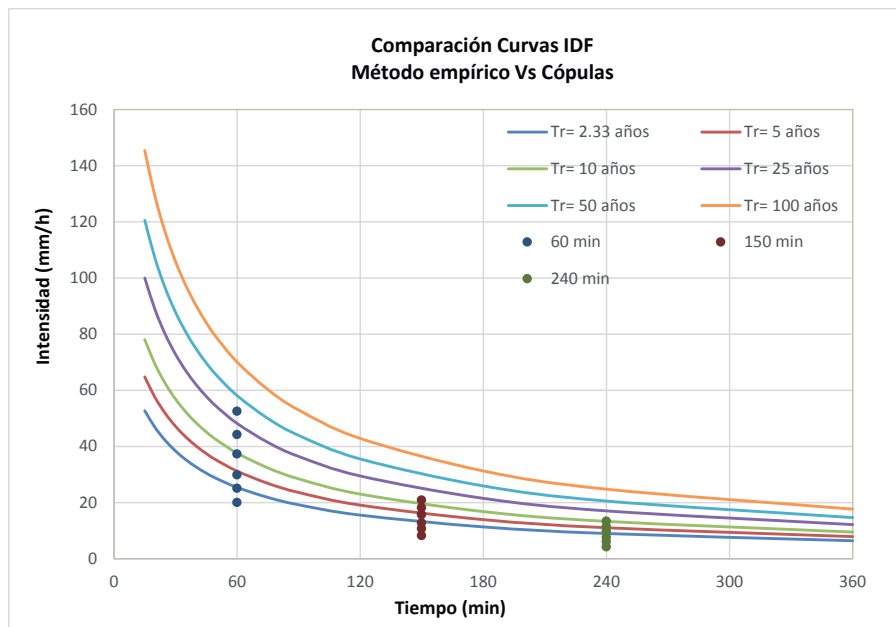


Figura 3-12.: Comparación de las curvas IDF aplicando el método de la cópula y el método empírico para la estación Santa Lucía.

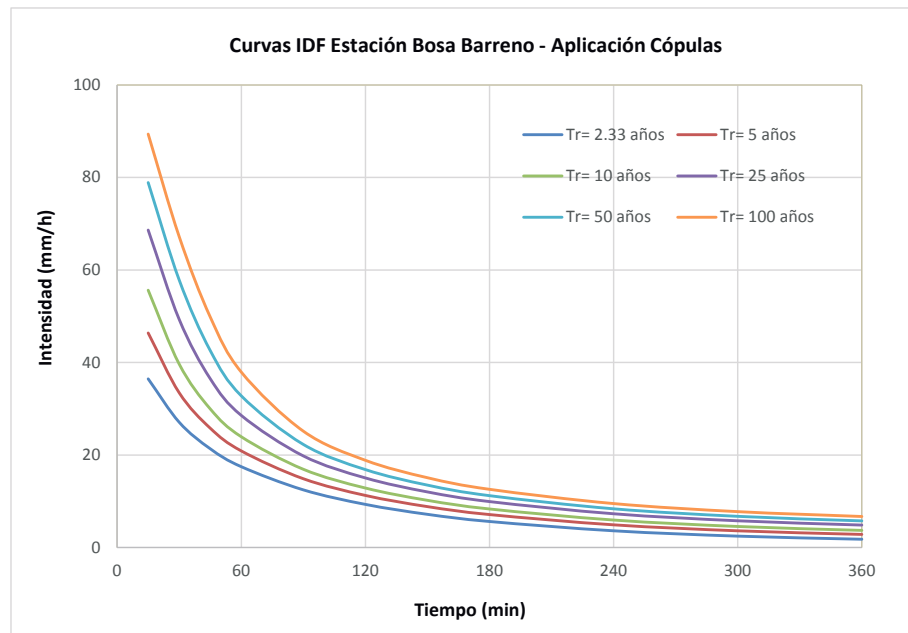


Figura 3-13.: Curvas IDF por medio de la aplicación del método de la cópula para la estación Bosa Barreno.

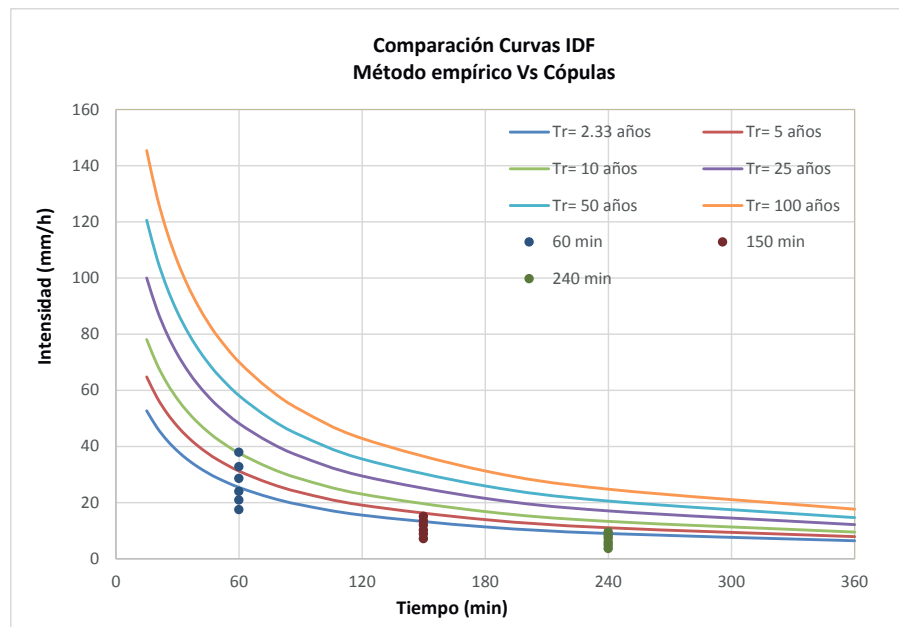


Figura 3-14.: Comparación de las curvas IDF aplicando el método de la cópula y el método empírico para la estación Bosa Barreno.

Tabla 3-2.: Comparación entre curvas IDF por medio de cópulas y Curvas IDF Empíricas

| Estación | Periodo de retorno (años) | 60 min | | | 150 min | | | 240 min | | |
|-----------------|---------------------------------|----------------------|--------------------|--------------------------|----------------------|--------------------|-------------|----------------------|--------------------|-------------|
| | | IDF | IDF | Dif. ¹ (%) | IDF | IDF | Dif. (%) | IDF | IDF | Dif. (%) |
| | | Empírica Ec. 2-16 | Cópula Ec. 3-56 | | Empírica Ec. 2-16 | Cópula Ec. 3-56 | | Empírica Ec. 2-16 | Cópula Ec. 3-56 | |
| El Bosque | 2,33 | 15,19 | 16,89 | 11,18 | 8,09 | 8,19 | 1,22 | 5,79 | 5,97 | 3,21 |
| | 5 | 19,68 | 23,91 | 21,52 | 10,48 | 11,19 | 6,78 | 7,50 | 8,28 | 10,50 |
| | 10 | 24,89 | 31,39 | 26,13 | 13,26 | 14,11 | 6,44 | 9,48 | 10,43 | 9,97 |
| | 25 | 33,95 | 43,80 | 28,99 | 18,09 | 18,95 | 4,75 | 12,94 | 13,78 | 6,56 |
| | 50 | 42,95 | 55,36 | 28,90 | 22,88 | 23,82 | 4,08 | 16,36 | 17,02 | 4,05 |
| | 100 | 54,32 | 68,91 | 26,85 | 28,94 | 30,14 | 4,13 | 20,70 | 21,20 | 2,42 |
| El Delirio | 2,33 | 25,39 | 26,45 | 4,21 | 13,08 | 13,28 | 1,53 | 8,97 | 9,60 | 6,98 |
| | 5 | 31,20 | 32,32 | 3,60 | 16,07 | 15,70 | 2,31 | 11,03 | 11,67 | 5,88 |
| | 10 | 37,62 | 38,23 | 1,62 | 19,38 | 17,70 | 8,65 | 13,30 | 13,31 | 0,09 |
| | 25 | 48,18 | 48,12 | 0,11 | 24,82 | 20,49 | 17,44 | 17,03 | 15,45 | 9,25 |
| | 50 | 58,10 | 57,83 | 0,46 | 29,93 | 22,87 | 23,56 | 20,53 | 17,18 | 16,33 |
| | 100 | 70,05 | 69,76 | 0,42 | 36,09 | 25,64 | 28,96 | 24,76 | 19,06 | 22,99 |
| Santa Lucía | 2,33 | 20,08 | 20,04 | 0,21 | 9,77 | 8,28 | 15,22 | 6,50 | 4,26 | 34,37 |
| | 5 | 24,30 | 25,10 | 3,30 | 11,82 | 10,75 | 9,09 | 7,86 | 6,13 | 21,96 |
| | 10 | 28,90 | 29,86 | 3,32 | 14,06 | 12,82 | 8,83 | 9,35 | 7,69 | 17,77 |
| | 25 | 36,34 | 37,30 | 2,66 | 17,68 | 15,70 | 11,20 | 11,76 | 9,79 | 16,75 |
| | 50 | 43,21 | 44,24 | 2,39 | 21,02 | 18,13 | 13,76 | 13,98 | 11,49 | 17,78 |
| | 100 | 51,39 | 52,52 | 2,20 | 25,00 | 20,89 | 16,44 | 16,62 | 13,36 | 19,66 |
| Bosa Barreno | 2,33 | 17,40 | 17,91 | 2,93 | 9,03 | 9,22 | 2,03 | 6,30 | 6,67 | 5,85 |
| | 5 | 20,89 | 20,58 | 1,47 | 10,84 | 10,37 | 4,40 | 7,56 | 7,71 | 1,98 |
| | 10 | 24,65 | 23,08 | 6,36 | 12,80 | 11,27 | 11,92 | 8,92 | 8,49 | 4,82 |
| | 25 | 30,69 | 27,07 | 11,79 | 15,93 | 12,46 | 21,80 | 11,11 | 9,47 | 14,73 |
| | 50 | 36,22 | 30,94 | 14,57 | 18,80 | 13,40 | 28,71 | 13,11 | 10,22 | 22,05 |
| | 100 | 42,74 | 35,85 | 16,12 | 22,19 | 14,42 | 35,00 | 15,47 | 10,99 | 28,96 |

¹ Nota: Porcentaje de diferencia = $\frac{|Intensidad\ 2-16 - Intensidad\ 3-56|}{Intensidad\ 2-16} * 100$

El porcentaje de diferencia fue calculado como:

$$Diferencia(\%) = \frac{|Intensidad\ (2 - 16) - Intensidad\ (3 - 56)|}{Intensidad\ (2 - 16)} \times 100 \quad (3-62)$$

3.13.2. Análisis de resultados de la construcción de curvas IDF

El problema de definir un modelo de probabilidad para las observaciones de dos variables dependientes como en este caso; intensidad y duración de la tormenta para una estación dada, se simplifica en gran medida mediante la distribución conjunta $H(x, y)$ en términos de sus marginales, $F_X(x)$ y $F_Y(y)$ y una función de dependencia asociada C , definida a través de $H(x, y) = C[F_X(x), F_Y(y)]$. Como se mostró en esta sección, el estudio de las distribuciones condicionales se facilita en gran medida mediante la adopción de cópulas.

De Michele and Salvadori (2003), fueron tal vez los primeros en aplicar cópulas en hidrología para analizar el comportamiento conjunto entre la duración y la intensidad de lluvia. En esta sección se mostró la aplicación para obtener curvas IDF a través del método de la cópula para cuatro estaciones diferentes, pertenecientes a la cuenca baja del río Tunjuelo.

La tabla **3-2** muestra las diferencias que se encuentran entre las curvas IDF construidas mediante el enfoque empírico y el método de la cópula para duraciones de 60, 150 y 240 minutos. Para 60 minutos se encuentra que la máxima diferencia entre los dos métodos es del 29% con un periodo de retorno de 25 años que corresponde a la estación el bosque, la cual cuenta con un registro histórico de 41 años. Para una duración de 150 minutos, la mayor diferencia se presenta en la estación Bosa Barreno para un periodo de retorno de 100 años con una diferencia entre los dos métodos de 35%. Esta estación cuenta con un registro histórico de 27 años. Para 240 minutos, se presentan una diferencia máxima entre los dos métodos de 35% para un periodo de retorno de 2,33 años que corresponde a la estación Santa Lucía, la cual cuenta con un registro histórico de 57 años. Para las tres duraciones, las diferencias de los periodos de retorno intermedios van hacia arriba y hacia abajo. Como se puede observar, estas diferencias no tienen un patrón sistemático particular.

Singh and Zhang (2007) obtienen curvas IDF por medio del método de la cópula y el método empírico para seis estaciones de precipitación en América del Norte. Compararon los dos tipos de curvas IDF y no encontraron diferencias significativas entre los dos métodos. Conclusiones similares también fueron hechas por Ariff et al. (2012) y Bezak et al. (2016), quienes también construyeron curvas IDF a través de cópulas.

3.14. Aplicación de las cópulas bivariadas Multisitio

Conceptos básicos como distribuciones de probabilidad conjunta, periodos de retorno condicional y periodos de retorno conjunto son importantes para tratar de entender los fenómenos hidrológicos bivariados y multivariados. Dentro de los ejemplos más comunes están las inundaciones y las tormentas. Un evento hidrológico bivariado o multivariado descrito por varias variables aleatorias, como sucede en la naturaleza, se debe analizar conjuntamente para obtener evaluaciones más completas.

Salvadori and De Michele (2010) introducen una nueva definición adecuada de periodo de retorno bivariado y multivariado por medio del enfoque de la cópula y muestran por qué la definición tradicional que se usa actualmente en la literatura no puede proporcionar resultados significativos. Una amplia variedad de situaciones relacionadas con los fenómenos extremos tiene un carácter multivariado inherente, como ha señalado Coles and Tawn (1991).

A continuación se presenta un ejemplo práctico para determinar el periodo de retorno conjunto de dos estaciones pertenecientes a la zona de estudio definida en la sección 2.4.1. De acuerdo con los datos hidrológicos de mayor registro histórico, las estaciones seleccionadas fueron: La estación El Delirio y la estación Santa Lucía, éstas cuentan con series hidrológicas mayores de 45 años. Antes de aplicar el método de la cópula, es necesario seleccionar una cópula adecuada para obtener el periodo de retorno conjunto. Como se vio en la sección 3.10, en este estudio se aplicaron dos métodos para encontrar la cópula de mejor ajuste.

3.14.1. Selección de la cópula en el área de estudio

1. Método Gráfico

Para aplicar el método gráfico, primero es necesario determinar el τ de Kendall como se mencionó en la sección 3.9. Este coeficiente de correlación se determinó mediante el lenguaje de programación Python (2015) (ver anexo B).

El correspondiente coeficiente de correlación entre los registros históricos de la estación El Delirio y la estación Santa Lucía es:

$$\tau = 0,6092585$$

Como el coeficiente de correlación τ de Kendall para la cópula Ali-Mikhail-Haq está en un

rango aproximado de $-0,1817 < \tau < 0,333$, esta cópula no se tuvo en cuenta para el análisis gráfico.

Las cópulas seleccionadas para este método fueron: la cópula Gumbel, la cópula Frank, la cópula Clayton y la cópula Joe. Como el método gráfico requiere de las funciones generadoras de cada cópula así como su parámetro θ , las distribuciones marginales de la intensidad para cada estación no son necesarias. El parámetro θ se obtuvo como se ilustró en la sección 3.9.

Tabla 3-3.: Parámetro θ para las estaciones El Delirio y Santa Lucía

| Cópula | θ |
|---------|----------|
| Gumbel | 2,56 |
| Frank | 8,18 |
| Clayton | 3,116 |
| Joe | 3,92 |

Mediante la aplicación del método gráfico mencionado en la sección 3.10, se obtuvieron las siguientes figuras (ver anexo B):

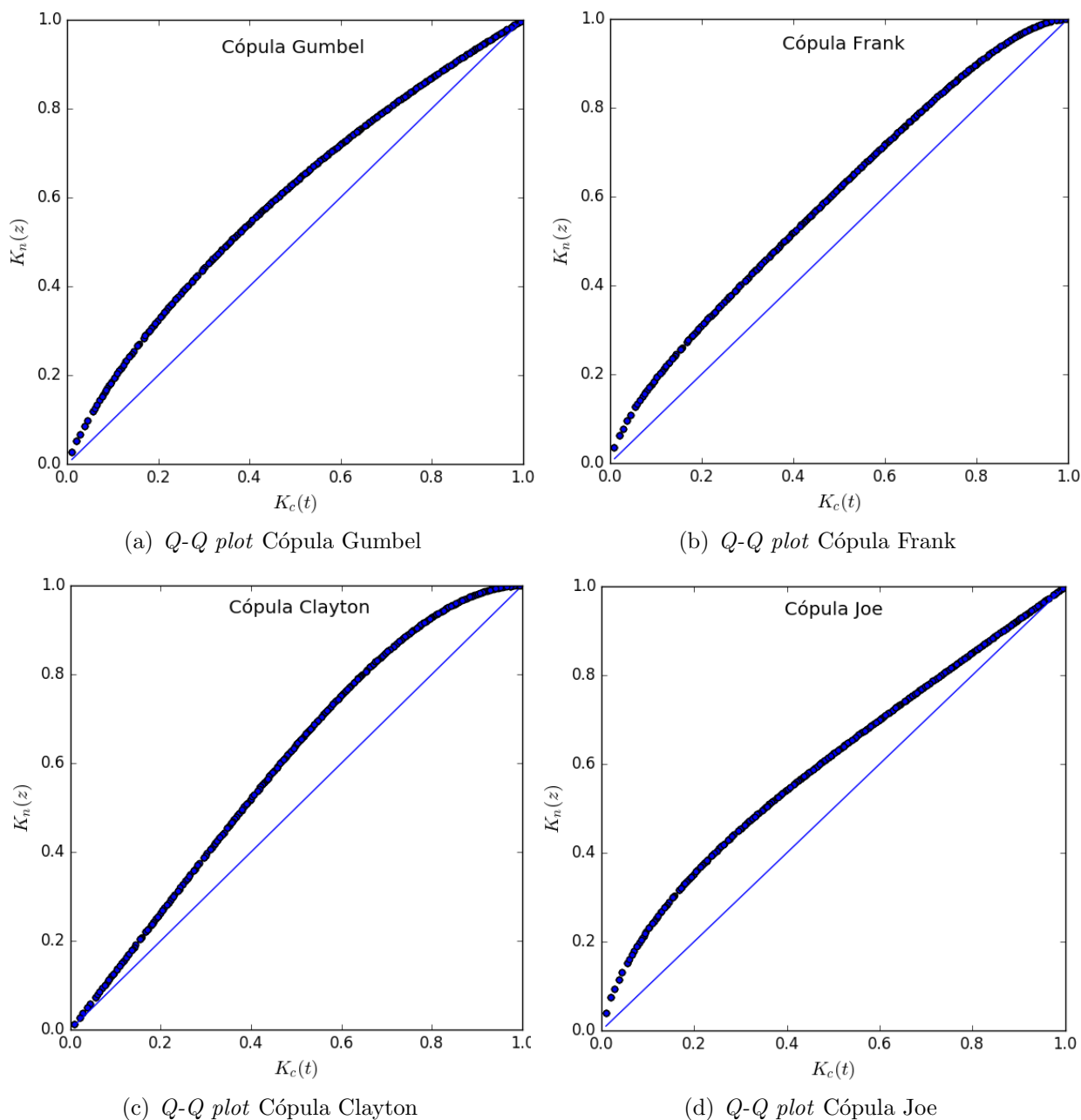


Figura 3-15.: Método gráfico para la selección de la cópula. Estimación No paramétrica contra la estimación paramétrica (método de la cópula).

En efecto, como se mencionó en la sección 3.10, si la gráfica concuerda con una línea recta que pasa por el origen en un ángulo de 45° la función generadora es satisfactoria. De lo contrario, la función cópula debe ser reidentificada. Gráficamente las cuatro cópulas analizadas están muy cerca de la línea recta en ángulo de 45° , por lo tanto, se podría concluir que

las cuatro cópulas se ajustan a los registros históricos, sin embargo, es necesario aplicar un método más completo donde se muestre y se seleccione la cópula de mejor ajuste.

2. Aplicación mediante el estadístico Cramer-von Mises usando el paquete `copula` en R

A continuación se presenta el modelo basado en el paquete “`copula`” de R mediante el estadístico Cramer-von-Mises para la selección de la cópula, mencionado en la sección 3.10.

Esta metodología se aplicó para las estaciones que cuentan con los mayores registros históricos de la cuenca baja del río Tunjuelo. Las estaciones seleccionadas fueron: La estación El Delirio y la estación Santa Lucía ya que éstas cuentan con registros históricos mayores de 45 años.

```
library("copula")
Conjunta<-read.table("Datos_Delirio_Lucia.txt", header=T)
myConjunta <- subset(Conjunta, select = c("Int1", "Int2"))
nrow(myConjunta)

## [1] 539
```

Inicialmente se cuentan 539 datos que corresponden a 49 años de registros para un evento de tormenta máximo de cada año a diferentes duraciones.

```
set.seed(123)
pseudoConjunta <- sapply(myConjunta, rank, ties.method = "random")
/(nrow(myConjunta) + 1)
pseudoConjunta.ave <- sapply(myConjunta,rank)/(nrow(myConjunta)+1)
par(mfrow = c(1,2), mgp = c(1.5, 0.5, 0), mar = c(3.5, 2.5, 0, 0))
plot(pseudoConjunta, sub = "(a) Correlacion aleatoria de filas")
plot(pseudoConjunta.ave, sub = "(b) Correlacion media de las filas")
```

La primera prueba que se realiza es una prueba multivariada de aleatoriedad seguida de una prueba de independencia mutua. Lo que se pretende es romper la presencia de lazos de cada par de datos entre filas para cada estación.

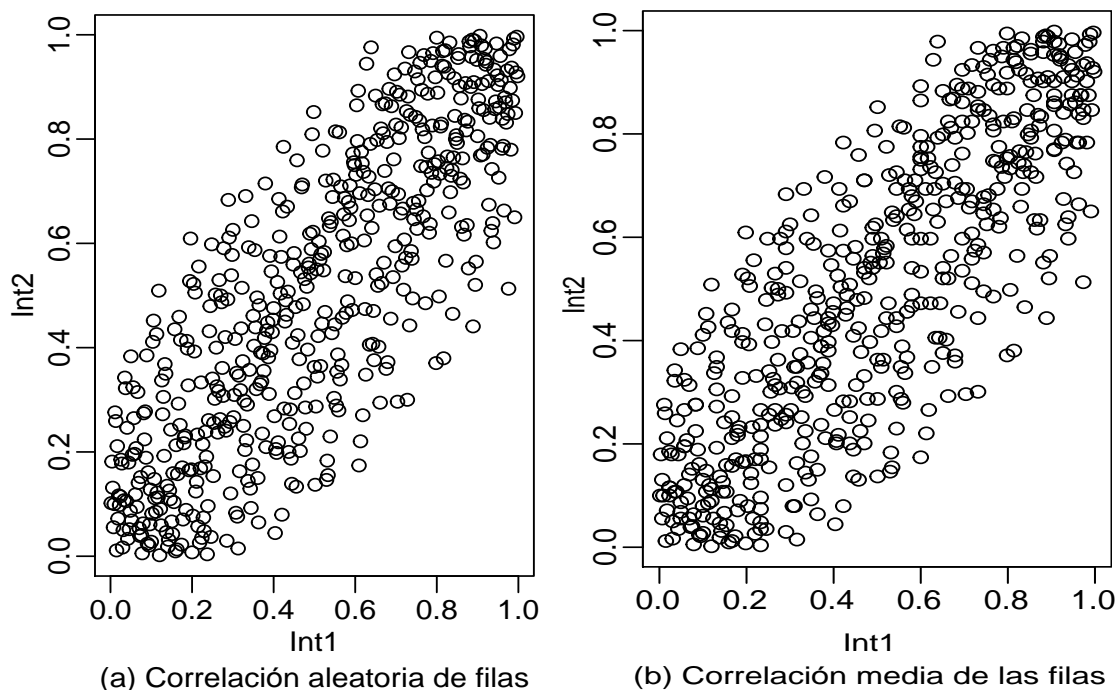


Figura 3-16.: Pseudo-observaciones construidas a partir de los datos de las estaciones El Delirio y Santa Lucía.

```
system.time(empsamp <- indepTestSim(nrow(pseudoConjunta),
  p = 2, N = 1000, verbose = 0))

##      user  system elapsed
## 39.83    0.00   39.97
```

La figura **3-16** muestra que no existe una fuerte evidencia en contra de la hipótesis nula de independencia, tiene sentido considerar diferentes familias de cópulas para modelar la dependencia entre las dos estaciones.

```
indepTest(pseudoConjunta, empsamp)

##
## Global Cramer-von Mises statistic: 3.811837 with p-value 0.0004995005
## Combined p-values from the Mobius decomposition:
## 0.0004995005 from Fisher's rule,
## 0.0004995005 from Tippett's rule.
```

El siguiente paso es llevar a cabo una prueba de bondad de ajuste como se mencionó en la sección 3.10 para las cópulas candidatas: Gumbel-Hougaard, Clayton, Frank y Joe. Para las cuatro familias en consideración, los tres métodos de estimación, “mpl” (máximo pseudo-likelihood), “itau” (inversión del tau de Kendall), y “irho” (inversión del rho de Spearman), fueron utilizados.

```
system.time(Conjunta.Gof.Gumbel.mult <- gofCopula(gumbelCopula(),
pseudoConjunta,estim.method="itau",simulation="mult", N = 1000))

##      user  system elapsed
##      4.97    0.00    4.99

Conjunta.Gof.Gumbel.mult

##
## Multiplier bootstrap goodness-of-fit test with 'method'="Sn",
## 'estim.method'="itau"
##
## data:  x
## statistic = 0.062262, parameter = 2.5639, p-value = 0.0004995

system.time(Conjunta.Gof.Frank.mult <- gofCopula(frankCopula(),
pseudoConjunta,estim.method="itau",simulation="mult", N = 1000))

##      user  system elapsed
##      5.26    0.02    5.30

Conjunta.Gof.Frank.mult

##
## Multiplier bootstrap goodness-of-fit test with 'method'="Sn",
## 'estim.method'="itau"
##
## data:  x
## statistic = 0.013805, parameter = 8.2018, p-value = 0.469

system.time(Conjunta.Gof.Clayton.mult<-gofCopula(claytonCopula(),
pseudoConjunta,estim.method="itau",simulation = "mult",N = 1000))

##      user  system elapsed
##      5.57    0.00    5.57
```

```

Conjunta.Gof.Clayton.mult

##
## Multiplier bootstrap goodness-of-fit test with 'method'="Sn",
## 'estim.method'="itau"
##
## data:  x
## statistic = 0.19378, parameter = 3.1277, p-value = 0.0004995

system.time(Conjunta.Gof.Joe.pb <- gofCopula(joeCopula(),
      pseudoConjunta,estim.method="itau", simulation = "pb",
      N = 1000, verbose = 0))

## user system elapsed
## 11.82 0.01 11.91

Conjunta.Gof.Joe.pb

##
## Parametric bootstrap goodness-of-fit test with 'method'="Sn",
## 'estim.method'="itau"
##
## data:  x
## statistic = 0.12979, parameter = 3.9514, p-value = 0.0004995

```

Tabla 3-4.: Resumen de la prueba de bondad de ajuste para las familias candidatas. Parámetro θ para las estaciones El Delirio y Santa Lucía

| Familias de Cópulas | Estadístico Sn | p - valor | θ |
|---------------------|----------------|-----------|----------|
| Gumbel | 0,06226 | 0,0004995 | 2,564 |
| Frank | 0,01380 | 0,469 | 8,202 |
| Clayton | 0,19378 | 0,0004995 | 3,128 |
| Joe | 0,12979 | 0,0004995 | 3,951 |

La tabla 3-4 presenta un resumen de los resultados obtenidos aplicando el modelo mencionado en la sección 3.10 para la prueba de bondad de ajuste mediante el estadístico S_n . La cópula de mejor ajuste es aquella que presenta el p-valor más alto, además entre más bajo el p-valor más independientes son los datos de registros históricos de las estaciones. Es claro que el área de influencia de las dos estaciones hacen parte de la cuenca baja del río Tunjuleo,

razón por la cual no hay garantía de que sean independientes, es por esto que entre más grande el p-valor, más cercana es la cópula de mejor ajuste. En este caso la cópula que mejor se ajusta a los datos conjuntos de las estaciones El delirio y Santa Lucía es la cópula Frank. Así mismo se realiza una última prueba para calcular el error estandar de la estimación del parámetro θ :

```
fitCopula(frankCopula(), pseudoConjunta, method = "itau")

## fitCopula() estimation based on 'inversion of Kendall's tau'
## and a sample of size 539.
##           Estimate Std. Error
## param      8.202      0.222
```

3.14.2. Periodo de retorno conjunto multisitio

Una vez obtenido el parámetro de la cópula de mejor ajuste, se encuentran los periodos de retorno conjunto para las dos estaciones mencionadas anteriormente (El delirio y Santa Lucía). Como se vio en la sección 3.11, el periodo de retorno bivariado conjunto está dado por:

$$T_{X,Y}(x, y) = \frac{1}{1 - H_{X,Y}(x, y)}$$

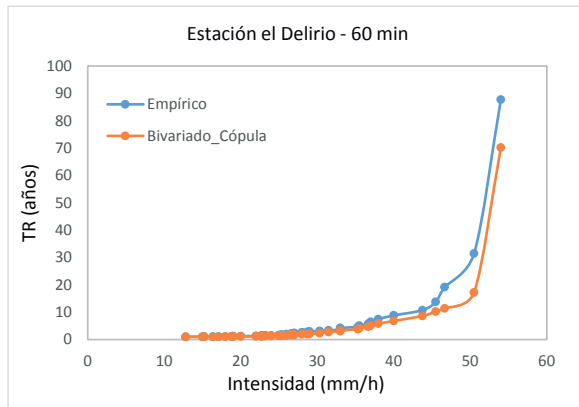
donde

$$H_{X,Y}(x, y) = -\frac{1}{\theta} \ln \left(\frac{[\exp(-\theta) - 1] + [\exp(-\theta F_X(x)) - 1] [\exp(-\theta F_Y(y)) - 1]}{\exp(-\theta) - 1} \right)$$

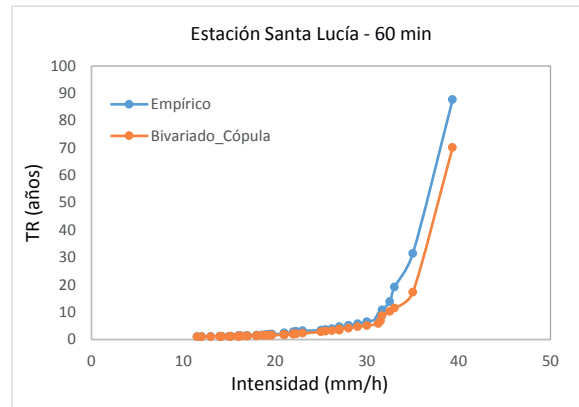
$F_X(x)$ = Probabilidad acumulada de la intensidad de la estación El Delirio

$F_Y(y)$ = Probabilidad acumulada de la intensidad de la estación Santa Lucía

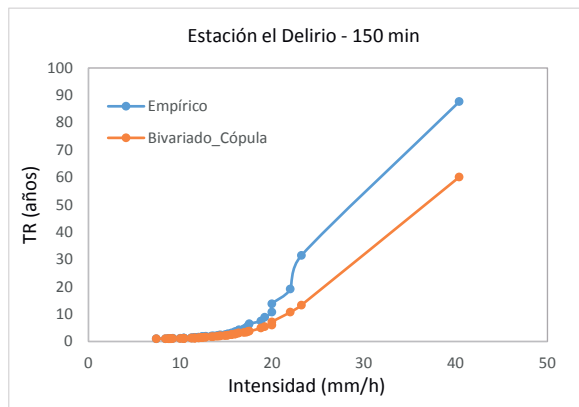
La figura **3-17** muestra los resultados obtenidos para el periodo de retorno conjunto multisitio al aplicar la metodología descrita anteriormente para la cópula Frank. De igual forma, se realizó el mismo procedimiento para las estaciones Quiba y Juan Rey ya que presentan el mismo registro histórico. (Ver figura **3-18**).



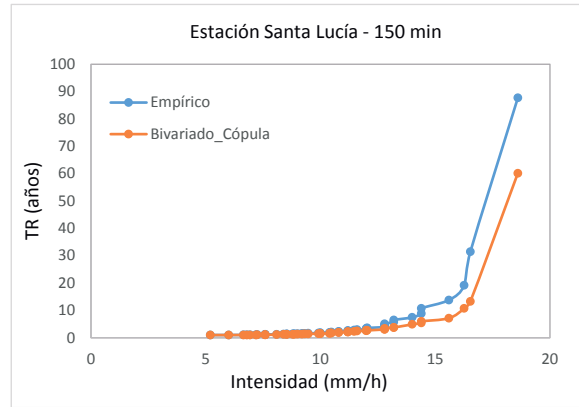
(a) T_R multisitio empírico y conjunto para una duración de 60 minutos.



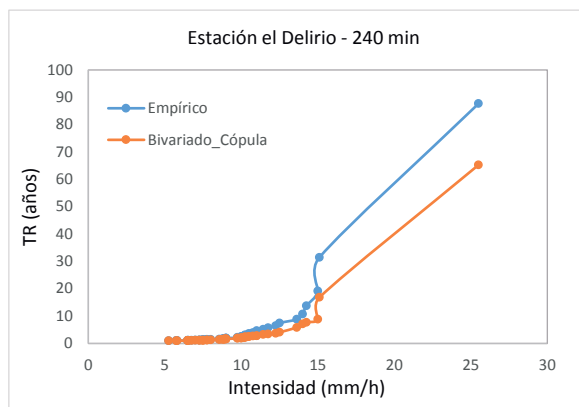
(b) T_R multisitio empírico y conjunto para una duración de 60 minutos.



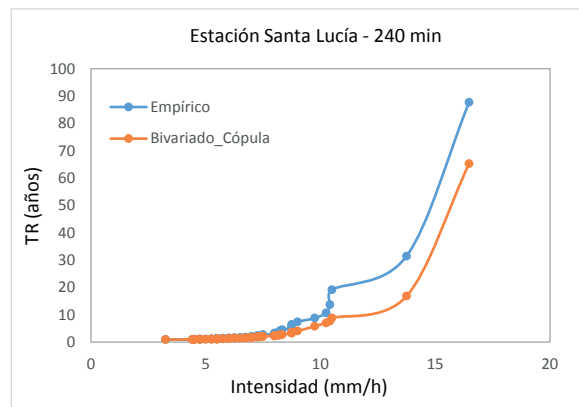
(c) T_R multisitio empírico y conjunto para una duración de 150 minutos.



(d) T_R multisitio empírico y conjunto para una duración de 150 minutos.

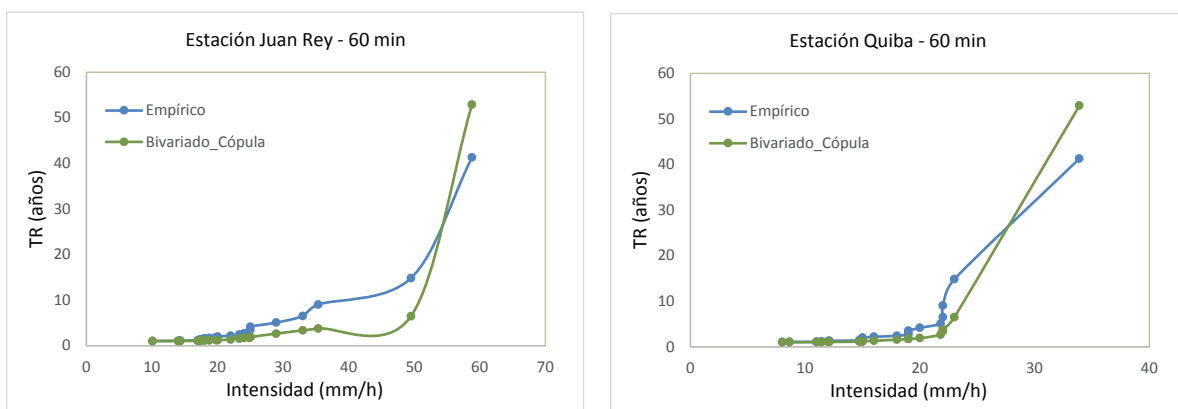


(e) T_R multisitio empírico y conjunto para una duración de 240 minutos.

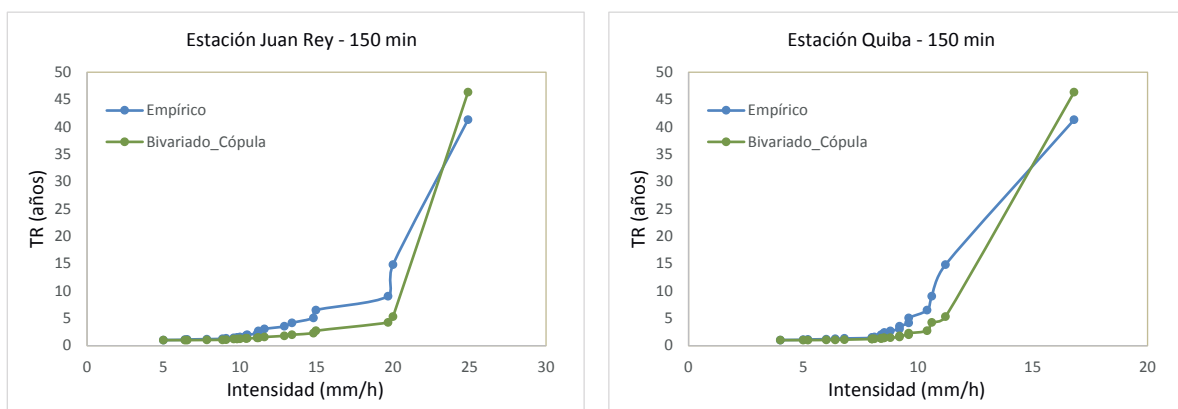


(f) T_R multisitio empírico y conjunto para una duración de 240 minutos.

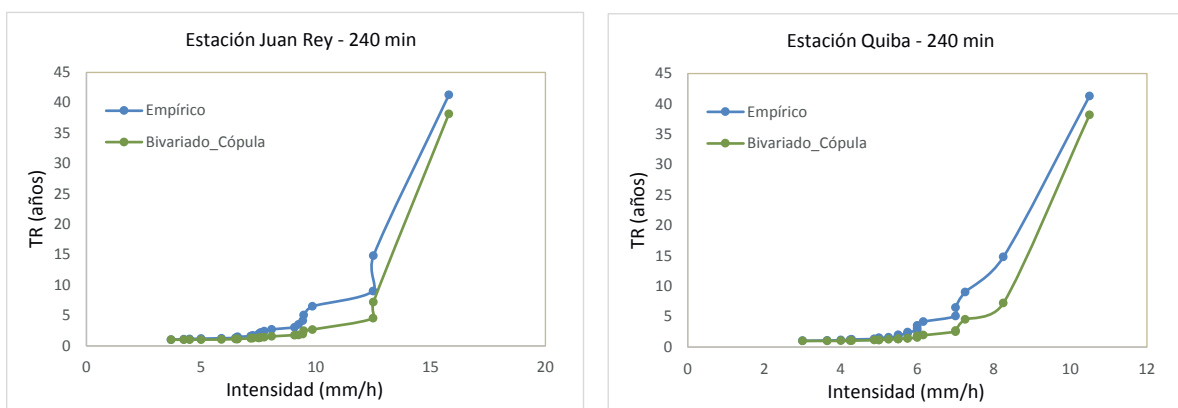
Figura 3-17.: Comparación entre el Periodo de retorno empírico y el periodo de retorno conjunto multisitio de las estaciones El Delirio y Santa Lucía para diferentes duraciones.



(a) T_R multisitio empírico y conjunto para una duración de 60 minutos. (b) T_R multisitio empírico y conjunto para una duración de 60 minutos.



(c) T_R multisitio empírico y conjunto para una duración de 150 minutos. (d) T_R multisitio empírico y conjunto para una duración de 150 minutos.



(e) T_R multisitio empírico y conjunto para una duración de 240 minutos. (f) T_R multisitio empírico y conjunto para una duración de 240 minutos.

Figura 3-18.: Comparación entre el Periodo de retorno empírico y el periodo de retorno conjunto multisitio de las estaciones Juan Rey y Quiba para diferentes duraciones.

3.14.3. Análisis de resultados de la aplicación de cópulas bivariadas multisitio

Como ya se ha mencionado, el periodo de retorno de un evento dado por lo general se define como el tiempo transcurrido medio entre dos realizaciones sucesivas del evento en sí. En la mayoría de situaciones, el análisis del período de retorno se refiere a casos univariados; desafortunadamente, esto puede conducir a una sobre o subestimación del riesgo asociado a un evento determinado, ejemplos de estas situaciones se encuentran en Raynal-Villasenor and Salas (1987) y De Michele et al. (2005).

Como se puede observar en las figuras **3-17** y **3-18**, los periodos de retorno conjunto son similares a los periodos de retorno empíricos determinados por medio de la posición de graficación Gringorten mencionada en la Tabla **2-1**. De estas dos figuras se puede ver que el periodo de retorno conjunto mediante la aplicación del método de la cópula tiende a disminuir a medida que aumenta la duración de la tormenta, lo cual puede conducir a sobreestimaciones del riesgo asociados al evento hidrológico, Esto resalta aún más la necesidad de realizar un análisis multivariado para entender el comportamiento de los eventos extremos de precipitación.

En efecto, se demuestra que los fenómenos hidrológicos se caracterizan por el comportamiento conjunto de varias variables aleatorias que por lo general no son independientes, en este caso, la intensidad y la duración para dos estaciones diferentes pero con el mismo registro histórico de datos. Como consecuencia, los eventos extremos, deben ser definidos en términos de dos o más variables. Como se mencionará en el capítulo 4, el método de la cópula puede ser fácilmente generalizable al caso multivariado.

Es importante resaltar que para determinar el periodo de retorno conjunto multisitio, es necesario aplicar el método de la cópula para una región hidrológicamente homogénea, por esta razón se escogieron dos estaciones pertenecientes a una misma cuenca, en este caso se seleccionaron las estaciones pertenecientes a la cuenca baja del río Tunjuelo. Teniendo en cuenta que las cópulas evalúan la simultaneidad de ocurrencia, se puede concluir que un evento de tormenta no necesariamente debe ocurrir el mismo día en cada estación, es claro que puede estar lloviendo en una estación y en la otra no, por esta razón se escoge el evento máximo de una tormenta que ocurra al año.

3.15. Aplicación de cópulas bivariadas en el análisis de series históricas

En muchas aplicaciones hidrológicas, como el diseño de drenaje urbano o el diseño de estructuras hidráulicas, se requiere de registros históricos de lluvia a diferentes escalas de tiempo, sin embargo, en algunas ocasiones no se cuenta con información suficiente y no se tienen series de tiempo de longitud apropiada. En tales casos, la elección de una cópula adecuada resulta problemático. Cuando los datos son escasos, las investigaciones se centran en el desarrollo de técnicas que hacen uso óptimo de los datos disponibles.

El objetivo de esta sección es analizar el tamaño de la muestra de una serie histórica, de tal forma que al ir aumentando el registro de los datos hasta llegar al tamaño actual de la muestra se pueda evaluar el desempeño de la cópula. Esto se logra dividiendo el conjunto de datos en intervalos de tiempo establecidos. Para la estación Santa Lucía que cuenta con el mayor registro histórico de la cuenca baja del río Tunjuelo, se establecieron los siguientes intervalos: [20, 30, 35, 40, 50 y 57] años. Estos diferentes escenarios pueden ser útiles para la evaluación del riesgo asociado a los problemas hidrológicos. La incorporación de este tipo de análisis verificará si una cópula particular, sobrestima o subestima el período de retorno.

Como se mostró en la sección 3.13.1, la cópula Frank es recomendada para la construcción de curvas IDF ya que la dependencia entre las variables aleatorias intensidad y duración presentan correlación negativa. Este tipo de comportamiento se representa adecuadamente mediante la cópula frank que tiene la siguiente expresión (ver sección 3.7.3):

$$C_{\theta}(u, v) = -\frac{1}{\theta} \ln \left(1 + \frac{(\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}{\exp(-\theta) - 1} \right), \quad \theta \neq 0$$

Esta expresión también puede ser escrita como:

$$C_{\theta}(u, v) = -\frac{1}{\theta} \ln \left(\frac{(\exp(-\theta) - 1) + (\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}{\exp(-\theta) - 1} \right), \quad \theta \neq 0 \quad (3-63)$$

donde:

$u = F_X(x) =$ Probabilidad acumulada de la intensidad de la lluvia (mm/h)

$v = F_Y(y) =$ Probabilidad acumulada de la duración de la tormenta (min)

En términos de sus distribuciones marginales, la función de distribución conjunta es:

$$H(x, y) = C[F_X(x), F_Y(y)] = -\frac{1}{\theta} \ln \left(\frac{[\exp(-\theta) - 1] + [\exp(-\theta F_X(x)) - 1][\exp(-\theta F_Y(y)) - 1]}{\exp(-\theta) - 1} \right)$$

Las curvas IDF se obtienen siguiendo el procedimiento descrito en la sección 3.13.1:

El periodo de retorno condicional correspondiente, puede ser expresado como se mencionó en la sección 3.12 (caso 1):

$$T(I | D = d) = \frac{1}{1 - C_{I|D=d}}$$

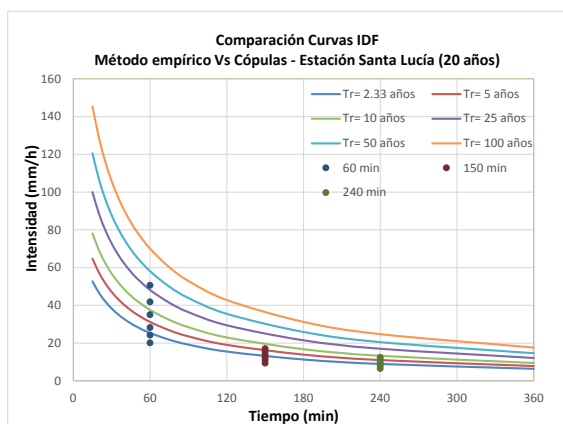
$$T(I | D = d) = \frac{1}{1 - \frac{[\exp(-\theta i) - 1][\exp(-\theta d)]}{[\exp(-\theta d) - 1][\exp(-\theta i) - 1] + [\exp(-\theta) - 1]}} \quad (3-64)$$

La expresión 3-64, es la curva IDF.

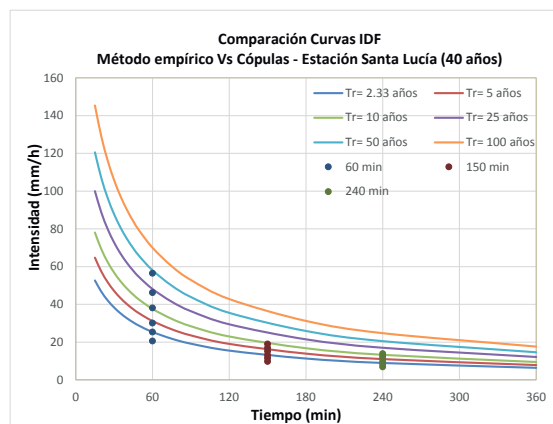
El parámetro θ de la cópula se determinó mediante el paquete “copula” de R, tal como se ilustró en la sección 3.14.1. Los resultados obtenidos al aplicar el estadístico Cramer-von Mises son los siguientes:

Tabla 3-5.: Parámetro θ para diferentes series históricas de la estación Santa Lucía

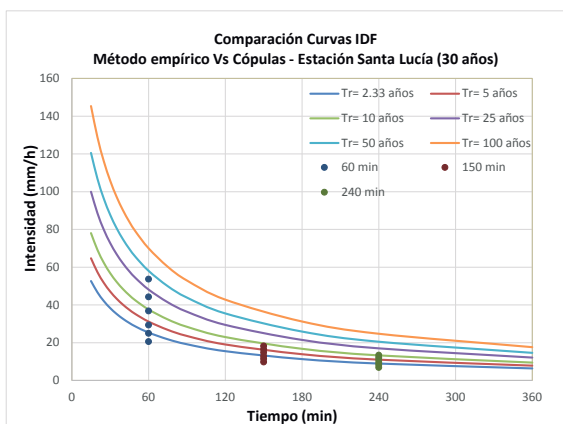
| Serie de tiempo (años) | τ | θ | p-valor | Error |
|---------------------------|---------|----------|---------|-------|
| 20 | -0,7596 | -14,786 | 0,7507 | 0,044 |
| 30 | -0,7444 | -13,781 | 0,7468 | 0,032 |
| 35 | -0,7406 | -13,549 | 0,7657 | 0,030 |
| 40 | -0,7318 | -13,031 | 0,8676 | 0,028 |
| 50 | -0,7405 | -13,540 | 0,9146 | 0,027 |
| 57 | -0,7383 | -13,410 | 0,9100 | 0,026 |



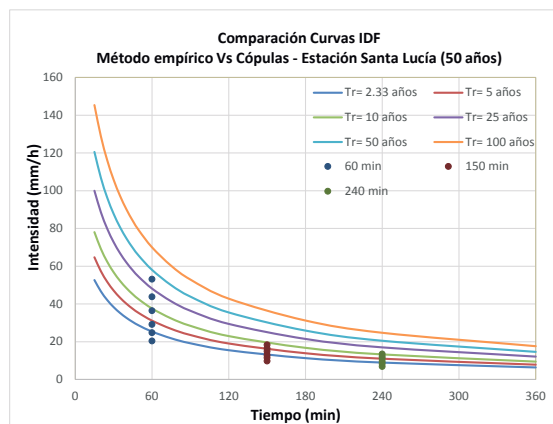
(a) Curvas IDF empíricas y mediante cópulas para una serie histórica de 20 años.



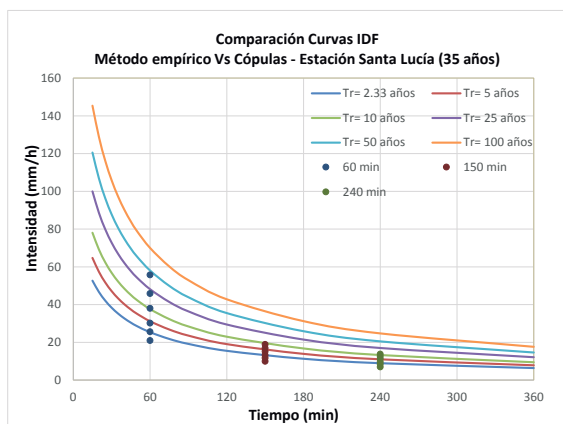
(b) Curvas IDF empíricas y mediante cópulas para una serie histórica de 40 años.



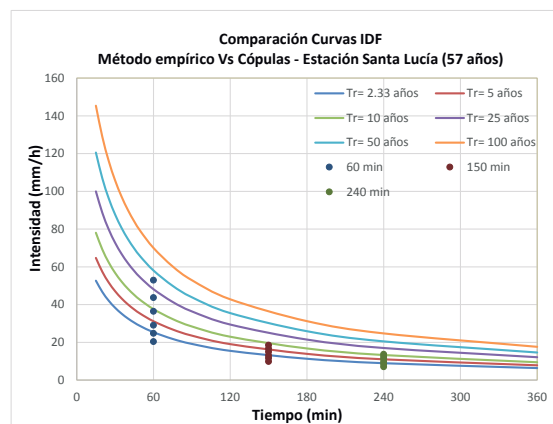
(c) Curvas IDF empíricas y mediante cópulas para una serie histórica de 30 años.



(d) Curvas IDF empíricas y mediante cópulas para una serie histórica de 50 años.



(e) Curvas IDF empíricas y mediante cópulas para una serie histórica de 35 años.



(f) Curvas IDF empíricas y mediante cópulas para una serie histórica de 57 años.

Figura 3-19.: Comparación de las curvas IDF aplicando el método de la cópula y el método empírico para diferentes series históricas para la estación Santa Lucía.

Tabla 3-6.: Comparación entre curvas IDF por medio de cópulas y Curvas IDF Empíricas para la estación Santa Lucía

| Serie de Tiempo (años) | Periodo de retorno (años) | 60 min | | | 150 min | | | 240 min | | |
|------------------------|---------------------------|-------------------|-----------------|-----------------------|-------------------|-----------------|----------|-------------------|-----------------|----------|
| | | IDF | IDF | Dif. ¹ (%) | IDF | IDF | Dif. (%) | IDF | IDF | Dif. (%) |
| | | Empírica Ec. 3-60 | Cópula Ec. 3-64 | | Empírica Ec. 3-60 | Cópula Ec. 3-64 | | Empírica Ec. 3-60 | Cópula Ec. 3-64 | |
| 20 | 2,33 | 20,08 | 20,11 | 0,17 | 9,77 | 9,38 | 3,95 | 6,50 | 6,50 | 0,09 |
| | 5 | 24,30 | 24,19 | 0,45 | 11,82 | 10,95 | 7,33 | 7,86 | 7,84 | 0,26 |
| | 10 | 28,90 | 28,23 | 2,31 | 14,06 | 12,24 | 12,96 | 9,35 | 8,88 | 4,99 |
| | 25 | 36,34 | 34,99 | 3,70 | 17,68 | 13,98 | 20,92 | 11,76 | 10,23 | 12,98 |
| | 50 | 43,21 | 41,80 | 3,26 | 21,02 | 15,42 | 26,64 | 13,98 | 11,29 | 19,23 |
| | 100 | 51,39 | 50,59 | 1,56 | 25,00 | 17,04 | 31,84 | 16,62 | 12,42 | 25,27 |
| 30 | 2,33 | 20,08 | 20,54 | 2,29 | 9,77 | 9,73 | 0,36 | 6,50 | 6,81 | 4,90 |
| | 5 | 24,30 | 24,93 | 2,61 | 11,82 | 11,45 | 3,16 | 7,86 | 8,27 | 5,22 |
| | 10 | 28,90 | 29,33 | 1,51 | 14,06 | 12,86 | 8,55 | 9,35 | 9,41 | 0,68 |
| | 25 | 36,34 | 36,75 | 1,14 | 17,68 | 14,79 | 16,33 | 11,76 | 10,90 | 7,29 |
| | 50 | 43,21 | 44,20 | 2,28 | 21,02 | 16,41 | 21,92 | 13,98 | 12,08 | 13,59 |
| | 100 | 51,39 | 53,67 | 4,44 | 25,00 | 18,25 | 26,98 | 16,62 | 13,35 | 19,67 |
| 35 | 2,33 | 20,08 | 20,97 | 4,44 | 9,77 | 9,90 | 1,39 | 6,50 | 6,92 | 6,61 |
| | 5 | 24,30 | 25,55 | 5,14 | 11,82 | 11,68 | 1,17 | 7,86 | 8,44 | 7,30 |
| | 10 | 28,90 | 30,16 | 4,38 | 14,06 | 13,15 | 6,44 | 9,35 | 9,62 | 2,91 |
| | 25 | 36,34 | 37,96 | 4,47 | 17,68 | 15,18 | 14,14 | 11,76 | 11,17 | 4,97 |
| | 50 | 43,21 | 45,80 | 5,98 | 21,02 | 16,89 | 19,67 | 13,98 | 12,41 | 11,25 |
| | 100 | 51,39 | 55,75 | 8,49 | 25,00 | 18,83 | 24,66 | 16,62 | 13,74 | 17,33 |
| 40 | 2,33 | 20,08 | 20,57 | 2,44 | 9,77 | 9,71 | 0,55 | 6,50 | 6,80 | 4,77 |
| | 5 | 24,30 | 25,26 | 3,97 | 11,82 | 11,54 | 2,41 | 7,86 | 8,34 | 6,10 |
| | 10 | 28,90 | 30,02 | 3,88 | 14,06 | 13,05 | 7,19 | 9,35 | 9,55 | 2,18 |
| | 25 | 36,34 | 38,10 | 4,84 | 17,68 | 15,15 | 14,31 | 11,76 | 11,15 | 5,17 |
| | 50 | 43,21 | 46,19 | 6,90 | 21,02 | 16,94 | 19,43 | 13,98 | 12,43 | 11,09 |
| | 100 | 51,39 | 56,42 | 9,79 | 25,00 | 18,99 | 24,02 | 16,62 | 13,83 | 16,83 |
| 57 | 2,33 | 20,08 | 20,04 | 0,21 | 9,77 | 8,28 | 15,22 | 6,50 | 4,26 | 34,37 |
| | 5 | 24,30 | 25,10 | 3,30 | 11,82 | 10,75 | 9,09 | 7,86 | 6,13 | 21,96 |
| | 10 | 28,90 | 29,86 | 3,32 | 14,06 | 12,82 | 8,83 | 9,35 | 7,69 | 17,77 |
| | 25 | 36,34 | 37,30 | 2,66 | 17,68 | 15,70 | 11,20 | 11,76 | 9,79 | 16,75 |
| | 50 | 43,21 | 44,24 | 2,39 | 21,02 | 18,13 | 13,76 | 13,98 | 11,49 | 17,78 |
| | 100 | 51,39 | 52,52 | 2,20 | 25,00 | 20,89 | 16,44 | 16,62 | 13,36 | 19,66 |

¹ Nota: Porcentaje de diferencia = $[(\text{Intensidad 3-60} - \text{Intensidad 3-64}) / \text{Intensidad 3-60}] * 100$

El porcentaje de diferencia fue calculado como:

$$Diferencia(\%) = \frac{|Intensidad(3-60) - Intensidad(3-64)|}{Intensidad(3-60)} \times 100 \quad (3-65)$$

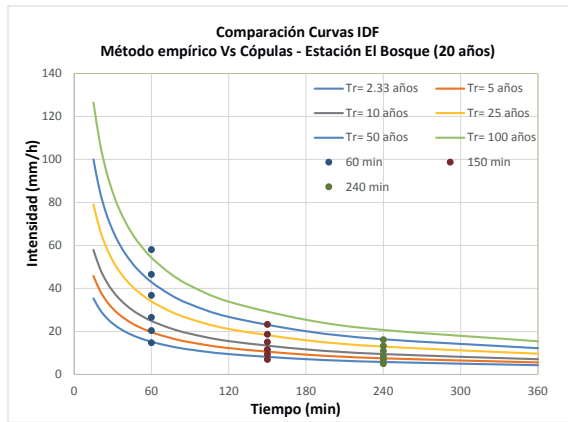
La figura **3-19** muestra las curvas IDF obtenidas para las diferentes series históricas establecidas de la estación Santa Lucía. La tabla **3-6** muestra el porcentaje de diferencia entre las curvas IDF obtenidas por medio del método empírico y las obtenidas a partir del método de la cópula.

De igual forma que para la estación Santa Lucía, se obtuvieron las curvas IDF empíricas y por medio de cópulas para la estación El Bosque para diferentes series históricas. Los resultados obtenidos para esta estación se presentan a continuación:

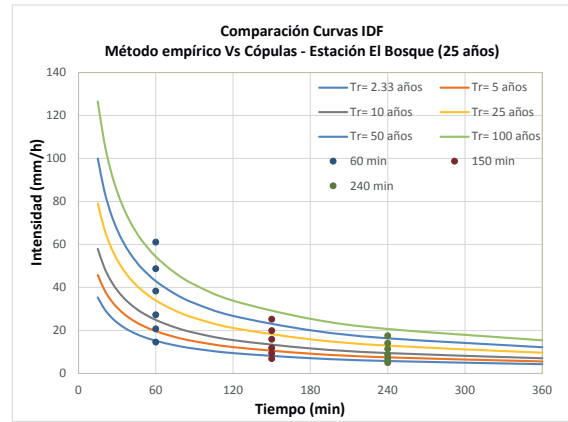
Tabla 3-7.: Parámetro θ para diferentes series históricas de la estación El Bosque

| Serie de tiempo (años) | τ | θ | p-valor | Error |
|------------------------|--------|----------|---------|-------|
| 20 | -0,589 | -7,651 | 0,0435 | 0,049 |
| 25 | -0,569 | -7,159 | 0,1364 | 0,037 |
| 30 | -0,562 | -6,983 | 0,0245 | 0,035 |
| 35 | -0,559 | -6,922 | 0,0405 | 0,032 |
| 40 | -0,561 | -6,978 | 0,0145 | 0,029 |

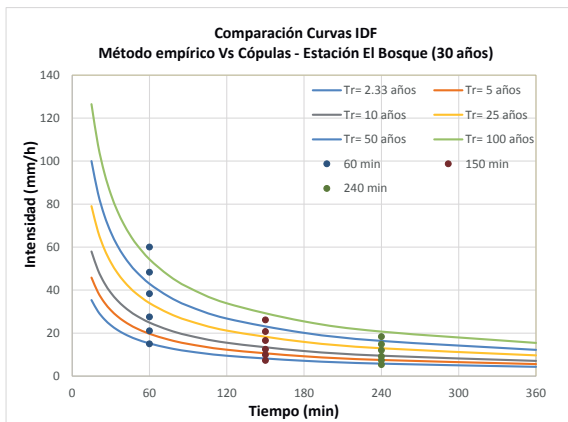
La figura **A-8** muestra las curvas IDF obtenidas para las diferentes series históricas establecidas de la estación El Bosque. La tabla **3-8** muestra el porcentaje de diferencia entre las curvas IDF obtenidas por medio del método empírico y las obtenidas a partir del método de la cópula.



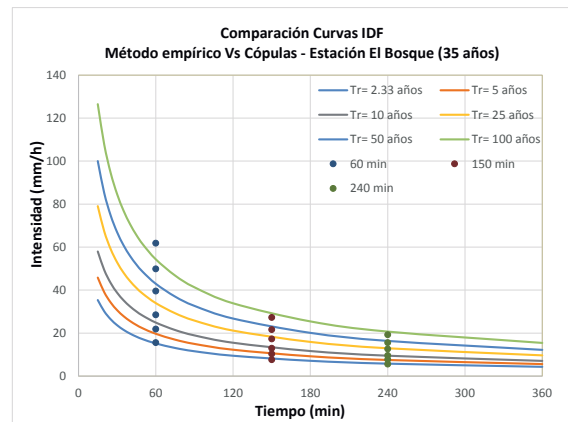
(a) Curvas IDF empíricas y mediante cópulas para una serie histórica de 20 años.



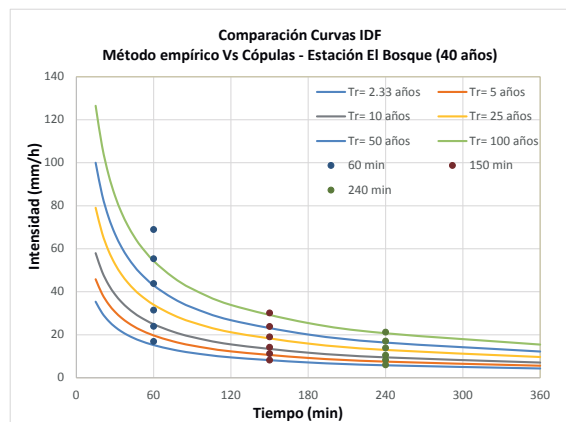
(b) Curvas IDF empíricas y mediante cópulas para una serie histórica de 25 años.



(c) Curvas IDF empíricas y mediante cópulas para una serie histórica de 30 años.



(d) Curvas IDF empíricas y mediante cópulas para una serie histórica de 35 años.



(e) Curvas IDF empíricas y mediante cópulas para una serie histórica de 40 años.

Figura 3-20.: Comparación de las curvas IDF aplicando el método de la cópula y el método empírico para diferentes series históricas para la estación El Bosque.

Tabla 3-8.: Comparación entre curvas IDF por medio de cópulas y Curvas IDF Empíricas para la estación El Bosque

| Serie de Tiempo (años) | Periodo de retorno (años) | 60 min | | | 150 min | | | 240 min | | |
|------------------------------|---------------------------------|----------------------|--------------------|-------------|----------------------|--------------------|-------------|----------------------|--------------------|-------------|
| | | IDF | IDF | Dif. (%) | IDF | IDF | Dif. (%) | IDF | IDF | Dif. (%) |
| | | Empírica Ec. 3-58 | Cópula Ec. 3-64 | | Empírica Ec. 3-58 | Cópula Ec. 3-64 | | Empírica Ec. 3-58 | Cópula Ec. 3-64 | |
| 20 | 2,33 | 15,19 | 14,68 | 3,32 | 8,09 | 6,97 | 13,89 | 5,79 | 5,01 | 13,39 |
| | 5 | 19,68 | 20,40 | 3,66 | 10,48 | 9,30 | 11,26 | 7,50 | 6,82 | 9,04 |
| | 10 | 24,89 | 26,50 | 6,49 | 13,26 | 11,51 | 13,19 | 9,48 | 8,45 | 10,93 |
| | 25 | 33,95 | 36,77 | 8,31 | 18,09 | 15,08 | 16,65 | 12,94 | 10,91 | 15,64 |
| | 50 | 42,95 | 46,49 | 8,25 | 22,88 | 18,62 | 18,64 | 16,36 | 13,22 | 19,19 |
| | 100 | 54,32 | 57,99 | 6,74 | 28,94 | 23,23 | 19,74 | 20,70 | 16,14 | 22,03 |
| 25 | 2,33 | 15,19 | 14,58 | 3,98 | 8,09 | 6,88 | 15,00 | 5,79 | 4,95 | 14,52 |
| | 5 | 19,68 | 20,69 | 5,15 | 10,48 | 9,39 | 10,41 | 7,50 | 6,87 | 8,39 |
| | 10 | 24,89 | 27,25 | 9,51 | 13,26 | 11,83 | 10,81 | 9,48 | 8,64 | 8,85 |
| | 25 | 33,95 | 38,29 | 12,77 | 18,09 | 15,85 | 12,38 | 12,94 | 11,41 | 11,83 |
| | 50 | 42,95 | 48,71 | 13,42 | 22,88 | 19,91 | 12,99 | 16,36 | 14,06 | 14,07 |
| | 100 | 54,32 | 61,04 | 12,36 | 28,94 | 25,23 | 12,83 | 20,70 | 17,47 | 15,57 |
| 30 | 2,33 | 15,19 | 14,95 | 1,57 | 8,09 | 7,27 | 10,11 | 5,79 | 5,31 | 8,27 |
| | 5 | 19,68 | 21,04 | 6,96 | 10,48 | 9,88 | 5,77 | 7,50 | 7,32 | 2,37 |
| | 10 | 24,89 | 27,52 | 10,58 | 13,26 | 12,39 | 6,53 | 9,48 | 9,17 | 3,27 |
| | 25 | 33,95 | 38,26 | 12,68 | 18,09 | 16,54 | 8,59 | 12,94 | 12,05 | 6,83 |
| | 50 | 42,95 | 48,26 | 12,37 | 22,88 | 20,69 | 9,58 | 16,36 | 14,81 | 9,45 |
| | 100 | 54,32 | 59,97 | 10,40 | 28,94 | 26,08 | 9,88 | 20,70 | 18,36 | 11,30 |
| 35 | 2,33 | 15,19 | 15,51 | 2,14 | 8,09 | 7,60 | 6,12 | 5,79 | 5,56 | 3,89 |
| | 5 | 19,68 | 21,82 | 10,91 | 10,48 | 10,32 | 1,59 | 7,50 | 7,66 | 2,23 |
| | 10 | 24,89 | 28,51 | 14,55 | 13,26 | 12,94 | 2,38 | 9,48 | 9,60 | 1,28 |
| | 25 | 33,95 | 39,56 | 16,51 | 18,09 | 17,27 | 4,52 | 12,94 | 12,62 | 2,44 |
| | 50 | 42,95 | 49,83 | 16,02 | 22,88 | 21,61 | 5,58 | 16,36 | 15,52 | 5,17 |
| | 100 | 54,32 | 61,82 | 13,80 | 28,94 | 27,22 | 5,95 | 20,70 | 19,23 | 7,10 |
| 40 | 2,33 | 15,19 | 16,89 | 11,18 | 8,09 | 8,19 | 1,22 | 5,79 | 5,97 | 3,21 |
| | 5 | 19,68 | 23,91 | 21,52 | 10,48 | 11,19 | 6,78 | 7,50 | 8,28 | 10,50 |
| | 10 | 24,89 | 31,39 | 26,13 | 13,26 | 14,11 | 6,44 | 9,48 | 10,43 | 9,97 |
| | 25 | 33,95 | 43,80 | 28,99 | 18,09 | 18,95 | 4,75 | 12,94 | 13,78 | 6,56 |
| | 50 | 42,95 | 55,36 | 28,90 | 22,88 | 23,82 | 4,08 | 16,36 | 17,02 | 4,05 |
| | 100 | 54,32 | 68,91 | 26,85 | 28,94 | 30,14 | 4,13 | 20,70 | 21,20 | 2,42 |

¹ Nota: Porcentaje de diferencia = [| Intensidad 3-58 - Intensidad 3-64 | / Intensidad 3-58]*100

$$Diferencia(\%) = \frac{|Intensidad (3 - 58) - Intensidad (3 - 64)|}{Intensidad (3 - 58)} \times 100 \quad (3-66)$$

3.15.1. Análisis de resultados de la aplicación de cópulas bivariadas en el análisis de series históricas

La tabla **3-6** muestra la diferencia porcentual encontrada entre las curvas IDF construidas mediante el enfoque empírico y el método de la cópula para duraciones de 60, 150 y 240 minutos. Las series históricas establecidas fueron [20, 30, 35, 40 y 57] años para la estación Santa Lucía. Para 60 minutos, se encontró una máxima diferencia de 10% entre los dos métodos para la serie histórica de 40 años con un periodo de retorno de 100 años. Para una duración de 150 minutos, la mayor diferencia de los métodos se presentó para la serie histórica de 20 años, con un valor aproximado de 32% para un periodo de retorno de 100 años. Para 240 minutos, se encontró una máxima diferencia de 34% para la serie histórica de 57 años con un periodo de retorno de 2,33 años. La tabla **3-5** muestra los diferentes valores del parámetro θ de la cópula para las series de tiempo establecidas. De esta tabla se puede observar que a medida que aumenta la serie histórica, el valor del coeficiente τ de kendall tiende a disminuir y por lo tanto el valor del parámetro de la cópula disminuye, así como su error en el cálculo de su obtención, esto quiere decir que las mayores diferencias de los dos métodos, se encuentran en las series de tiempo menores.

La tendencia muestra que mientras mayor sea el tamaño de la muestra, menor es la diferencia encontrada de los dos métodos. Esto no quiere decir que entre mayor sea el número de datos, la curva IDF mediante el método de la cópula se aproxima cada vez más a la curva IDF empírica, lo que muestra la tabla **3-6** a medida que aumenta el tamaño de la muestra, es que los valores encontrados por medio del enfoque de la cópula, oscilan alrededor de un valor desconocido al cual intentan converger para cada uno de los periodos de retorno establecidos.

Para tener mayor seguridad en estos comportamientos, se realizó el mismo procedimiento para la estación El Bosque. Los resultados se presentan en la tabla **3-8**. De esta tabla se puede observar que para 60 minutos la máxima diferencia entre los dos métodos es de 29% para un periodo de retorno de 25 años con la serie histórica de 40 años. Para una duración de 150 minutos, la mayor diferencia de los métodos se presentó para la serie histórica de 20 años, con un valor aproximado de 20% para un periodo de retorno de 100 años. Para 240 minutos, se encontró que la mayor diferencia es de 22% para la serie histórica de 20 años con un periodo de retorno de 100 años. Al igual que la estación Santa Lucía, la tabla **3-7** muestra que el τ de kendall tiende a disminuir a medida que aumenta la serie histórica para la estación El Bosque, de igual forma, el error en el cálculo del parámetro θ de la cópula disminuye.

La tendencia muestra que a medida que aumenta el periodo de retorno la diferencia de los dos métodos aumenta, este comportamiento es más marcado en la estación Santa Lucía. Una vez más se muestra que a medida que aumenta el tamaño de la muestra, menor es la diferencia entre los dos métodos.

3.16. Conclusiones del capítulo

Bajas dependencias del Tau de Kendall validan el supuesto de independencia y por lo tanto la distribución conjunta se reduce a un simple producto de las marginales. Como se vio en este capítulo, los taus de Kendall presentan una correlación negativa para cada una de las cuatro estaciones analizadas indicando que las variables aleatorias intensidad y duración son discordantes, es decir; a medida que una variable aumenta, la otra disminuye, lo que implica que las tormentas más cortas tienen intensidades más altas. Se comprueba entonces, que la construcción de distribuciones conjuntas que dependen de las características de las lluvias extremas es inevitable.

El proceso de construcción de la distribución conjunta bivariada a través de cópulas se puede descomponer en dos partes: las distribuciones marginales y la estructura de dependencia. La aplicación de la distribución condicional de la intensidad, dada una duración $D = d$ conocida, para la construcción de curvas IDF por medio de cópulas, produjo estimaciones cualitativamente similares a los obtenidos mediante el enfoque univariado convencional para duraciones de 360 minutos. Las diferencias encontradas por ambos métodos dependen de la duración de la tormenta y el periodo de retorno, sin embargo, no siguen un patrón sistemático particular.

Aunque las curvas IDF empíricas no tienen mucha base teórica, presentan resultados consistentes con respecto a las curvas IDF construidas por el método de la cópula, sin embargo, el método empírico requiere del análisis de varias ecuaciones de naturaleza empírica, tomando mayor tiempo en los cálculos y resultando menos eficiente. El método de la cópula es apoyado por una base teórica fuerte de probabilidad que proporciona una forma general para la distribución conjunta de intensidad de las tormentas con su duración siendo más significativo para la hidrología ya que las curvas IDF se pueden obtener directamente a partir de observaciones sin adoptar las relaciones empíricas.

De acuerdo con los resultados obtenidos del análisis de series históricas por medio de cópulas, si se cuenta con un registro de 20 años, el método de la cópula es más adecuado para representar el riesgo asociado al evento que el método univariado, por lo tanto se recomienda su aplicación para series históricas menores o iguales a 20 años. Por otra parte, las diferencias en los valores de intensidad por medio del método de la cópula para una serie histórica de 20 años con respecto a una serie de 40 años no son muy significativas para duraciones largas y periodos de retorno menores de 50 años. Para series históricas de 20 años con periodos de retorno mayores o iguales a 100 años y duraciones cortas, el método de la cópula subestima el riesgo asociado al evento.

4. Cópulas multivariadas

Como se definió en el capítulo anterior, una cópula es una función de distribución multivariada definida sobre el cubo de unidad $[0, 1]^n$ con marginales distribuidas uniformemente. Antes de introducir los conceptos de la cópula, es necesario mencionar algunas definiciones. (Ver Paloma (2006)).

4.1. Preliminares

Sean S_1, \dots, S_n subconjuntos no vacíos de $\overline{\mathbb{R}} = [-\infty, +\infty]$. Sea H una función real de n variables tales que $DomH = S_1 \times \dots \times S_n$ y para $a \leq b$ ($a_k \leq b_k \forall k$) sea $B = [a, b] = [a_1, b_1] \times \dots \times [a_n, b_n]$ cuyos vértices están en el $DomH$.

Definición 7. Llamamos H -volumen de B a

$$V_H(B) = \sum sign(c)H(c)$$

donde la suma se toma sobre todos los vértices c , B y $sign(c)$

Definición 8. Una función real H de n variables es n -decreciente si $V_H(B) \geq 0 \forall B$ cuyos vértices están en el $DomH$. Sea H la función de variable real n -dimensional con $DomH = S_1 \times \dots \times S_n$ donde S_k tiene un elemento más pequeño a_k . Supongamos que H verifica la siguiente propiedad:

Propiedad 1. $H(t) = 0 \forall t \in DomH$ tal que $t_k = a_k$ para al menos un k .

Definición 9. Una función de distribución n -dimensional es una función H con dominio $\overline{\mathbb{R}}^n$ tal que H es n -decreciente, verifica la propiedad 1. Las marginales de una función de distribución n -dimensional son funciones de distribución que se denotarán F_1, \dots, F_n .

4.2. Definición de cópula N-dimensional

Una cópula n -dimensional es una función C con dominio $[0, 1]^n$ tal que:

- I) C es creciente y verifica la propiedad 1.
- II) C tiene marginales C_k , $k = 1, 2, \dots, n$ que satisfacen $C_k(u) = u \forall u \in [0, 1]$

Para cualquier cópula n -dimensional C , $n \geq 3$, las marginales k -dimensionales de C son cópulas k -dimensionales. De igual forma que las cópulas bivariadas, una cópula n -dimensional C es una función con las siguientes propiedades (Zhang, 2005):

- Para cada $\mathbf{u} \in [0, 1]^n$, $C(\mathbf{u}) = 0$ si al menos una coordenada de \mathbf{u} es 0, y $C(\mathbf{u}) = u_k$ si todas las coordenadas de \mathbf{u} son iguales a 1 exepcto u_k .
- Para cada \mathbf{a} y \mathbf{b} en $[0, 1]^n$ tales que $a_i \leq b_i \forall i = 1 \dots n$ se verifica que $V_C[\mathbf{a}, \mathbf{b}] \geq 0$.

Puesto que las cópulas son funciones de distribución conjuntas (sobre $[0, 1]^n$), una cópula C induce una medida de probabilidad sobre $[0, 1]^n$ a través de:

$$V_C([0, u_1] \times \dots \times [0, u_n]) = C(u_1, \dots, u_n)$$

4.3. Teorema de Sklar N-dimensional

El teorema de Sklar bidimensional visto en la sección 3.3, se puede extender fácilmente al caso n -dimensional (Ver Zhang (2005)).

Teorema 4. Sea H una función de distribución n -dimensional con marginales F_1, \dots, F_n entonces existe una cópula n -dimensional C tal que para todo $F_1, \dots, F_n \in \overline{\mathbb{R}}^n$

$$H(x_1, \dots, x_n) = C(u_1, u_2, \dots, u_n) = C(F_1(x_1), \dots, F_n(x_n)) \quad (4-1)$$

Si F_1, \dots, F_n son todas continuas, entonces C es única, por tanto está unívocamente determinada sobre $\text{Ran}F_1 \times \dots \times \text{Ran}F_n$.

Inversamente, si C es una cópula n -dimensional y F_1, \dots, F_n son funciones de distribución, entonces la función H definida anteriormente es una función de distribución n -dimensional con marginales F_1, \dots, F_n (Paloma, 2006).

Considere el vector aleatorio X_1, \dots, X_n con funciones de distribución marginal $F_{X_i}(x) = P_{X_i}(X_i \leq x_i)$, donde n es el número de variables aleatorias y x_i es el valor de la variable

aleatoria $X_i(X_1, \dots, X_N)$. La distribución conjunta del vector aleatorio X_1, \dots, X_n es expresado como:

$$H_{X_1, \dots, X_N}(x_1, x_2, \dots, x_n) = P[X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n,] \text{ o simplemente } H.$$

Las cópulas son funciones que conectan las distribuciones de probabilidad multivariada a sus distribuciones de probabilidad marginales unidimensionales. Por lo tanto, la distribución de probabilidad multivariada, H , se expresa en términos de sus marginales y la función de dependencia asociada, C como : $C(F_{x_1}(x_1), F_{x_2}(x_2), \dots, F_{x_n}(x_n)) = H_{X_1, X_2, \dots, X_N}(x_1, x_2, \dots, x_n)$ donde C es la distribución conjunta del vector aleatorio X_1, \dots, X_n a través de sus marginales $F_{x_1}(x_1), F_{x_2}(x_2), \dots, F_{x_n}(x_n)$ (Zhang, 2005).

4.4. Cópulas arquimedianas N-dimensionales

De la misma forma que el teorema de Sklar, las cópulas arquimedianas se pueden extender al caso n-dimensional (ver Paloma (2006)).

Definición 9. Se dice que una cópula C es arquimediana si es de la forma:

$$C(u_1, \dots, u_n) = \begin{cases} \varphi^{-1}(\varphi(u_1) + \dots + \varphi(u_n)) & \text{si } \sum_{i=1}^n \varphi(u_i) \leq \varphi(0) \\ 0 & \text{En otro caso} \end{cases} \quad (4-2)$$

con $\varphi : [0, 1] \rightarrow [0, \infty]$ una función continua, estrictamente decreciente en $[0, \varphi(0)]$ con $\varphi(1) = 0$, $\varphi'(u) < 0$ y $\varphi''(u) > 0 \forall 0 \leq u \leq 1$. $\varphi(u)$ se denomina *generador de la cópula*.

Esta cópula se puede expresar como:

$$C_{\theta}^N(u) = \varphi^{[-1]}(\varphi(u_1) + \dots + \varphi(u_n)) \quad (4-3)$$

Y para $N \geq 3$ se tiene:

$$C_{\theta}^N(u_1, \dots, u_N) = C_{\theta}(C_{\theta}^{N-1}(u_1, \dots, u_{N-1}), u_N) \quad (4-4)$$

Las propiedades para las cópulas arquimedianas multivariadas son las mismas que para el caso bivariado.

4.5. Familias de cópulas arquimedianas N-dimensionales

Mediante el mismo procedimiento para la generación de las cópulas arquimedianas de dos dimensiones, las cópulas arquimedianas n-dimensionales se obtienen de forma similar y se expresan de la siguiente forma:

$$C_{\theta}^N(\mathbf{u}) = \varphi^{[-1]}(\varphi(u_1) + \dots + \varphi(u_n))$$

donde el superíndice N denota la dimensión y \mathbf{u} denota el vector variable. De la misma forma que para el caso bivariado, las cópulas arquimedianas n-dimensionales pueden ser representadas como sigue:

Cópula multivariada Gumbel-Hougaard

$$C_{\theta}^N(\mathbf{u}) = C_{\theta}^N(F_{x_1}(x_1), F_{x_2}(x_2), \dots, F_{x_n}(x_n)) = H(x_1, x_2, \dots, x_N)$$

$$C_{\theta}^N(\mathbf{u}) = \exp\left(-\left((-\ln u_1)^{\theta} + (-\ln u_2)^{\theta} + \dots + (-\ln u_N)^{\theta}\right)^{1/\theta}\right) \quad (4-5)$$

Con función de generación $\varphi(t) = (-\ln t)^{\theta}$

Cópula multivariada Frank

$$C_{\theta}^N(\mathbf{u}) = C_{\theta}^N(F_{x_1}(x_1), F_{x_2}(x_2), \dots, F_{x_n}(x_n)) = H(x_1, x_2, \dots, x_N)$$

$$C_{\theta}^N(\mathbf{u}) = -\frac{1}{\theta} \ln \left(1 + \frac{(\exp(-\theta u_1) - 1)(\exp(-\theta u_2) - 1) \dots (\exp(-\theta u_N) - 1)}{(\exp(-\theta) - 1)^{N-1}} \right), \quad \theta \neq 0 \quad (4-6)$$

con

$$\varphi(t) = -\ln \left(\frac{\exp(-\theta t) - 1}{\exp(-\theta) - 1} \right)$$

Cópula multivariada Cook-Johnson (Clayton)

$$C_{\theta}^N(\mathbf{u}) = C_{\theta}^N(F_{x_1}(x_1), F_{x_2}(x_2), \dots, F_{x_n}(x_n)) = H(x_1, x_2, \dots, x_N))$$

$$C_{\theta}^N(\mathbf{u}) = (u_1^{-\theta} + u_2^{-\theta} + \dots + u_N^{-\theta} - N + 1)^{-1/\theta} \quad \theta \geq 0 \quad (4-7)$$

con

$$\varphi(t) = \frac{1}{\theta}(t^{-\theta} - 1)$$

Cópula multivariada Ali-Mikhail-Haq

$$C_{\theta}^N(\mathbf{u}) = C_{\theta}^N(F_{x_1}(x_1), F_{x_2}(x_2), \dots, F_{x_n}(x_n)) = H(x_1, x_2, \dots, x_N))$$

$$C_{\theta}^N(\mathbf{u}) = \frac{u_1 u_2 \dots u_N}{1 - \theta(1 - u_1)(1 - u_2) \dots (1 - u_N)} \quad , \quad \theta \in [-1, 1] \quad (4-8)$$

con

$$\varphi(t) = \ln \frac{1 - \theta(1 - t)}{t}$$

4.6. Estimación del parámetro θ para cópulas N-dimensionales

Con el fin de estimar el parámetro θ de la cópula, se pueden considerar dos condiciones. En primer lugar, si las marginales apropiadas ya están disponibles, entonces se expresa simplemente la máxima verosimilitud para los datos. La estimación resultante de θ sería entonces la marginal dependiente; la misma metodología de máxima verosimilitud, que generalmente se aplica para la estimación de los parámetros de las distribuciones de probabilidad univariadas, se ve afectada indirectamente por el método de cópula. En segundo lugar, si se contemplan las estimaciones no paramétricas para las marginales, la estimación del parámetro θ de la cópula será la marginal libre. Por lo tanto, este parámetro puede ser encontrado utilizando un método de estimación semiparamétrica desarrollado por Genest et al. (1995). Esta estimación puede expresarse paso a paso como se muestra a continuación (Zhang, 2005):

1. Sea $\{(X_{1k}, \dots, X_{Nk}) : k = 1, \dots, N\}$ una muestra aleatoria, extraída de una distribución

multivariada como:

$$H(x_1, \dots, x_N) = C_\theta^N(F_1(x_1), \dots, F_N(x_N)) \quad (4-9)$$

2. La función log-likelihood de la cópula en la ecuación 4-9 puede ser expresada como:

$$L(\theta) = \sum_{k=1}^n \log [c_\theta^N \{F_{1n}(X_{1k}), \dots, F_{Nn}(X_{Nk})\}] \quad (4-10)$$

Donde c_θ es la función de densidad de la cópula que tiene el mismo significado que la función de densidad de probabilidad de variables aleatorias univariadas.

F_{in} es $n/(n+1)$ veces la función de distribución empírica marginal de la variable i -ésima con el fin de evitar la dificultad de la no acotación de $\log(C_\theta(u_1, \dots, u_N))$ ya que algunos de los u_i 's tienden a 1 (Zhang, 2005).

3. De acuerdo con la propiedad del estimador semiparamétrico θ es consistente y asintóticamente normal bajo las mismas condiciones de estimación de la máxima verosimilitud que es una propiedad asintótica. Con el fin de maximizar el log-likelihood es necesario derivar la ecuación 4-10 con respecto a θ e igualar a cero:

$$\frac{1}{n} \frac{\partial}{\partial \theta} = \frac{1}{n} \sum_{k=1}^n l_\theta \{\theta, F_{1n}(X_{1k}), F_{2n}(X_{2k}), \dots, F_{Nn}(X_{Nk})\} = 0 \quad (4-11)$$

donde L denota la función log-likelihood, y l_θ denota la derivada de L con respecto al parámetro θ .

4. Resolviendo la ecuación 4-11 se puede obtener el parámetro θ de la cópula.

La ventaja de la estimación semiparamétrica para obtener el parámetro de la cópula θ es que este puede ser aplicado directamente a cópulas n -dimensionales ($N \geq 3$) mientras que la estimación no paramétrica, a través del τ de Kendall sólo se puede aplicar para cópulas de dos dimensiones.

4.7. Periodo de retorno conjunto y posiciones de graficación para análisis de frecuencia multivariado

Periodo de retorno trivariado

Sean tres variables aleatorias denotadas como X , Y y Z . Estas variables puede ser dependientes. La distribución de probabilidad conjunta H puede ser expresada como:

$$H_{X,Y,Z}(x, y, z) = P(X \leq x, Y \leq y, Z \leq z) = \int_{-\infty}^z \int_{-\infty}^y \int_{-\infty}^x h_{X,Y,Z}(u, v, w) \, dudvdw \quad (4-12)$$

donde x, y, z son los valores de las variables aleatorias X, Y y Z respectivamente, y P es la probabilidad de no excedencia.

El periodo de retorno conjunto de un evento (x, y, z) , $T(x, y, z)$ puede ser definido como:

$$T(x, y, z) = \frac{1}{1 - H(x, y, z)} \quad (4-13)$$

Los periodos de retorno condicional y las distribuciones condicionales se presentan a continuación:

Sea $h_{X,Y,Z}(x, y, z)$ la función de densidad de probabilidad conjunta (PDF) de variables aleatorias X, Y y Z , y sean $f_X(x)$, $f_Y(y)$ y $f_Z(z)$ las PDF marginales de X, Y y Z respectivamente, entonces los periodos de retorno condicional se pueden expresar mediante los siguientes cuatro casos (Zhang, 2005):

Caso 1.

- La función de densidad de probabilidad conjunta condicional (CPDF) de X y Y dado $Z = z$ se puede escribir como:

$$h_{X,Y|Z=z}(x, y | Z = z) = \frac{f(x, y, z)}{f_Z(z)} \quad (4-14)$$

- Del mismo modo, el condicional CDF (CCDF) de X y Y dado $Z = z$ es:

$$H_{X,Y|Z=z}(x, y | Z = z) = \int_{-\infty}^x \int_{-\infty}^y h_{X,Y|Z=z}(u, v | z) \, dudv$$

$$H_{X,Y|Z=z}(x, y | Z = z) = \frac{\int_{-\infty}^x \int_{-\infty}^y h_{X,Y,Z}(u, v, z) dudv}{f_Z(z)} \quad (4-15)$$

- El periodo de retorno condicional correspondiente puede ser expresado como:

$$T_{X,Y|Z=z}(x, y | Z = z) = \frac{1}{1 - H_{X,Y|Z=z}(x, y | Z = z)} \quad (4-16)$$

Caso 2.

- El condicional PDF de X , dado $Y = y$ y $Z = z$ puede ser escrito como:

$$h_{X|Y=y,Z=z}(x | Y = y, Z = z) = \frac{h(x, y, z)}{h_{Y,Z}(y, z)} \quad (4-17)$$

- El condicional CDF (CCDF) de X , dado $Y = y$ y $Z = z$ es:

$$H_{X|Y=y,Z=z}(x | Y = y, Z = z) = \int_{-\infty}^x h_{X|Y=y,Z=z}(u | y, z) du$$

$$H_{X|Y=y,Z=z}(x | Y = y, Z = z) = \frac{\int_{-\infty}^x h_{X,Y,Z}(u, y, z) du}{h_{Y,Z}(y, z)} \quad (4-18)$$

- El periodo de retorno condicional correspondiente se puede expresar como:

$$T_{X|Y=y,Z=z}(x | Y = y, Z = z) = \frac{1}{1 - H(x | Y = y, Z = z)} \quad (4-19)$$

Caso 3.

- El CCDF de X y Y dado $Z \leq z$ puede ser expresado como:

$$H'_{X,Y|Z \leq z}(x, y | Z \leq z) = H(x, y | Z \leq z) = \frac{H(x, y, x)}{F_Z(z)} \quad (4-20)$$

- El periodo de retorno condicional bajo estas condiciones puede ser expresado como:

$$T'_{X,Y|Z \leq z}(x, y | Z \leq z) = \frac{1}{1 - H'(x, y | Z \leq z)} = \frac{F_Z(z)}{F_Z(z) - H(x, y, z)} \quad (4-21)$$

Caso 4.

- El CCDF de X dado $Y \leq y$ y $Z \leq z$ es:

$$H'_{X|Y \leq y, Z \leq z}(x | Y \leq y, Z \leq z) = H(x | Y \leq y, Z \leq z) = \frac{H_{X,Y,Z}(x, y, z)}{H_{Y,Z}(y, z)} \tag{4-22}$$

- El periodo de retorno condicional bajo estas condiciones es:

$$T'_{X|Y \leq y, Z \leq z}(x | Y \leq y, Z \leq z) = \frac{1}{1 - H'_{X|Y \leq y, Z \leq z}(x | Y \leq y, Z \leq z)}$$

$$T'(x | Y \leq y, Z \leq z) = \frac{H_{Y,Z}(y, z)}{H_{Y,Z}(y, z) - H_{X,Y,Z}(x, y, z)} \tag{4-23}$$

Posiciones de graficación trivariada

La formulación de las posiciones de graficación trivariada pueden obtenerse mediante la extensión del caso bivariado directamente como (Zhang, 2005):

$$F(x_i, y_i, z_i) = P(X \leq x_i, Y \leq y_i, Z \leq z_i) = \frac{\sum_{m=1}^i \sum_{l=1}^i \sum_{p=1}^i n_{mlp} - 0,44}{N + 0,12} \tag{4-24}$$

donde $F(x_i, y_i, z_i)$ se obtiene de las observaciones trivariadas (x_i, y_i, z_i) ya sea por x o y o z ;

n_{mlp} es el número de (x_i, y_i, z_i) observaciones contadas como: $x_j \leq x_i$, $y_j \leq y_i$, y $z_j \leq z_i$, $i = 1, \dots, n$, $1 \leq j \leq i$; y n es el tamaño de la muestra.

4.8. Distribuciones condicionales trivariadas a través de cópulas y correspondientes periodos de retorno

La distribución conjunta condicional utilizando el método de la cópula se presenta a continuación:

Sean X, Y, Z tres variables aleatorias con marginales $u = F_X(x)$, $v = F_Y(y)$ y $w = F_Z(z)$. Las funciones de distribución condicional bajo diferentes condiciones pueden ser expresadas utilizando la cópula arquimediana trivariada mediante los siguientes cuatro casos (Zhang, 2005):

Caso 1.

- La función de distribución condicional de X y Y dado $Z = z$ puede ser expresada como:

$$H(x, y | Z = z) = \frac{\partial H(x, y, z) / \partial z}{f_Z(z)}$$

$$H(x, y | Z = z) = C_\theta^3(U \leq u, V \leq v | W = w) = \frac{\partial C_\theta^3(u, v, w)}{\partial w} \quad (4-25)$$

- El correspondiente periodo de retorno es dado por:

$$T(x, y | Z = z) = \frac{1}{1 - C_\theta^3(U \leq u, V \leq v | W = w)} \quad (4-26)$$

Caso 2.

- La función de distribución condicional de X dado $Y = y$ y $Z = z$, puede ser expresada como:

$$H(x | Y = y, Z = z) = \frac{\partial^2 H(x, y, z) / (\partial y \partial z)}{h_{Y,Z}(y, z)} = C_\theta^3(u | V = v, W = w)$$

$$H(x | Y = y, Z = z) = \frac{\partial C_\theta^3(u, v, w)}{\partial v \partial w} \quad (4-27)$$

- El periodo de retorno es:

$$T(x | Y = y, Z = z) = \frac{1}{1 - C_\theta^3(u | V = v, W = w)} \quad (4-28)$$

Caso 3.

- La función de distribución condicional de X y Y dado $Z \leq z$, puede ser expresada como:

$$H(x, y | Z \leq z) = \frac{H(x, y, x)}{F_Z(z)} = C_{\theta}^3(u, v | W \leq w) = \frac{C_{\theta}^3(u, v, w)}{w} \quad (4-29)$$

- El periodo de retorno es:

$$T(x, y | Z \leq z) = \frac{1}{1 - C_{\theta}^3(u, v | W \leq w)} \quad (4-30)$$

Caso 4.

- La función de distribución condicional de X dado $Y \leq y$ y $Z \leq z$, puede ser expresada como:

$$H(x | Y \leq y, Z \leq z) = \frac{H(x, y, x)}{H'(y, z)} = C_{\theta}^3(u | V \leq v, W \leq w) = \frac{C_{\theta}^3(u, v, w)}{C_{\theta'}^2(v, w)} \quad (4-31)$$

- El periodo de retorno es:

$$T(x | Y \leq y, Z \leq z) = \frac{1}{1 - C_{\theta}^3(u | V \leq v, W \leq w)} \quad (4-32)$$

donde H' es la distribución conjunta de las variables aleatorias Y y Z . $C_{\theta'}^2$ es la función cópula bivariada correspondiente y θ' es el parámetro de la cópula bivariada.

4.9. Aplicación de las cópulas Multivariadas en el área de estudio

Como se ha subrayado con frecuencia en la literatura hidrológica, el análisis estadístico de los eventos extremos multivariados no es fácil de obtener debido a (1) la complejidad de los fenómenos, (2) el tamaño de la muestra reducida de los conjuntos actuales de datos multivariados, y (3) falta de disponibilidad de distribuciones de probabilidad multivariada adecuadas (Katz et al., 2002). Desafortunadamente, un buen número de distribuciones multivariadas presentes en la literatura, que son extensiones directas de las distribuciones univariadas, sufren de varias limitaciones y restricciones.

El periodo de diseño en el caso univariado suele ser identificado sin ambigüedad y es ampliamente utilizado en la práctica de la ingeniería, en el caso multivariado no es así. De hecho, el problema de identificación de eventos de diseño en un contexto multivariado es de importancia fundamental, pero de naturaleza problemática. Recientemente, varios esfuerzos se han dedicado a las cuestiones de diseño (véase, por ejemplo Serfling (2002); Belzunce et al. (2007); Chebana and Ouarda (2009); Chaouch and Goga (2010). La pregunta que surge es la siguiente: ¿Cómo es posible calcular el evento de diseño crítico en el caso multivariado?

En esta sección se intentará dar solución a este interrogante mediante dos aplicaciones; una de ellas será analizada para la zona de estudio del caso bivariado, es decir; en la cuenca baja del río Tunjuelo y la segunda aplicación se realizará para una subregión de la Costa Caribe Colombiana conocida como la Mojana.

4.9.1. Construcción de curvas IDF conjuntas multisitio mediante el método de la cópula

La construcción de Curvas IDF conjuntas se pueden aplicar para los 4 casos mencionados en la sección 4.8, sin embargo, para este trabajo se realizó el cálculo para los casos 1 y 4 ya que éstos son más utilizados en problemas de hidrología.

1. Caso 1

Las estaciones seleccionadas para aplicar este caso fueron la estación Juan Rey y la Estación Quiba, ya que estas tienen los mismos años de registro histórico. (ver tabla 2-3)

La determinación del parámetro de la cópula θ mediante la metodología mencionada en la sección 4.6 se aplicó para cada una de las familias de cópulas: Gumbel-Hougaard, Clayton,

Frank, Joe y Ali-Mikhail-Haq, resultando esta última la única cópula maximizable a los datos.

El parámetro θ de la cópula se determinó mediante la siguiente metodología:

1. Sea $\{(X_k, Y_k, Z_k) : k = 1, \dots, n\}$ una muestra aleatoria, extraída de una distribución multivariada como:

$$H(x, y, z) = C_{\theta}^3(F(x), F(y), F(z)) \quad (4-33)$$

2. La función log-likelihood de la cópula en la ecuación 4-33 puede ser expresada como:

$$L(\theta) = \sum_{k=1}^n \log [c_{\theta}^3 \{F(X_k), F(Y_k), F(Z_k)\}] \quad (4-34)$$

Donde c_{θ} es la función de densidad de la cópula que tiene el mismo significado que la función de densidad de probabilidad de variables aleatorias univariadas. (ver sección 3.3.1).

Si la cópula seleccionada Ali-Mikhail-Haq es:

$$C_{\theta}^3(u, v, w) = \frac{uvw}{1 - \theta(1-u)(1-v)(1-w)} \quad , \quad \theta \in [-1, 1)$$

entonces, su función de densidad es:

$$c_{\theta}^3 = \frac{\partial^3 C_{\theta}^3}{\partial u \partial v \partial w} \quad (4-35)$$

$u = F_X(x) = \text{Probabilidad acumulada de la Intensidad de la Estación Juan Rey (mm/h)}$
 $\rightarrow \text{Distribución ajustada: LogNormal}$

$v = F_Y(y) = \text{Probabilidad acumulada de la Intensidad de la Estación Quiba (mm/h)}$
 $\rightarrow \text{Distribución ajustada: LogNormal}$

$w = F_Z(z) = \text{Duración conjunta (min)}$
 $\rightarrow \text{Distribución ajustada: LogNormal}$

3. Con el fin de maximizar el log-likelihood es necesario derivar la ecuación 4-34 con respecto a θ e igualar a cero:

$$\frac{1}{n} \frac{\partial}{\partial \theta} = \frac{1}{n} \sum_{k=1}^n l_{\theta} \{\theta, F(X_k), F(Y_k), F(Z_k)\} = 0 \quad (4-36)$$

donde L denota la función log-likelihood, y l_θ denota la derivada de L con respecto al parámetro θ .

4. Resolviendo la ecuación 4-36 se puede obtener el parámetro θ de la cópula. (Ver anexo C).

De acuerdo con la sección 4.8, para el caso 1 se tiene:

• La función de distribución condicional de X y Y dado $Z = z$ puede ser expresada como:

$$H(x, y | Z = z) = \frac{\partial H(x, y, z) / \partial z}{f_Z(z)}$$

$$H(x, y | Z = z) = C_\theta^3(U \leq u, V \leq v | W = w) = \frac{\partial C_\theta^3(u, v, w)}{\partial w} \quad (4-37)$$

donde:

$X =$ intensidad de la estación Juan Rey (mm/h)

$Y =$ intensidad de la estación Quiba (mm/h)

$Z =$ duración (min)

Si la cópula Ali-Mikhail-Haq es:

$$C_\theta^3(u, v, w) = \frac{uvw}{1 - \theta(1-u)(1-v)(1-w)} \quad , \quad \theta \in [-1, 1)$$

$$\theta = 0,15$$

entonces, su función de distribución condicional es:

$$H(u, v | W = w) = \frac{\partial C_\theta^3(u, v, w)}{\partial w} = -\frac{\theta uvw(1-u)(1-v)}{A^2} + \frac{uv}{A}$$

$$A = -\theta(1-u)(1-v)(1-w) + 1 \quad (4-38)$$

$u = F_X(x) =$ Probabilidad acumulada de la Intensidad de la Estación Juan Rey (mm/h)

$v = F_Y(y) =$ Probabilidad acumulada de la Intensidad de la Estación Quiba (mm/h)

$w = F_Z(z) =$ Duración conjunta (min)

- El correspondiente periodo de retorno es dado por:

$$T(x, y | Z = z) = \frac{1}{1 - H(u, v | W = w)} \quad (4-39)$$

Así, para una duración de 15 minutos se cuentan con los registros históricos de las intensidades de las estaciones Juan Rey y Quiba y por medio de la ecuación 4-39 se puede obtener la figura 4-1, la cual representa el periodo de retorno conjunto de estas dos estaciones para una duración dada. (Ver anexo C).

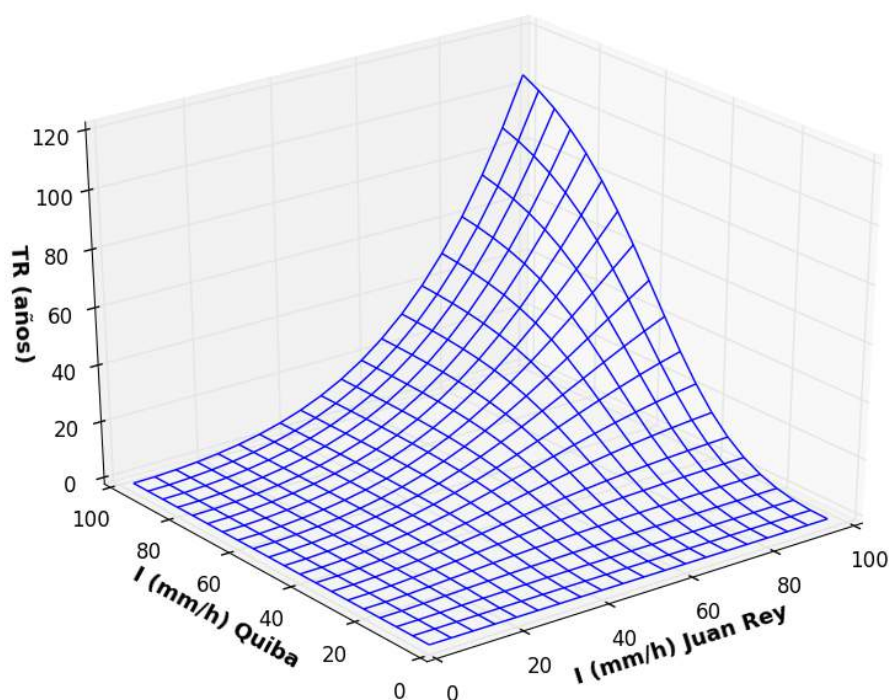
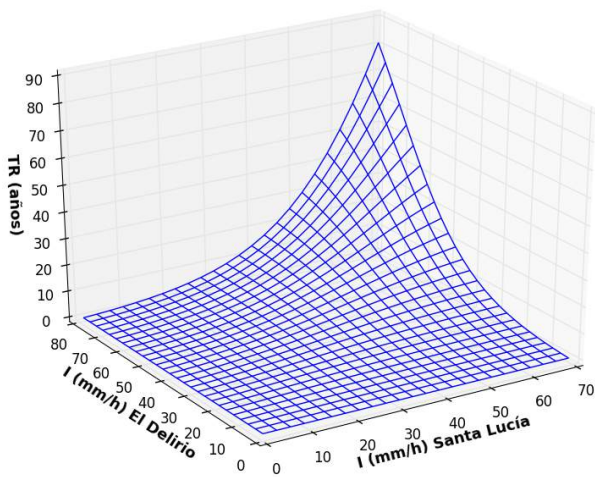
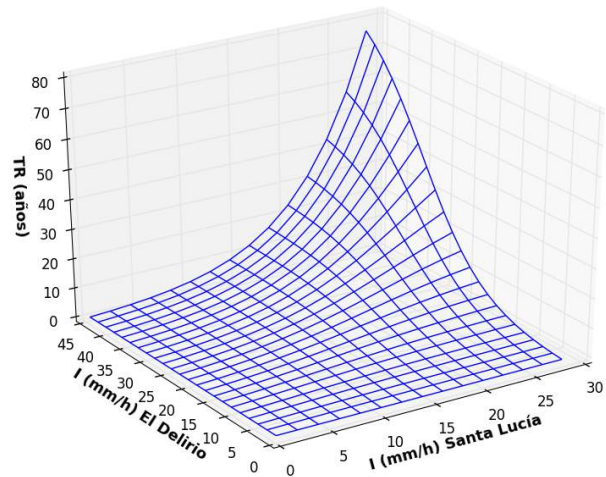


Figura 4-1.: Periodo de retorno conjunto para una duración de 15 minutos. Estaciones Juan Rey y Quiba.

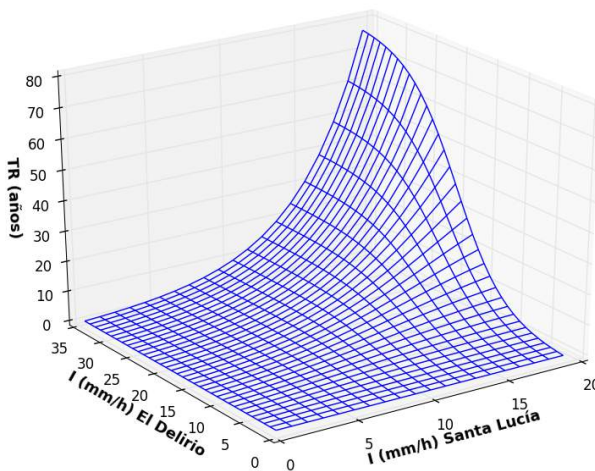
De igual forma, se obtuvieron los periodos de retorno conjunto para diferentes duraciones para las estaciones El delirio y Santa Lucía, las cuales cuentan con el mayor registro histórico de las estaciones seleccionadas para el estudio. (Ver figura 4-2).



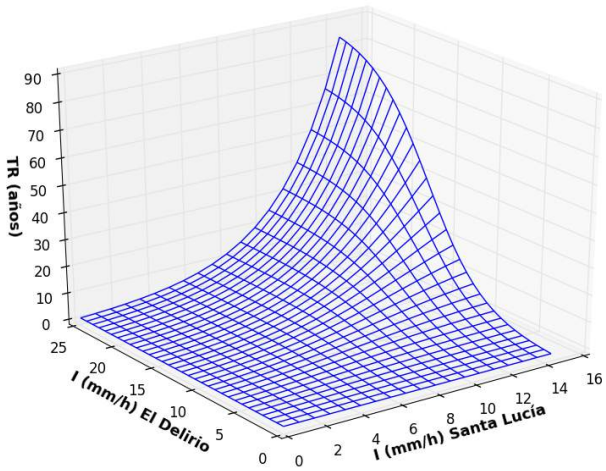
(a) Periodo de retorno conjunto para una duración de 30 minutos.



(b) Periodo de retorno conjunto para una duración de 90 minutos.



(c) Periodo de retorno conjunto para una duración de 180 minutos.



(d) Periodo de retorno conjunto para una duración de 300 minutos.

Figura 4-2.: Periodo de retorno conjunto para diferentes duraciones. Estaciones El Delirio y santa lucía.

2. Caso 4

El caso 4 se puede aplicar para diferentes escenarios, uno de ellos puede ser para la construcción de superficies IDFD (Intensidad - duración - frecuencia - precipitación) para una misma estación, otro escenario se puede aplicar, como en ese trabajo, para la construcción de curvas IDF conjuntas multisitio. Este caso en particular, se enfocó en métodos de interpolación a partir de puntos. Como se encuentra en la literatura, los métodos de interpolación pueden ser globales o locales. Los primeros utilizan toda la muestra para estimar el valor en cada punto nuevo. Los segundos utilizan sólo los puntos de muestreo más cercanos (Sarría, 2006).

Los métodos globales asumen la dependencia de la variable a interpolar de otras variables de apoyo, un ejemplo de éstos son los métodos de regresión que generalmente utilizan X e Y (longitud y latitud) como variables de apoyo y alguna variable cuantitativa espacialmente distribuida como por ejemplo la altitud. Los métodos locales se basan en la utilización de los puntos más cercanos al punto de interpolación para estimar la variable Z . Ejemplos de este método son el inverso de la distancia, kriging e interpolación por splines y mediante TIN (Redes irregulares de triángulos) (Sarría, 2006).

Usualmente, este análisis espacial se realiza con sistemas de información geográfica (SIG) que proporcionan herramientas de análisis para calcular actividades de geoprocesamiento como la interpolación de datos.

El modelo de la cópula para el caso 4, dependerá de la selección del método de interpolación. La aplicación para el caso 4 de la sección 4.8 se basó en el método local conocido como inverso de la distancia, el cual se aplicó al centroide de la cuenca baja del río Tunjuelo. (Ver figura 4-3).

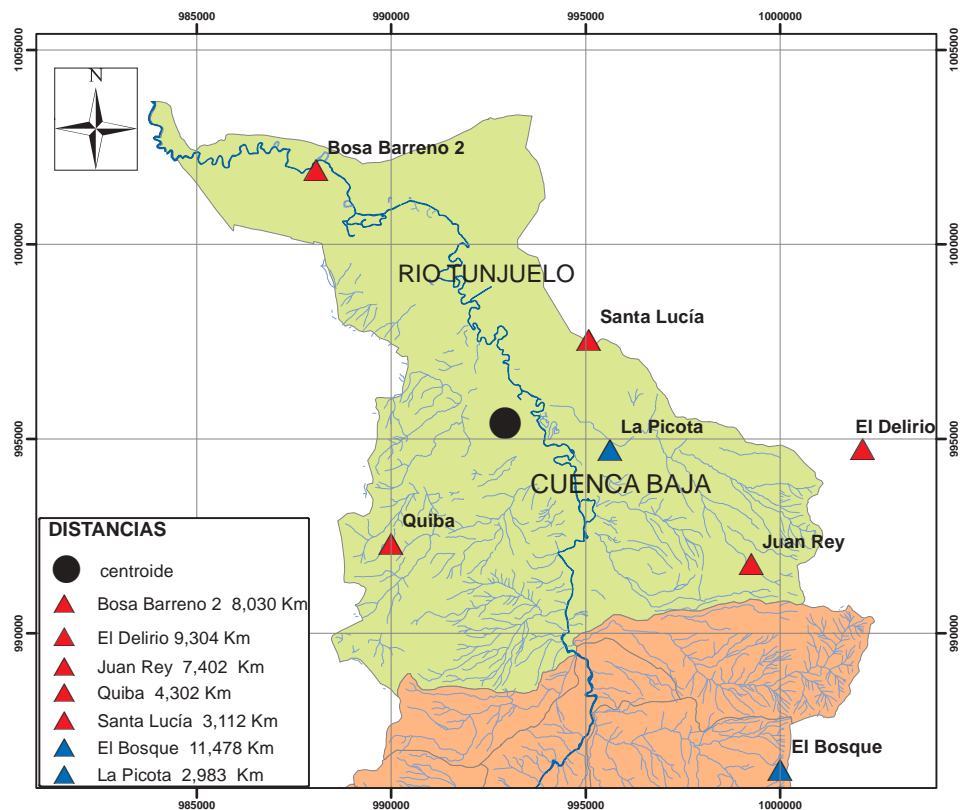


Figura 4-3.: Distancias al centroide de cada estación en la cuenca baja del río Tunjuelo

Antes de aplicar el método de la cópula, es necesario seleccionar una cópula adecuada para obtener el periodo de retorno condicional. De igual forma que en el caso bivariado, se realizó una prueba de aleatoriedad seguida de una prueba de independencia mutua y finalmente una prueba de bondad y ajuste por medio del paquete “copula” de R. Estas pruebas se realizaron para un conjunto de tres series de datos; el primero pertenece a los registros históricos de la estación El Delirio, el segundo son de la estación Santa Lucía y el tercero pertenece al método del inverso de la distancia aplicado a estas dos series.

Así, si se quiere construir una curva IDF conjunta por ejemplo en el centroide de la cuenca se tendría entonces:

$Int1 = \text{Intensidad para la estación El Delirio (mm/h)}$

$Int2 = \text{Intensidad para la estación Santa Lucía (mm/h)}$

$Int3 = \text{Intensidad para el centroide de la cuenca (mm/h)}$

donde

$$Int3 = \frac{\frac{1}{r_1^2}I_1 + \frac{1}{r_2^2}I_2}{\frac{1}{r_1^2} + \frac{1}{r_2^2}} \quad (4-40)$$

$$r_1 = 9,3km$$

$$r_2 = 3,11km$$

Selección de la cópula Trivariada mediante la aplicación del estadístico Cramer-von Mises usando el paquete copula en R

A continuación se presenta el modelo basado en el paquete “copula” de R mediante el estadístico Cramer-von-Mises para la selección de la cópula trivariada, mencionado en la sección 3.10.

La metodología propuesta por Kojadinovic et al. (2010) se aplicó para las estaciones El Delirio, Santa Lucía y el inverso de la distancia entre ellas ya que éstas cuentan con el mayor registro histórico.

```
setwd("~/Cpulas/RStudio/Conjunta/Trivariado")
library("copula")
Trivariado<-read.table("Datos_Delirio_Lucia_Trivariado.txt", header=T)
nrow(Trivariado)

## [1] 539

apply(Trivariado[, 2:4], 2, function(x) length(unique(x)))

## Int1 Int2 Int3
## 266 310 495

pseudoTrivariado <- apply(Trivariado [, 2:4], 2, rank)/(nrow(Trivariado) + 1)
```

Por cuestiones de la simplicidad, se pasan por alto los vínculos a medida que se sigue con el código, aunque fácilmente se puede proceder como en el caso bivariado, a expensas de más cálculos.

Como se están tratando con datos de series de tiempo hidrológicos, se podría esperar que el supuesto de independencia de la serie no se sostiene. La primera etapa es para realizar una prueba de aleatoriedad. Las herramientas estándar para esto son la prueba de Ljung and Box (1978) y su extensión multivariada estudiada por Hosking (1980) (véase también Johansen (1995)).

Como se describió en la sección 3.10, se aplica la prueba de aleatoriedad de los rendimientos al cuadrado, que se transforman en filas dentro de la función.

```
set.seed(123)
system.time(TrivariadoMultSerialIndepTest<-multSerialIndepTest(Trivariado[,2:4]^2,
  lag.max = 4, verbose = 0))

##      user  system elapsed
## 294.12    0.00 294.68

TrivariadoMultSerialIndepTest

##
## Global Cramer-von Mises statistic: 1.118285 with p-value 0.0004995005
## Combined p-values from the Mobius decomposition:
## 0.0004995005 from Fisher's rule,
## 0.0004995005 from Tippett's rule.

dependogram(TrivariadoMultSerialIndepTest)
dependogram(TrivariadoMultSerialIndepTest,print = TRUE)
```

Para visualizar los datos de la prueba de independencia multivariada vista en la sección 3.10, se puede usar el dependograma de la figura 4-4

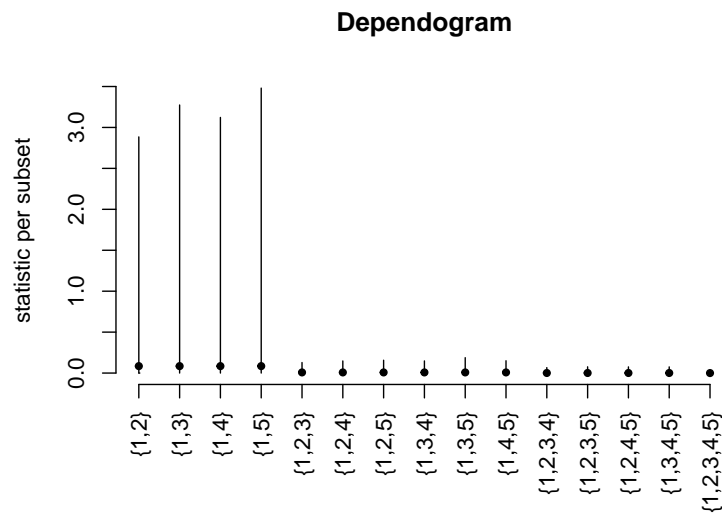


Figura 4-4.: Dependograma que resume los resultados de la prueba de independencia de la serie hidrológica para las estaciones El Delirio y Santa Lucía con $lag.max = 4$.

Los p-valores calculados dentro de la función (`TrivariadoMultSerialIndepTest`) indican una fuerte evidencia en contra de la hipótesis nula de independencia de la serie.

El siguiente paso es poner a prueba la independencia mutua de las tres series usando, por ejemplo, la prueba descrita en la sección 3.10.

```
system.time(empsamp <- indepTestSim(nrow(pseudoTrivariado), p = 3, N = 1000,
                                     verbose = 0))

##      user  system elapsed
##  84.83    0.00   85.13

TrivariadoMultIndepTest <- indepTest(pseudoTrivariado, empsamp)
TrivariadoMultIndepTest

##
## Global Cramer-von Mises statistic: 9.637795 with p-value 0.0004995005
## Combined p-values from the Mobius decomposition:
##   0.0004995005 from Fisher's rule,
##   0.0004995005 from Tippett's rule.

dependogram(TrivariadoMultIndepTest)
```

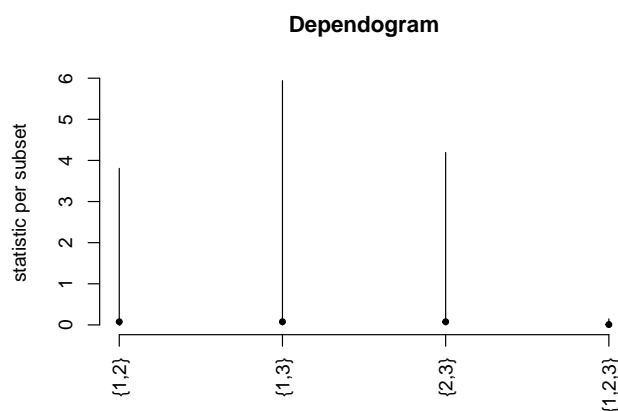


Figura 4-5.: Dependograma que resume los resultados de la prueba de independencia mutua de los datos trivariados para las estaciones El Delirio y Santa Lucía.

```
## Copula Gumbel

system.time(g.copula.mult <- gofCopula(gumbelCopula(dim = 3), pseudoTrivariado,
  estim.method = "mpl", simulation = "mult" ))

##      user  system elapsed
##    4.46    0.00    4.47

g.copula.mult

##
## Multiplier bootstrap goodness-of-fit test with 'method'="Sn",
## 'estim.method'="mpl"
##
## data:  x
## statistic = 0.9969, parameter = 2.7549, p-value = 0.0004995

fitCopula(gumbelCopula(dim = 3), pseudoTrivariado, method = "mpl")

## fitCopula() estimation based on 'maximum pseudo-likelihood'
## and a sample of size 539.
##      Estimate Std. Error
## param    2.755    0.126
## The maximized loglikelihood is 739.1
## Optimization converged
## Number of loglikelihood evaluations:
## function gradient
##      27      9
```

Existe una fuerte evidencia contra la hipótesis nula de independencia mutua, como era de esperar, el dependograma que se muestra en la figura 4-5, indica que la independencia es rechazada por todos los subconjuntos de las variables aleatorias.

El siguiente paso es llevar a cabo una prueba de bondad de ajuste como se mencionó en la sección 3.10 para las familias candidatas: Gumbel-Hougaard, Clayton, Frank y Joe.

```

## Copula Clayton

system.time(Clay.copula.mult <- gofCopula(claytonCopula(dim = 3),
      pseudoTrivariado, estim.method = "mpl", simulation = "mult" ))

##      user  system elapsed
##      2.14   0.00   2.14

Clay.copula.mult

##
## Multiplier bootstrap goodness-of-fit test with 'method'="Sn",
## 'estim.method'="mpl"
##
## data:  x
## statistic = 2.7779, parameter = 2.3092, p-value= 0.0004995

fitCopula(claytonCopula(dim = 3), pseudoTrivariado, method = "mpl")

## fitCopula() estimation based on 'maximum pseudo-likelihood'
## and a sample of size 539.
##      Estimate Std. Error
## param      2.309      0.137
## The maximized loglikelihood is 597.9
## Optimization converged
## Number of loglikelihood evaluations:
## function gradient
##      25      5

```

Como se puede observar, para las cópulas Gumbel y Clayton se tiene que:

Cópula Gumbel:

$$\theta = 2,75 \quad , \quad p\text{-valor} = 0,0004995$$

Cópula Clayton:

$$\theta = 2,30 \quad , \quad p\text{-valor} = 0,0004995$$

```

##Cópula Frank

system.time(Frank.copula.mult <- gofCopula(francCopula(dim = 3),
      pseudoTrivariado, estim.method = "mpl", simulation = "mult" ))

##      user  system elapsed
##      2.22   0.00   2.21

Frank.copula.mult

##
## Multiplier bootstrap goodness-of-fit test with 'method'="Sn",
## 'estim.method'="mpl"
##
## data:  x
## statistic = 0.35519, parameter = 11.18, p-value = 0.0004995

fitCopula(francCopula(dim = 3), pseudoTrivariado, method = "mpl")

## fitCopula() estimation based on 'maximum pseudo-likelihood'
## and a sample of size 539.
##      Estimate Std. Error
## param    11.18     0.359
## The maximized loglikelihood is 840.8
## Optimization converged
## Number of loglikelihood evaluations:
## function gradient
##      30      9

```

Para la cópula Frank y la cópula Joe se tiene:

Cópula Frank:

$$\theta = 11,18$$

$$p\text{-valor} = 0,0004995$$

Cópula Joe:

$$\theta = 3,25$$

$$p\text{-valor} = 0,0004995$$

```

##Cópula JOE

system.time(Joe.copula.mult <- gofCopula(joeCopula(dim = 3), pseudoTrivariado,
  estim.method = "mpl", simulation = "mult" ))

##    user  system elapsed
##    6.36    0.00    6.40

Joe.copula.mult

##
## Multiplier bootstrap goodness-of-fit test with 'method'="Sn",
## 'estim.method'="mpl"
##
## data: x
## statistic = 2.4237, parameter = 3.2533, p-value = 0.0004995

fitCopula(joeCopula(dim = 3), pseudoTrivariado, method = "mpl")

## fitCopula() estimation based on 'maximum pseudo-likelihood'
## and a sample of size 539.
##      Estimate Std. Error
## param   3.253    0.198
## The maximized loglikelihood is 594.1
## Optimization converged
## Number of loglikelihood evaluations:
## function gradient
##      21      6

```

Tabla 4-1.: Resumen de la prueba de bondad de ajuste para las familias candidatas. Parámetro θ trivariado para las estaciones El delirio, Santa Lucía y el inverso de la distancia

| Familias de Cópulas | θ | Estadístico Sn | p - valor | Error |
|---------------------|----------|----------------|-----------|-------|
| Gumbel | 2,7549 | 0,9969 | 0,0004995 | 0,126 |
| Clayton | 2,3092 | 2,7779 | 0,0004995 | 0,137 |
| Frank | 11,18 | 0,35519 | 0,0004995 | 0,359 |
| Joe | 3,2533 | 2,4237 | 0,0004995 | 0,198 |

La tabla 4-1 presenta un resumen de los resultados obtenidos aplicando el modelo mencionado en la sección 3.10 para la prueba de bondad de ajuste mediante el estadístico S_n . La

cópula de mejor ajuste es aquella que presenta el p-valor más alto. Como el p-valor es el mismo para todas las cópulas se escogió la cópula que presentó el menor error. En este caso la cópula de mejor ajuste es la cópula Gumbel-Hougaard.

De acuerdo con la sección 4.8, para el caso 4 se tiene:

- La función de distribución condicional de X dado $Y \leq y$ y $Z \leq z$, puede ser expresada como:

$$H(x | Y \leq y, Z \leq z) = \frac{H(x, y, z)}{H'(y, z)} = C_{\theta}^3(u | V \leq v, W \leq w) = \frac{C_{\theta}^3(u, v, w)}{C_{\theta'}^2(v, w)} \quad (4-41)$$

donde:

$X =$ Intensidad del inverso de la distancia para Y y Z (mm/h)

$Y =$ Intensidad de la estación Santa Lucía (mm/h)

$Z =$ Intensidad de la estación El delirio (mm/h)

Como la cópula de mejor ajuste es la Gumbel-Hougaard entonces se tiene:

$$C_{\theta}^3(u, v, w) = \exp \left(- \left((-\ln u)^{\theta} + (-\ln v)^{\theta} + (-\ln w)^{\theta} \right)^{1/\theta} \right) \quad (4-42)$$

$\theta = 2,75 \rightarrow$ tabla **4-1**

$$C_{\theta}^2(v, w) = \exp \left(- \left((-\ln v)^{\theta} + (-\ln w)^{\theta} \right)^{1/\theta} \right) \quad (4-43)$$

$\theta = 2,56 \rightarrow$ tabla **3-4**

La función de distribución condicional es:

$$H(u | V \leq v, W \leq w) = \frac{C_{\theta}^3(u, v, w)}{C_{\theta'}^2(v, w)} = \frac{\exp \left(- \left((-\ln u)^{\theta} + (-\ln v)^{\theta} + (-\ln w)^{\theta} \right)^{1/\theta} \right)}{\exp \left(- \left((-\ln v)^{\theta} + (-\ln w)^{\theta} \right)^{1/\theta} \right)} \quad (4-44)$$

$u = F_X(x) =$ Probabilidad acum. Intensidad del inverso de la distancia para v y w (mm/h)

$v = F_Y(y) =$ Probabilidad acum. Intensidad de la Estación Santa Lucía (mm/h)

$w = F_Z(z) = \text{Probabilidad acum. Intensidad de la Estación El Delirio (mm/h)}$

- El periodo de retorno es:

$$T(x | Y \leq y, Z \leq z) = \frac{1}{1 - H(u | V \leq v, W \leq w)} \quad (4-45)$$

Teniendo las intensidades de las estaciones Santa Lucía y El Delirio para cada una de las duraciones establecidas [15, 30, 45, 60, 90, 120, 150, 180, 240, 300, 360] minutos, y aplicando la ecuación 4-45 se puede construir una curva IDF conjunta para el centroide de la cuenca baja del río Tunjuelo. La ecuación 4-45 arroja resultados del periodo de retorno condicional de la distribución X dados Y y Z . Este caso no proporciona una curva IDF directa como en el caso bivariado, por esto es necesario realizar una regresión no lineal a los datos de la intensidad en el centroide de la cuenca, las duraciones establecidas y el periodo de retorno de la ecuación 4-45.

Al realizar la regresión no lineal a estos datos, se encontró que la ecuación de mejor ajuste y que presenta menor error es la que corresponde al modelo propuesto por Mijares (1992). (ver ecuación 2-13).

$$i = \frac{kT^m}{(d+c)^n} = \frac{6314,7 T^{0,26}}{(d+54,1)^{1,3}} \rightarrow \text{curva IDF conjunta. Método de la cópula} \quad (4-46)$$

De igual forma, se obtuvo la curva IDF empírica que resulta de aplicar directamente el método del inverso de la distancia al centroide de la cuenca de los datos nativos de las estaciones Santa Lucía y el Delirio. La curva IDF empírica se obtiene como se ilustró en la sección 2.3.2. La ecuación de mejor ajuste y que presenta menor error es la correspondiente al modelo propuesto por Mijares (1992).

$$i = \frac{kT^m}{(d+c)^n} = \frac{1110,3 T^{0,24}}{(d+20,67)^{0,95}} \rightarrow \text{curva IDF conjunta. Método empírico} \quad (4-47)$$

La curva IDF conjunta obtenida utilizando la ecuación 4-46 se presenta en la figura 4-6.

Una vez obtenidas las curvas IDF conjuntas por medio del método de la cópula (ec. 4-46), se realiza la comparación con el método empírico (ec. 4-47) como se muestra en la figura 4-7. El porcentaje de diferencia se encuentra en la tabla 4-2.

De igual forma, se obtuvieron las curvas IDF conjuntas para el caso 4 mediante el método de la cópula y el método empírico para las estaciones Juan rey y Quiba. (ver figuras 4-8 y 4-9). Los resultados obtenidos se presentan en la tabla 4-2

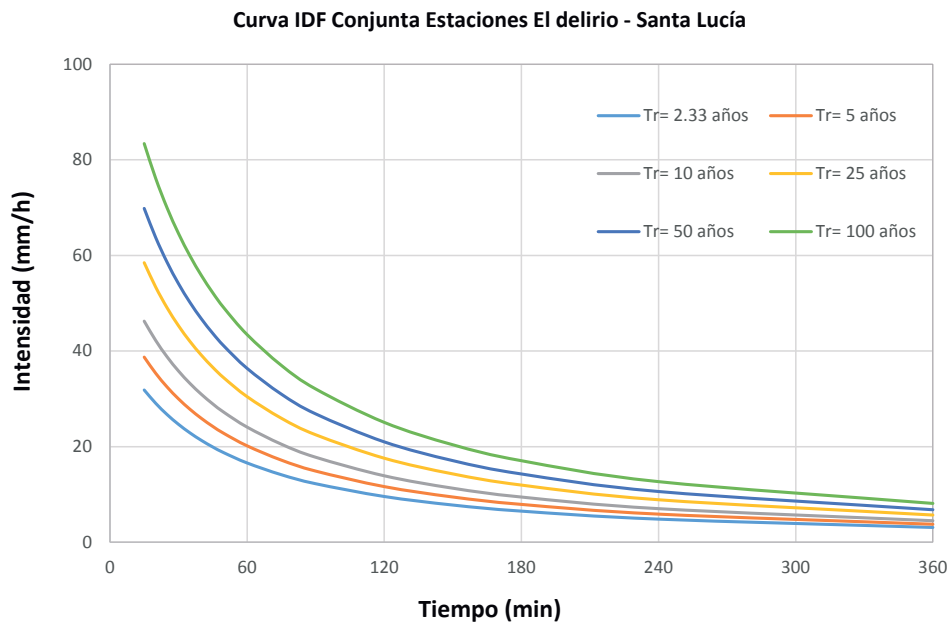


Figura 4-6.: Curvas IDF conjuntas por medio de la aplicación del método de la cópula para las estaciones El delirio y Santa Lucía.

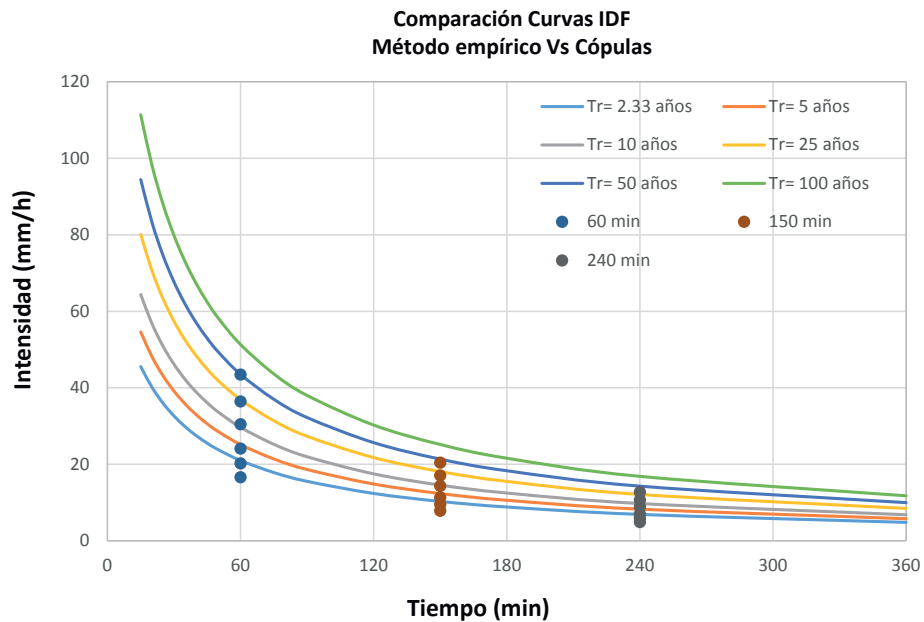


Figura 4-7.: Comparación entre las curvas IDF conjuntas aplicando el método de la cópula y el método empírico para las estaciones El delirio y Santa Lucía.

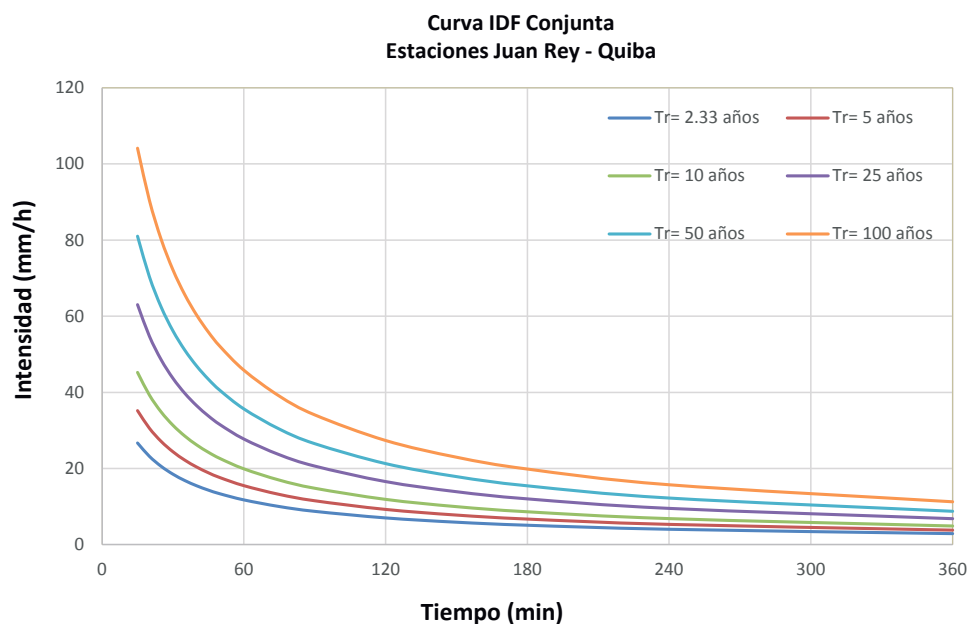


Figura 4-8.: Curvas IDF conjuntas por medio de la aplicación del método de la cópula para las estaciones Juan Rey y Quiba.

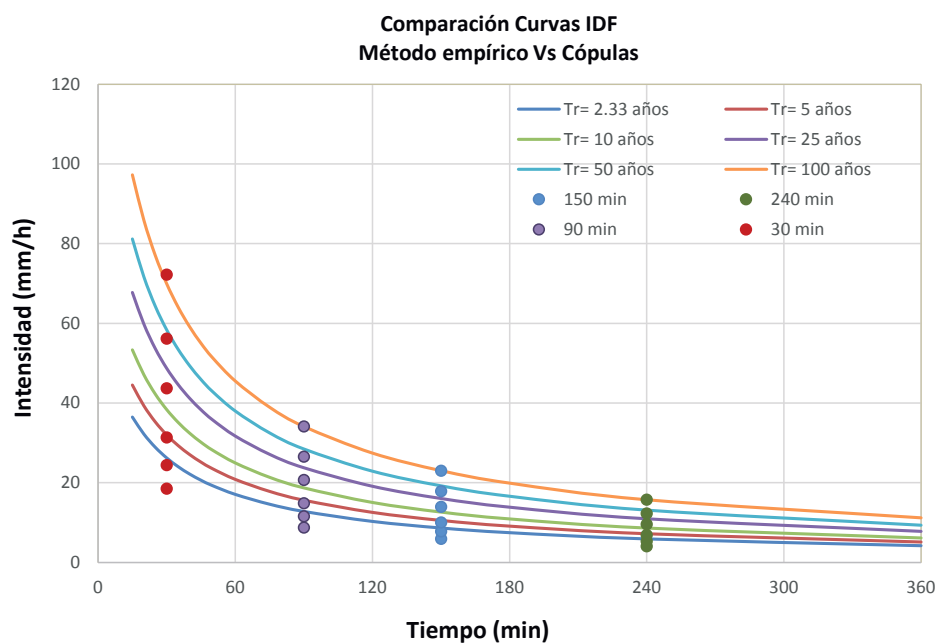


Figura 4-9.: Comparación entre las curvas IDF conjuntas aplicando el método de la cópula y el método empírico para las estaciones Juan Rey y Quiba.

La tabla 4-2 presenta el porcentaje de diferencia al aplicar las ecuaciones 4-46 y 4-47 para tres diferentes duraciones.

Tabla 4-2.: Comparación entre curvas IDF conjuntas multisitio por medio de cópulas y aplicando el método empírico

| Estaciones | Periodo de retorno (años) | 60 min | | | 150 min | | | 240 min | | |
|--------------------------------|---------------------------------|-----------------------------|---------------------------|--------------------------|-----------------------------|---------------------------|-------------|-----------------------------|---------------------------|-------------|
| | | IDF Empírica Ec. 2-13 | IDF Cópula Ec. 4-45 | Dif. ¹ (%) | IDF Empírica Ec. 2-13 | IDF Cópula Ec. 4-45 | Dif. (%) | IDF Empírica Ec. 2-13 | IDF Cópula Ec. 4-45 | Dif. (%) |
| Santa Lucía y El Delirio | 2,33 | 20,96 | 16,60 | 20,84 | 10,29 | 7,79 | 24,26 | 6,88 | 4,85 | 29,56 |
| | 5 | 25,14 | 20,18 | 19,75 | 12,34 | 9,47 | 23,21 | 8,25 | 5,89 | 28,59 |
| | 10 | 29,65 | 24,10 | 18,74 | 14,55 | 11,31 | 22,25 | 9,73 | 7,04 | 27,69 |
| | 25 | 36,88 | 30,47 | 17,39 | 18,10 | 14,30 | 20,96 | 12,10 | 8,90 | 26,49 |
| | 50 | 43,49 | 36,38 | 16,35 | 21,34 | 17,08 | 19,97 | 14,27 | 10,62 | 25,57 |
| | 100 | 51,29 | 43,44 | 15,30 | 25,17 | 20,40 | 18,96 | 16,83 | 12,69 | 24,63 |
| Juan Rey y Quiba | 2,33 | 17,07 | 11,75 | 31,19 | 8,64 | 5,89 | 31,85 | 5,90 | 4,03 | 31,63 |
| | 5 | 20,83 | 15,49 | 25,66 | 10,54 | 7,76 | 26,38 | 7,19 | 5,31 | 26,14 |
| | 10 | 24,96 | 19,91 | 20,26 | 12,63 | 9,97 | 21,03 | 8,62 | 6,83 | 20,77 |
| | 25 | 31,71 | 27,74 | 12,52 | 16,04 | 13,90 | 13,35 | 10,95 | 9,52 | 13,08 |
| | 50 | 37,99 | 35,65 | 6,16 | 19,22 | 17,86 | 7,06 | 13,12 | 12,23 | 6,76 |
| | 100 | 45,52 | 45,82 | 0,66 | 23,03 | 22,96 | 0,30 | 15,72 | 15,72 | 0,01 |

¹ Nota: Porcentaje de diferencia = $\frac{|Intensidad\ 2-13 - Intensidad\ 4-45|}{Intensidad\ 2-13} * 100$

El porcentaje de diferencia fue calculado como:

$$Diferencia(\%) = \frac{|Intensidad\ (2 - 13) - Intensidad\ (4 - 45)|}{Intensidad\ (2 - 13)} \times 100 \quad (4-48)$$

4.9.2. Análisis de resultados de la construcción de curvas IDF conjuntas multisitio

El caso 1 estudiado en la sección 4.8, representa una metodología para encontrar curvas IDF conjuntas multisitio a través del método de la cópula. La figura 4-1 presenta el periodo de retorno condicional conjunto para las estaciones Juan rey y Quiba para una duración de 15 minutos. La figura 4-2 presenta el periodo de retorno conjunto para diferentes duraciones de las estaciones El Delirio y Santa Lucía. Se puede observar que a medida que aumenta la duración, disminuye la intensidad y el periodo de retorno se mantiene constante. Esto es consistente con las curvas IDF empíricas univariadas. Una vez más se demuestra que usando las funciones cópula arquimedianas, el análisis de las distribuciones conjuntas se puede realizar por medio de las distribuciones marginales por separado, además se pueden construir nuevos modelos presentando nuevas formulaciones, para este caso curvas IDF conjuntas multisitio.

Para el caso 4, se presenta el porcentaje de diferencia en la tabla 4-2 de las curvas IDF conjuntas multisitio, obtenidas a partir del enfoque de la cópula y el método empírico para las duraciones de 60, 150 y 240 minutos. Para 60 minutos se encuentra que la máxima diferencia entre los dos métodos es de 32 % para un periodo de retorno de 2,33 años correspondiente a las estaciones de Juan Rey y Quiba, las cuales cuentan con un registro histórico de 22 años. Para una duración de 150 minutos, la mayor diferencia se presenta en las estaciones Juan Rey y Quiba para un periodo de retorno de 2,33 años con una diferencia de 32 %. Para una duración de 240 minutos, se presenta una diferencia máxima entre los dos métodos de 32 % para un periodo de retorno de 2,33 años correspondiente a las estaciones Juan Rey y Quiba. Para las tres duraciones evaluadas las diferencias tienden a disminuir a medida que aumenta el periodo de retorno. Cuando las duraciones aumentan las diferencias van hacia arriba y hacia abajo.

De la tabla 4-2 se puede ver que las diferencias entre las dos metodologías para la construcción de curvas IDF conjuntas tienden a disminuir cuando el periodo de retorno aumenta, esto es debido a que en los dos enfoques se realizó un método de regresión no lineal de naturaleza empírico. El caso 4, representa el periodo de retorno condicional de un lugar o zona (en este caso el centroide de la cuenca) dadas dos variables aleatorias, que corresponden a las intensidades de dos estaciones. La regresión no lineal realizada es simplemente para comparar las dos metodologías. La curva IDF conjunta multisitio dependerá del método de interpolación usado. Se recomienda realizar un estudio más profundo con más variables aleatorias para métodos de interpolación usados en hidrología.

4.9.3. Análisis de frecuencia multivariado mediante el método de la cópula

La severidad de un evento hidrológico, se expresa mediante una probabilidad de no exceedencia, o lo que es equivalente, en términos del periodo de retorno. El uso de cópulas en hidrología, así como en otras ciencias geofísicas y ambientales, es reciente y de rápido crecimiento. En esta sección se muestra cómo, aprovechando los últimos desarrollos teóricos, se pueden construir modelos multivariados fácilmente a través de cópulas, dando lugar a nuevas formulaciones. Además, se presenta la noción de periodo de retorno multivariado en términos de cópulas. Este enfoque es bastante general y de base física.

Para este análisis se seleccionó la subregión de la Mojana ya que se caracteriza por ser un complejo de humedales que son fundamentales en la amortiguación de inundaciones.

Localización de la Mojana

La Mojana es una subregión de la Costa Caribe colombiana, caracterizada por ser una zona de humedales productivos que pertenecen a la Depresión Momposina. Actúa como regulador de tres grandes ríos: el Magdalena, el Cauca y el San Jorge. (Aguilera, 2004).

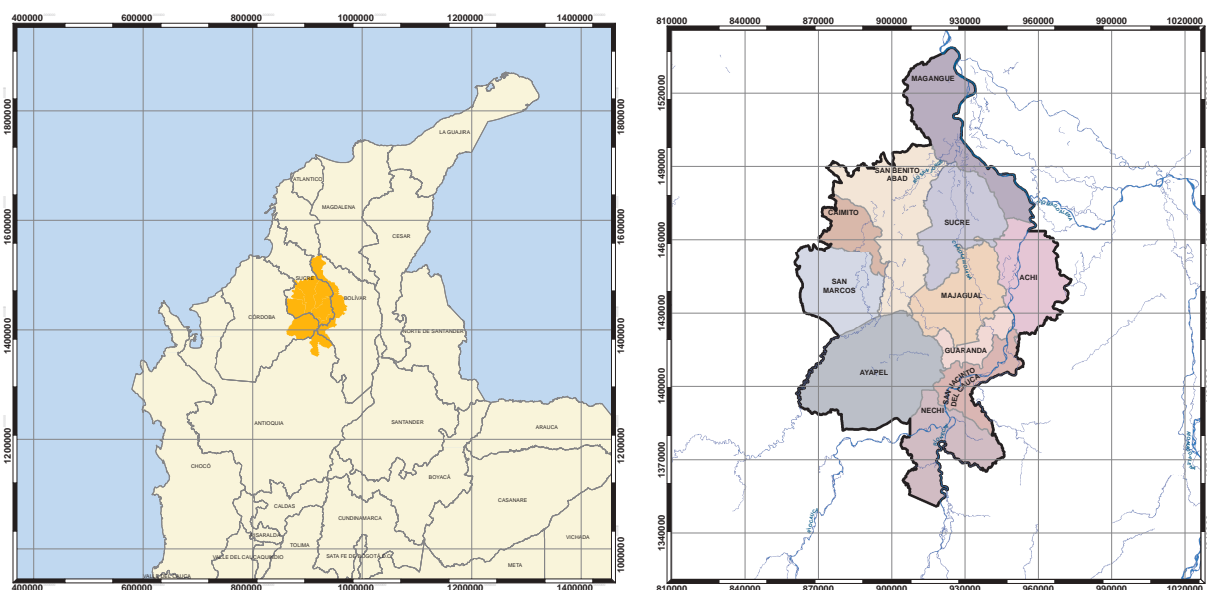


Figura 4-10.: Localización general de la Mojana (Adaptación, 2016).

La subregión de La Mojana Está delimitada geográficamente al oriente con el río Cauca, al occidente con el río San Jorge y la ciénaga de Ayapel, al nororiente con el brazo de Loba del

río Magdalena, y al sur con las tierras altas de Cauca y la serranía de Ayapel. (Aguilera, 2004).

Recopilación de información

La base de datos de información hidrológica fue suministrada por el Fondo Adaptación (Adaptación, 2016). Las estaciones seleccionadas se encuentran ubicadas en la región de la Mojana cuyas características se encuentran en la tabla 4-3. La localización de las estaciones se encuentra en la figura 4-11.

Tabla 4-3.: Estaciones hidrológicas seleccionadas pertenecientes a la región de la Mojana

| Nombre | Código | Corriente | Tipo | Entidad | Periodo |
|--------------|----------|---------------|------|---------|-------------|
| La Coquera | 26247020 | Cauca | LG | IDEAM | 1966 - 2015 |
| La Esperanza | 27037010 | Nechi | LG | IDEAM | 1966 - 2015 |
| Montelíbano | 25017010 | San Jorge | LG | IDEAM | 1973 - 2014 |
| Armenia | 25027360 | Brazo de Loba | LM | IDEAM | 1973 - 2014 |
| La Raya | 25027910 | Caño Caribona | LM | IDEAM | 1975 - 2015 |

Un evento de inundación puede ser estudiado en diferentes maneras. El rasgo más comúnmente evaluado es el flujo máximo del evento. Sin embargo, la duración o el volumen pueden ser de gran interés.

Un evento de inundación en una confluencia de ríos, puede ser el resultado de altas descargas en sólo uno o varios de los flujos aguas arriba. Si las descargas de estas corrientes son independientes, el comportamiento probabilístico del flujo aguas abajo, es igual a la suma de las corrientes aguas arriba que se puede obtener por convolución. Sin embargo, ambos afluentes pueden llegar a tener flujos simultáneos. En este caso, el cálculo anterior puede conducir a una fuerte subestimación del riesgo. (Renard and Lang, 2007). Es por esto que la región de la Mojana requiere de distribuciones multivariadas conjuntas para el análisis de frecuencia.

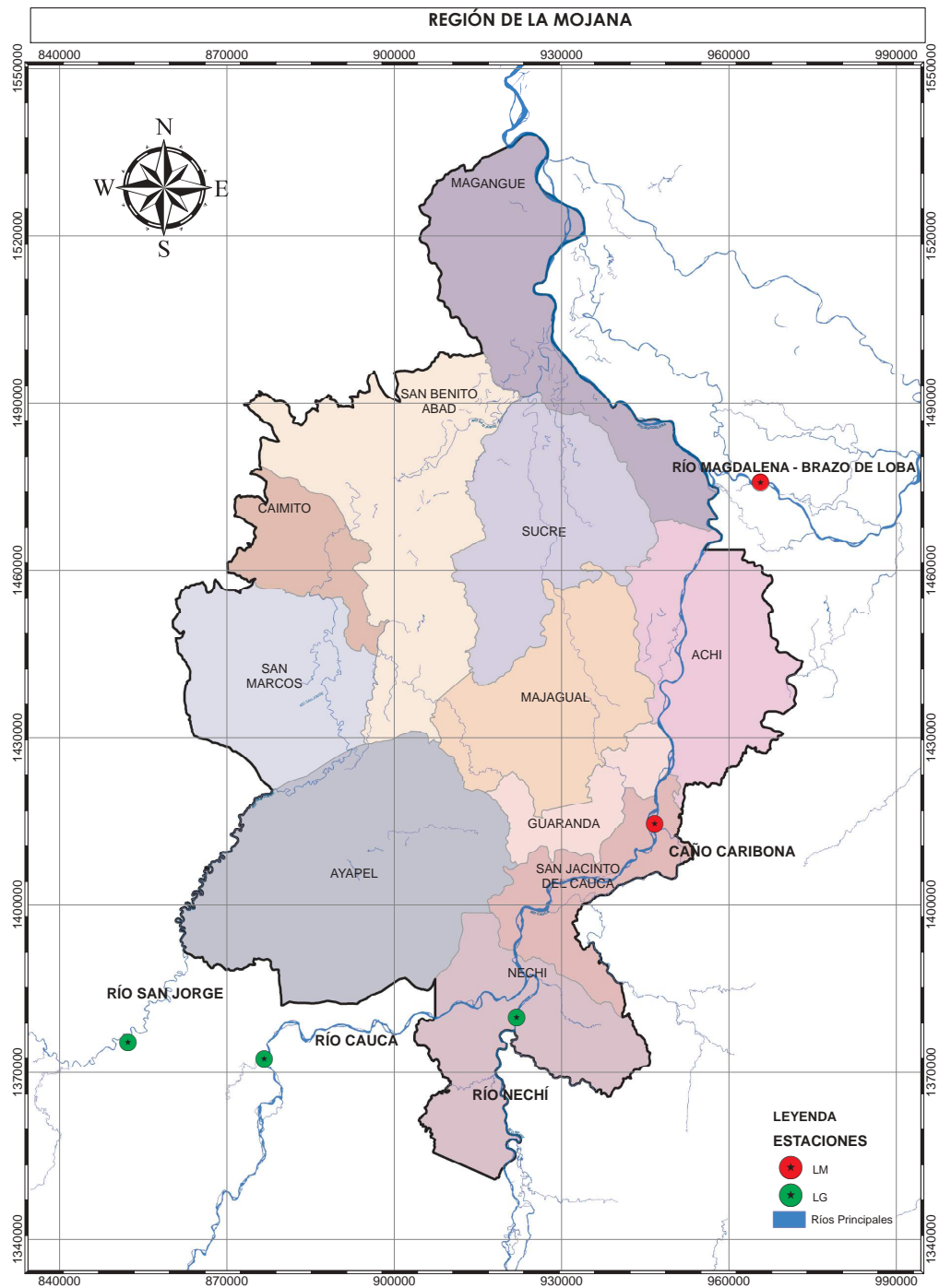


Figura 4-11.: Ubicación de las estaciones pertenecientes a la región de la Mojana. (Adaptación, 2016).

4.9.4. Periodo de retorno conjunto multivariado multisitio

Como se vio en la sección 4.7, el periodo de retorno conjunto trivariado está dado por:

$$T(x, y, z) = \frac{1}{1 - H(x, y, z)} \quad (4-49)$$

donde

$$H(x, y, z) = \exp \left(- \left((-\ln F_X(x))^\theta + (-\ln F_Y(y))^\theta + (-\ln F_Z(z))^\theta \right)^{1/\theta} \right) \quad (4-50)$$

$F_X(x)$ = Probabilidad acumulada de una estación LM o LG

$F_Y(y)$ = Probabilidad acumulada de una estación LM o LG

$F_Z(z)$ = Probabilidad acumulada de una estación LM o LG

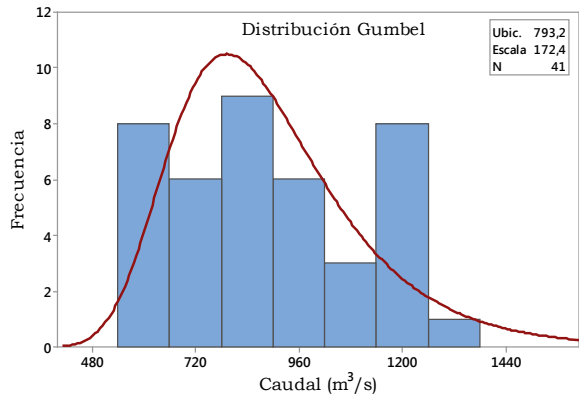
Para realizar este estudio, es necesario obtener las distribuciones de probabilidad marginales para cada una de las estaciones. La tabla 4-4 muestra las distribuciones de probabilidad de mejor ajuste a los registros obtenidas mediante la prueba Kolmogorov-Smirnov.

Tabla 4-4.: Distribuciones de probabilidad marginales asociadas a cada una de las estaciones pertenecientes a la región de la Mojana

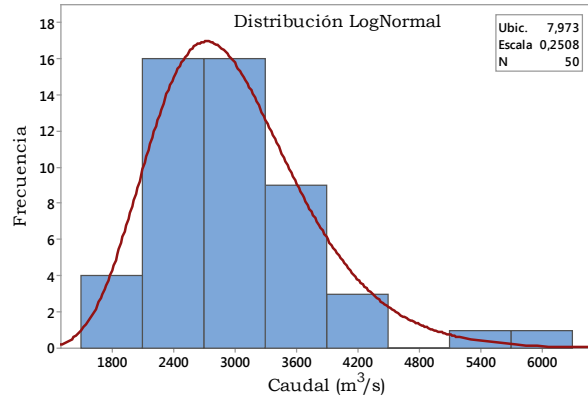
| Estación | Corriente | Distribución |
|--------------|---------------|--------------|
| Montelíbano | San Jorge | Gumbel |
| La Coquera | Cauca | LogNormal |
| La Esperanza | Nechi | LogNormal |
| La raya | Caño Caribona | Gumbel |
| Armenia | Brazo de Loba | Normal |

De igual forma que para el caso 4 visto en esta sección, el parámetro de la cópula se obtuvo por medio del paquete “copula” de R de acuerdo con el estadístico Cramer-von-Mises. La cópula que mejor se ajusta a los datos es aquella que tiene el mayor p-valor. Los resultados obtenidos, al aplicar la metodología propuesta por Kojadinovic et al. (2010) se presentan en las tablas 4-5 y 4-6.

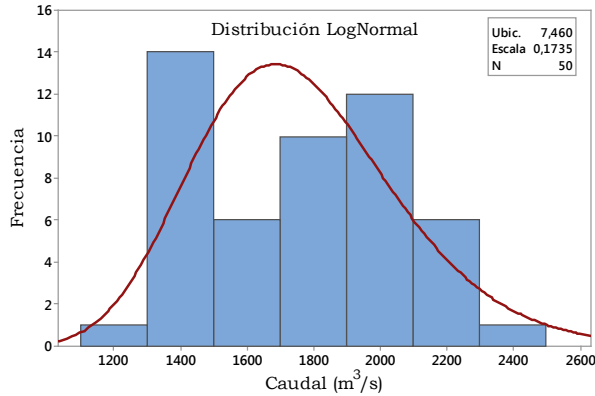
Una vez obtenidas las distribuciones marginales y el parámetro de la cópula, y al aplicar la ecuación 4-49 se obtienen las figuras 4-13 - 4-17.



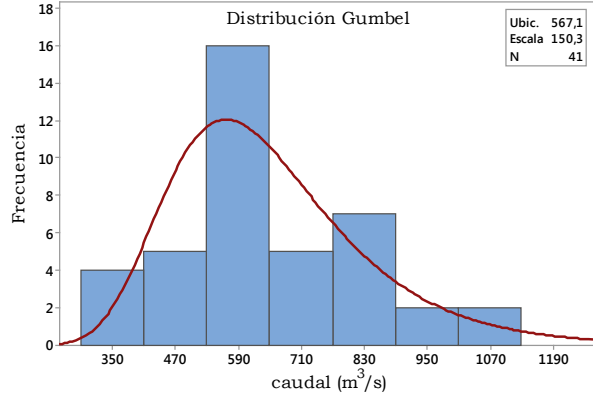
(a) Distribución Gumbel ajustada para la estación Montelíbano - San Jorge.



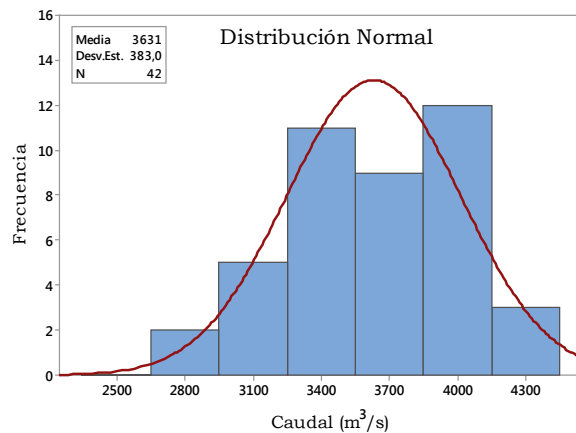
(b) Distribución LogNormal ajustada para la estación La Coquera - Cauca.



(c) Distribución LogNormal ajustada para la estación La Esperanza - Nechi.



(d) Distribución Gumbel ajustada para la estación La Raya - Caño Caribona.



(e) Distribución Normal ajustada para la estación Armenia - Brazo de Loba.

Figura 4-12.: Histograma de frecuencia ajustado a diferentes distribuciones de probabilidad mediante la prueba Kolmogorov-Smirnov para las estaciones pertenecientes a la región de la Mojana.

Tabla 4-5.: Parámetro θ bivariado para las estaciones pertenecientes a la región de la Mojana

| | Cópula | θ | p-valor | | Cópula | θ | p-valor |
|-----------|---------|----------|---------|-----------|---------|----------|---------|
| Cauca | Gumbel | 1,839 | 0,260 | Nechi | Gumbel | 1,117 | 0,187 |
| + | Frank | 4,994 | 0,410 | + | Frank | 0,952 | 0,461 |
| Nechi | Clayton | 1,679 | 0,029 | Caribona | Clayton | 0,234 | 0,252 |
| | Joe | 2,548 | 0,308 | | Joe | 1,205 | 0,059 |
| San Jorge | Gumbel | 1,429 | 0,303 | San Jorge | Gumbel | 1,178 | 0,125 |
| + | Frank | 2,917 | 0,462 | + | Frank | 1,387 | 0,318 |
| Cauca | Clayton | 0,857 | 0,054 | Br. Loba | Clayton | 0,356 | 0,053 |
| | Joe | 1,772 | 0,083 | | Joe | 1,314 | 0,051 |

Tabla 4-6.: Parámetro θ multivariado para las estaciones pertenecientes a la región de la Mojana

| | Cópula | θ | p-valor | Error | | Cópula | θ | p-valor | Error |
|--------------|---------|----------|---------|-------|---------------|---------|----------|---------|-------|
| San Jorge | Gumbel | 1,394 | 0,977 | 0,157 | Br. Loba | Gumbel | 1,356 | 0,217 | 0,114 |
| + Brazo Loba | Frank | 2,489 | 0,644 | 0,889 | + Caribona | Frank | 2,378 | 0,123 | 0,658 |
| + Caribona | Clayton | 0,491 | 0,130 | 0,178 | + Nechi | Clayton | 0,626 | 0,111 | 0,154 |
| | Joe | 1,635 | 0,942 | 0,211 | | Joe | 1,491 | 0,049 | 0,172 |
| San Jorge | Gumbel | 1,348 | 0,00649 | 0,100 | San Jorge | Gumbel | 1,398 | 0,436 | 0,112 |
| + Cauca | Frank | 2,373 | 0,00350 | 0,573 | + Cauca | Frank | 2,584 | 0,152 | 0,664 |
| + Nechi | Clayton | 0,482 | 0,00050 | 0,124 | + Nechi | Clayton | 0,567 | 0,053 | 0,131 |
| | Joe | 1,530 | 0,00150 | 0,147 | + Caribona | Joe | 1,618 | 0,316 | 0,143 |
| Cauca | Gumbel | 1,410 | 0,08641 | 0,109 | San Jorge + | Gumbel | 1,403 | 0,608 | 0,102 |
| + Nechi | Frank | 2,814 | 0,08641 | 0,739 | Cauca + Nechi | Frank | 2,525 | 0,253 | 0,541 |
| + Caribona | Clayton | 0,766 | 0,04446 | 0,231 | + Caribona | Clayton | 0,552 | 0,065 | 0,093 |
| | Joe | 1,540 | 0,00550 | 0,137 | + Brazo Loba | Joe | 1,641 | 0,372 | 0,146 |

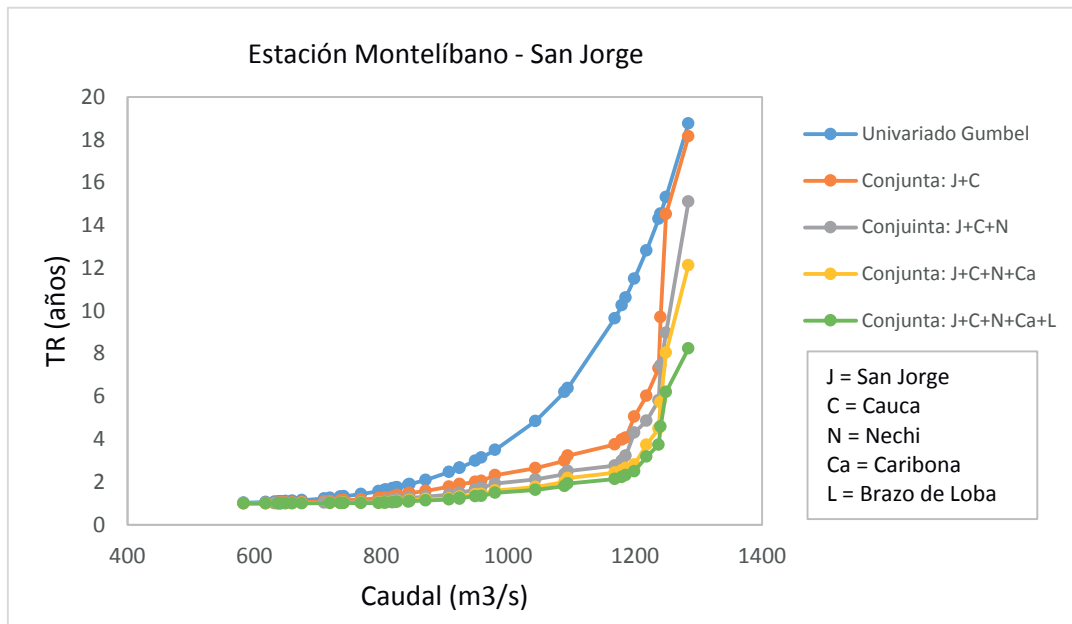


Figura 4-13.: Periodo de retorno multivariado multisitio para la estación Montelíbano - San Jorge.

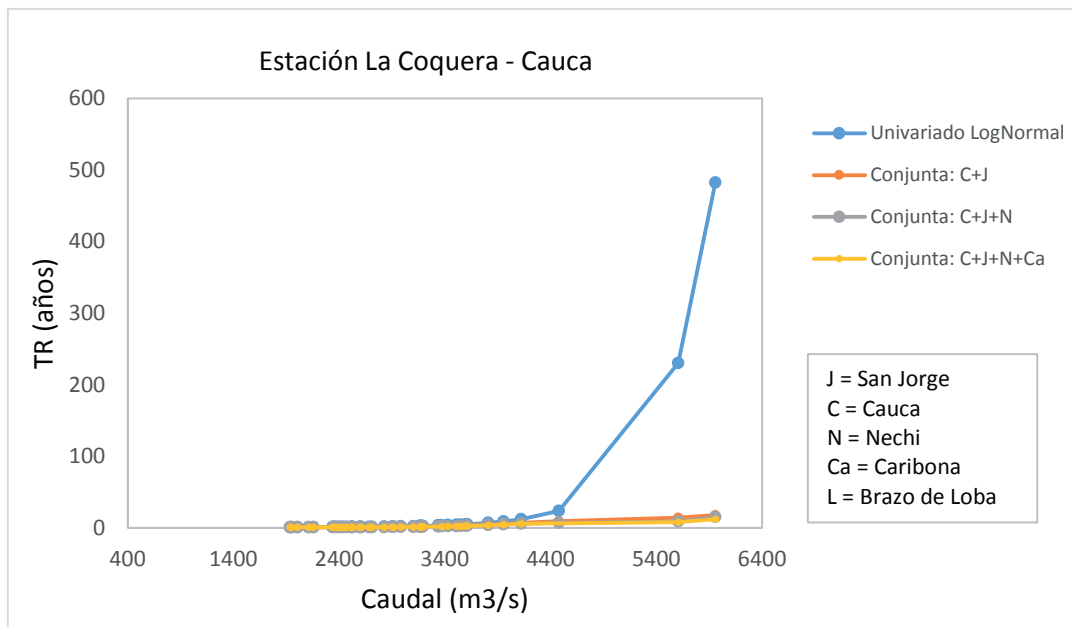


Figura 4-14.: Periodo de retorno multivariado multisitio para la estación La Coquera - Cauca.

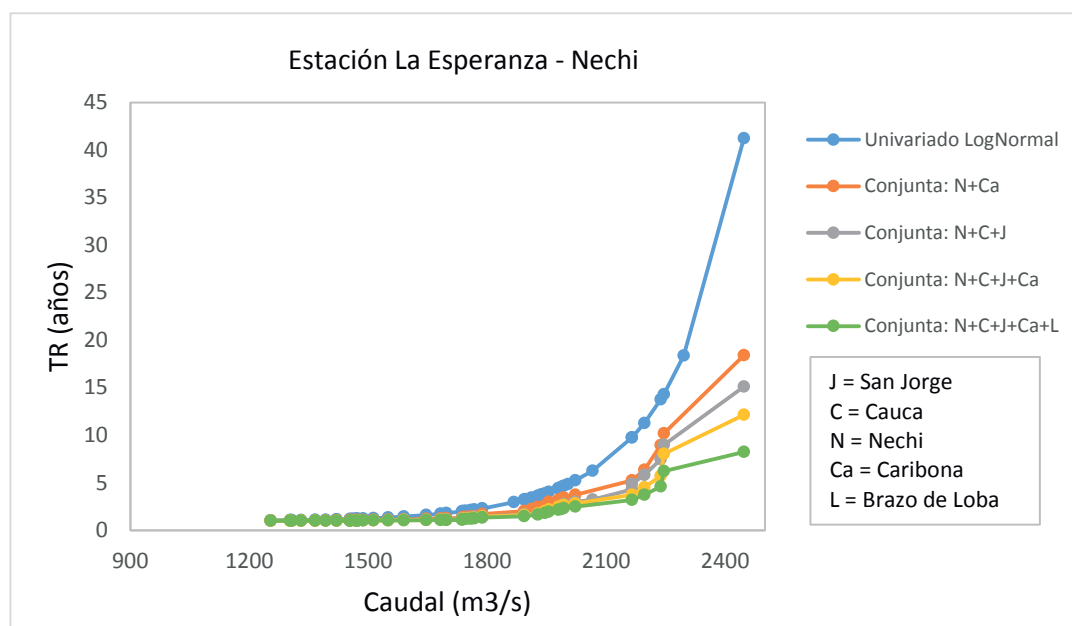


Figura 4-15.: Periodo de retorno multivariado multisitio para la estación La Esperanza - Nechi.

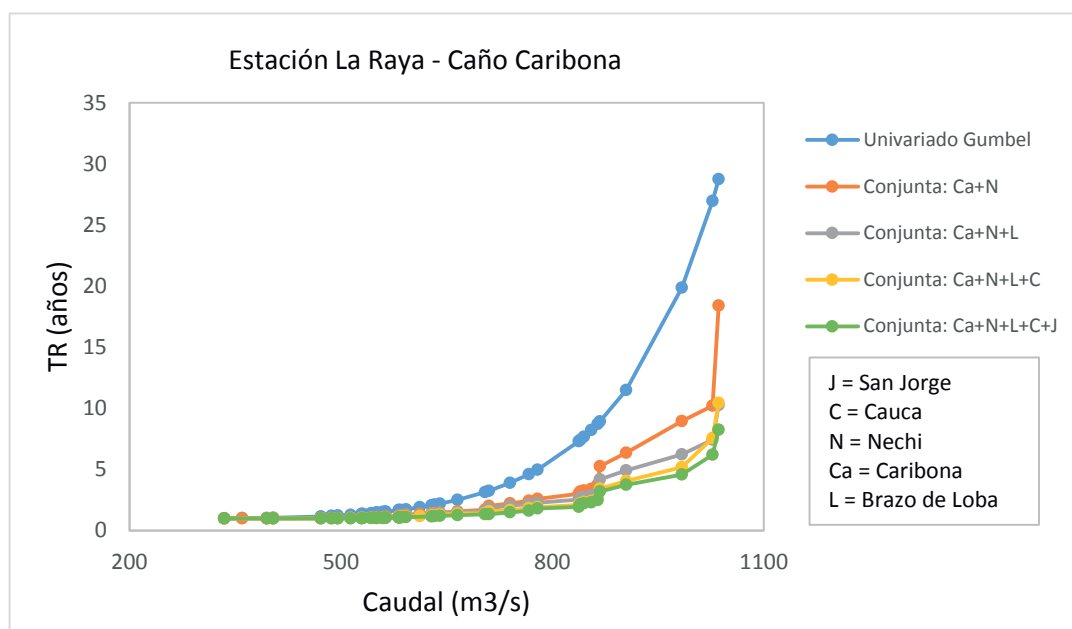


Figura 4-16.: Periodo de retorno multivariado multisitio para la estación La Raya - Caño Caribona.

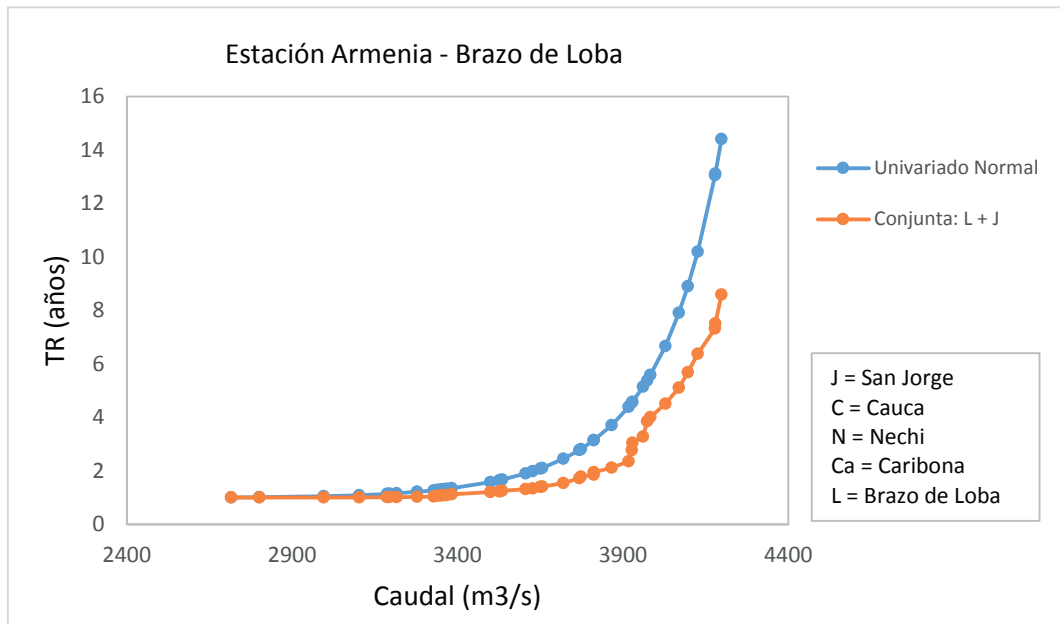


Figura 4-17.: Periodo de retorno multivariado multisitio para la estación Armenia - Brazo de Loba.

4.9.5. Análisis de resultados del periodo de retorno conjunto multivariado multisitio

El principal reto de la aplicación de modelos estocásticos multivariados se debe a la complejidad matemática de la distribución de probabilidad conjunta que abarca el conocimiento de las distribuciones marginales y estructuras de dependencia. En estudios previos (Eagleson, 1972; Díaz-Granados et al., 1984), el uso de la distribución conjunta fue acompañado generalmente con el supuesto de independencia entre diferentes variables (es decir, ignorando la posible estructura de dependencia). La importancia de la estructura de la dependencia en el análisis de frecuencia de crecidas se acentúa por Singh and Singh (1991), Bacchi et al. (1994) y Goel et al. (2000) entre otros.

Como ya se ha mencionado, la cópula de mejor ajuste es aquella que presenta el mayor p-valor de acuerdo con el estadístico Cramer-von-Mises. La tabla 4-5 muestra los resultados obtenidos del parámetro θ de las diferentes cópulas bivariadas estudiadas para cuatro combinaciones diferentes de las estaciones seleccionadas en la región de la Mojana. Como puede observarse, la cópula Frank en todos los casos presenta el mayor p-valor y por lo tanto es la cópula que mejor se ajusta a los registros históricos para el caso bivariado. Una vez más esta cópula ha demostrado ser un modelo adecuado para caracterizar eventos hidrológicos extremos bivariados ya que abarca todo el rango de correlación de los datos que se encuentran en $[-1,1]$.

La tabla 4-6 presenta los resultados obtenidos del parámetro θ para las diferentes cópulas extendidas al caso multivariado $N \geq 3$ siendo N el número de estaciones. Este parámetro se determinó para 6 diferentes combinaciones aplicadas en las estaciones seleccionadas. De esta tabla se puede concluir que la cópula Gumbel-Hougaard es adecuada para eventos hidrológicos extremos multivariados ya que presenta el mayor p-valor en todas las combinaciones. Los periodos de retorno trivariado multisitio se obtuvieron por medio de la ecuación 4-49, (esta ecuación es fácil de extender al caso multivariado) y los resultados se presentan en las figuras 4-13 - 4-17. La figura 4-13 presenta los periodos de retorno conjuntos para cada una de las observaciones, esta figura muestra que a medida que se adiciona una estación, el periodo de retorno va disminuyendo, esta forma es útil para saber si el método de la cópula puede proporcionar resultados razonables en una región grande, o si su aplicación se limita a unas pocas estaciones seleccionadas.

La figura 4-14 muestra que el periodo de retorno univariado ajustado a la distribución Log-Normal es cercano a los 500 años para el evento máximo registrado, al aplicar el método de la cópula, el periodo de retorno conjunto baja hasta un valor cercano a los 20 años, esto es

debido a que el evento máximo registrado para la estación La Coquera - Cauca se presentó en el mes de noviembre de 2010, temporada correspondiente al fenómeno de La Niña 2010-2011, una de las olas invernales más fuertes de la historia de nuestro país. De acuerdo con el ONI (Índice oceánico de El Niño), el evento 2010-2011 inició en el mes de julio de 2010 y terminó hasta mayo de 2011. (Euscátegui and Hurtado, 2011). El evento registrado en 2010 por la estación La Coquera - Cauca no es la única la razón por la cual el periodo de retorno conjunto de 20 años no presenta resultados razonables, la principal razón de este comportamiento es debido a que ninguna de las otras cuatro estaciones registraron el evento del fenómeno de La Niña 2010-2011. El método de la cópula se puede aplicar siempre y cuando en todas las estaciones se tenga registro de este tipo de eventos. Para este caso en particular, como sólo una estación registró el evento del 2010-2011 es recomendable aplicar el método de la cópula en este lugar o en esta estación pero con el conocimiento de la información completa sobre el evento de inundación (pico de crecida, el volumen de inundación, la duración de inundaciones, la forma del hidrograma, etc.)

Las figuras **4-15** y **4-16**, presentan el mismo comportamiento de la figura **4-13**. A medida que se adicionan estaciones, el periodo de retorno conjunto disminuye. Finalmente, para saber si el método de la cópula presenta resultados razonables para una región extensa como la Mojana, se realizó la comparación de los periodos de retorno conjuntos de las dos estaciones más alejadas, éstas son la estación Montelíbano - San Jorge y la estación Armenia - Brazo de Loba. La figura **4-17**, muestra que el periodo de retorno conjunto es menor con respecto al ajustado por medio de la distribución Normal. Esta gráfica demuestra que la región es estadísticamente homogénea y que presenta tendencias similares, por lo tanto el método de la cópula es recomendable. En conclusión, el periodo de retorno conjunto tiende a disminuir a medida que se adicionan estaciones y siempre está por debajo del periodo de retorno univariado. El procedimiento univariado, generalmente sobreestima el riesgo asociado al evento.

4.10. Conclusiones del capítulo

El uso de cópulas en hidrología cada vez es más frecuente y aún son muchos los resultados que se pueden descubrir. En este capítulo se mostró un nuevo enfoque para la construcción de curvas IDF conjuntas multisitio a través de un método de interpolación local. Como se presentó en la tabla 4-2, las diferencias no son muy significativas, llegando un máximo de 32 %, lo que puede dar lugar a una sub o sobreestimación del riesgo para el caso empírico. La importancia fundamental para aplicar un nuevo método es ser capaz de vincular las distribuciones marginales a través de una cópula para obtener una distribución conjunta que describa las características principales del evento hidrológico, en este caso se aplicó un método de interpolación para obtener una curva IDF conjunta multisitio. El método de la cópula dependerá del método de interpolación.

De acuerdo con los resultados obtenidos para el caso de estudio de la Mojana, el alcance del método de la cópula para extensiones similares a esta región es válido ya que presenta resultados razonables. Por otra parte, incluso si la cópula parece bien adaptada a los valores observados, la precaución es necesaria antes de usarla para fenómenos como el ENSO (EL Niño-Oscilación del Sur), varios criterios pueden ser utilizados para este fin, sin embargo sólo reflejarán una característica particular de los datos, es necesario analizar cada estación seleccionada por separado y comprobar si registraron y fueron afectadas por este tipo de fenómenos. Cuando una sola estación registra el evento, como en este caso, es recomendable aplicar el método de la cópula con información más precisa como el pico de crecida, el volumen de inundación y su duración.

El caso de estudio de la Mojana ha demostrado que el método de la cópula para especificar la función de distribución multivariada es de gran alcance, ya que evita el requisito de que las distribuciones marginales sean del mismo tipo que se asumen en la mayoría de los estudios de distribuciones multivariadas empíricos. Además, evita las fórmulas complejas que surgen para muchas funciones de distribución multivariada. Aun así, no todas las cópulas arquimedianas son apropiadas para la representación de las distribuciones de probabilidad multivariada, los periodos de retorno conjunto de la región de la Mojana para más de dos estaciones están representados por la cópula de mejor ajuste Gumbel-Hougaard que podría ser una candidata adecuada para el análisis de frecuencia hidrológico multivariado.

Muchas propuestas se han hecho recientemente para las pruebas de bondad y ajuste para la selección de cópulas; el criterio de información de Akaike (AIC), el criterio el criterio de información bayesiano (BIC) y el error cuadrático medio (RMSE) son ampliamente utilizados para la selección de cópulas apropiadas así como otros múltiples modelos trivariados. Aunque

se han propuesto varias pruebas de bondad de ajuste, no existen pautas generales para la selección de la cópula paramétrica óptima. Como se vio en la sección 3.10, el paquete “copula” de R se puede utilizar para pruebas de bondad y ajuste mediante el estadístico Cramer-von Mises en la construcción de un modelo de datos para la selección de la cópula. El problema de cómo elegir un modelo que se adapte a una cópula para un conjunto determinado de datos es una limitación del enfoque predominante de la cópula.

5. Conclusiones y recomendaciones

5.1. Conclusiones Generales

Este documento presenta un marco general del método de la cópula para el estudio de los periodos de retorno de eventos hidrológicos, proporcionando un medio simple, pero eficaz para realizar análisis de riesgo. Como resultado fundamental, se muestra cómo el uso de cópulas puede simplificar en gran medida el estudio de las distribuciones condicionales que introducen el concepto de periodo de retorno condicional utilizado para la construcción de curvas IDF.

Las distribuciones multivariadas presentes en la literatura se pueden escribir de una manera directa en términos de cópulas adecuadas. En este trabajo, las distribuciones de frecuencia de precipitaciones bivariadas se obtuvieron mediante el método de la cópula sin asumir la misma forma de las distribuciones marginales. Esto no es posible de obtener mediante los métodos clásicos estándar.

El método de la cópula tiene la flexibilidad en la construcción de distribuciones de probabilidad conjunta, proporcionando una evaluación probabilística más precisa de las variables aleatorias que se quieren estudiar ya que se basan en estructuras de dependencia estableciendo información más confiable de las tormentas de diseño y los riesgos asociados.

Se encontró que la dependencia de variables aleatorias de la intensidad y duración para las estaciones pluviográficas presentan correlación negativa, además, para el caso de estudio de la Mojana, la dependencia de correlación bivariada de las estaciones LM y LG es positiva. La familia de cópulas arquimedianas Frank ha demostrado ser un modelo adecuado para la caracterización de estos comportamientos de dependencia.

En las pruebas de bondad y ajuste, como en cualquier contexto inferencial, entre mayor el tamaño de la muestra, mayor es la ayuda para distinguir los modelos que desempeñan un papel importante en la fiabilidad de los procesos utilizados para la selección de la cópula. Para cópulas multidimensionales estas pruebas pueden proporcionar más apoyo para una

gama más amplia de cópulas que pueden describir asociaciones de variables hidrológicas que pueden estar correlacionadas, sin embargo, se requieren esfuerzos computacionales más complicados. Este trabajo demuestra que la selección de la cópula por medio del estadístico Cramer-von Mises es válido tanto para el enfoque paramétrico como para el semiparamétrico.

5.2. Recomendaciones

Con el fin de recomendar un modelo adecuado para el diseño, se estudiaron cinco familias de cópulas, observando que las familias Clayton y Gumbel se desempeñan bien en los casos de dependencia positivos y que la familia Frank es adecuada para toda la gama de dependencia. Sobre la base de los resultados y observaciones, la familia de la cópula Frank fue adoptado en este estudio y se recomienda para la construcción de curvas IDF.

El análisis de frecuencia univariado de crecidas es ampliamente utilizado en estudios hidrológicos que a menudo estudia estadísticamente el pico o el volumen de inundación. La gravedad de una inundación se define no sólo por el valor máximo de crecida, sino también por otros aspectos del evento extremo como su volumen y duración. Para un análisis más completo se recomienda estudiar el evento de inundación descrito como un evento multivariado cuyas principales características se pueden resumir por su pico, volumen y duración, que se correlacionan mutuamente. Es claro que el enfoque estadístico multivariado por medio de cópulas es necesario.

Un paso importante en el proceso de modelado es la elección de la función cópula de mejor ajuste. Es necesario estudiar más familias de cópulas que puedan reproducir mejores estructuras de dependencia de las variables hidrológicas multivariadas. En este trabajo se estudiaron cinco familias de cópulas arquimedianas, sin embargo, en la literatura existen diversas familias que se han aplicado a estudios hidrológicos, ejemplos de éstas son las cópulas de valor extremo, las cópulas elípticas o la cópula normal, entre otras.

La elección de una cópula no es sencilla de modelar y obtener, especialmente si el conjunto de datos no es lo suficientemente informativo para proporcionar indicaciones pertinentes acerca de las propiedades de dependencia. Uno de los principales problemas tiene que ver con la incertidumbre, pues ésta es parte fundamental para la evaluación precisa de los periodos de retorno. Las principales fuentes de incertidumbre en las curvas IDF se deben a la calidad y cantidad de los datos, conduciendo a incertidumbres en los parámetros de los múltiples modelos utilizados. Es importante cuantificar estas incertidumbres y utilizarlas en las decisiones

de diseño y análisis de riesgos. Es recomendable el enfoque bayesiano el cual proporciona un marco coherente para la incorporación de estas incertidumbres. (Algunos ejemplos de estos estudios se encuentran en Chandra et al. (2015), Van de Vyver (2015a), Huard et al. (2010), Prodanovic and Simonovic (2007), Coles and Tawn (1996)).

A. Anexo: Capítulo 2. Distribuciones de Probabilidad y Análisis de frecuencia

1. Distribuciones de probabilidad usadas en este trabajo

Distribución Exponencial

La distribución exponencial se utiliza para describir los tiempos de interarribo de choques aleatorios a sistemas hidrológicos. (Chow et al., 1994). La función de densidad de probabilidad (PDF) es expresada como sigue:

Una variable aleatoria X tiene una distribución exponencial si su función de densidad está dada por:

$$f(x, \lambda) = \begin{cases} \lambda e^{-\lambda x}, & \text{si } x > 0; \lambda > 0, \\ 0, & \text{de otra manera,} \end{cases} \quad (\text{A-1})$$

donde λ es igual a $1/\beta$

Y su función de distribución de probabilidad acumulada (CDF) está dada por:

$$P(0 \leq X \leq x) = 1 - e^{-\lambda x}$$

$$F(x) = 1 - e^{-\lambda x} \quad (\text{A-2})$$

En otros textos se puede encontrar que la función de densidad de probabilidad (PDF) exponencial es expresada como (Zhang, 2005):

$$f(x) = \frac{1}{\alpha} \exp\left(-\frac{x - \beta}{\alpha}\right) \quad (\text{A-3})$$

donde su función de distribución de probabilidad acumulada (CDF) es:

$$F(x) = 1 - \exp\left(-\frac{x - \beta}{\alpha}\right) \quad (\text{A-4})$$

donde α es el parámetro de forma y β es el parámetro de ubicación o localización de la distribución exponencial. Estos parámetros pueden ser estimados por el método de los momentos (MOM), estimación de máxima verosimilitud (MLE), y momentos lineales (LM). Ver Chow et al. (1994).

La distribución exponencial puede ser aplicada a una serie de datos de caudal como se muestra en la figura **A-1**

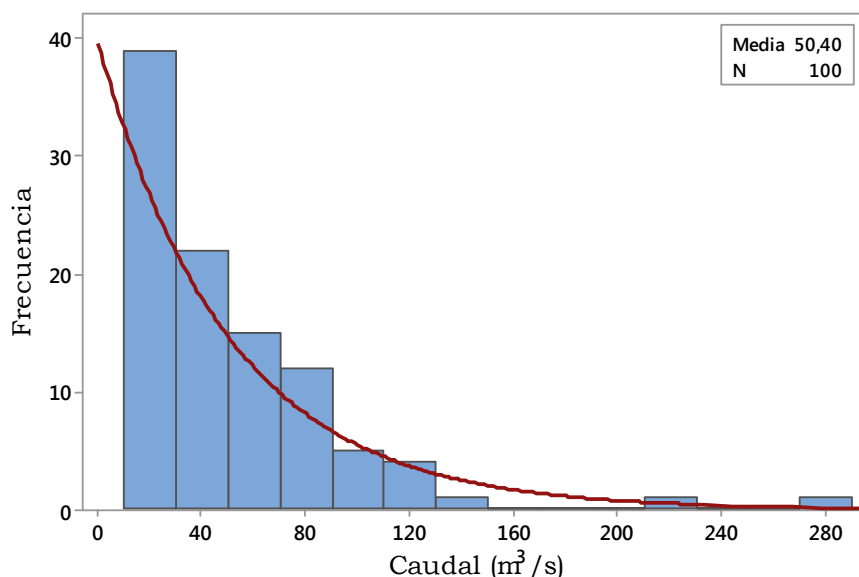


Figura A-1.: Distribucion Exponencial.

Distribución Normal

La distribución normal surge del *teorema del límite central* Chow et al. (1994). Es sin duda con la distribución Lognormal las más usadas en ciencias e ingeniería. La función de densidad de probabilidad de una variable aleatoria normal es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right), \quad \text{para } -\infty < x < \infty \quad (\text{A-5})$$

donde $\mu = \bar{x}$, $\sigma = s_x$

La distribución normal es simétrica con respecto a su media μ y x admite valores entre $-\infty < x < \infty$ por lo tanto, no siempre es satisfactoria para el modelado de fenómenos físicos e hidrológicos, tales como eventos extremos. La función de distribución de probabilidad acumulada es:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \quad (\text{A-6})$$

Generalmente la distribución normal tiene forma de campana, como se muestra en la figura **A-2**, y se puede aplicar a una serie de datos hidrológicos:

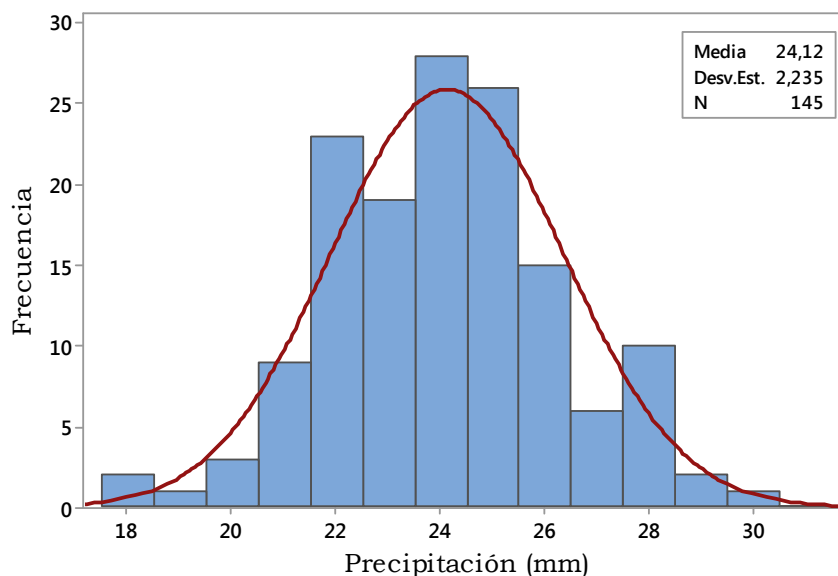


Figura A-2.: Distribución Normal.

Distribución Lognormal

Un modelo utilizado con frecuencia para distribuciones asimétricas es la distribución Lognormal. Si una variable aleatoria $Y = \log X$ está normalmente distribuida, entonces se dice que X está distribuida en forma lognormal. La PDF de la distribución lognormal se expresa como:

$$f(x) = \frac{1}{x\sigma_y\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu_y)^2}{2\sigma_y^2}\right), \quad \text{para } x > 0 \quad (\text{A-7})$$

donde $\mu_y = \bar{y}$, $\sigma_y = s_y$ y $y = \ln x$.

La CDF de la distribución lognormal se expresa como:

$$F(x) = \frac{1}{\sigma_y \sqrt{2\pi}} \int_0^x \frac{1}{x} \exp\left(-\frac{(\ln x - \mu_y)^2}{2\sigma_y^2}\right) dx \quad (\text{A-8})$$

Donde μ_y y σ_y son la media y la desviación estándar de logaritmo natural de x .

A diferencia de la distribución normal, la distribución logarítmica normal es ligeramente sesgada a la derecha como se muestra en la figura **A-3** y se aplica también para datos de caudal y análisis de la escorrentía.

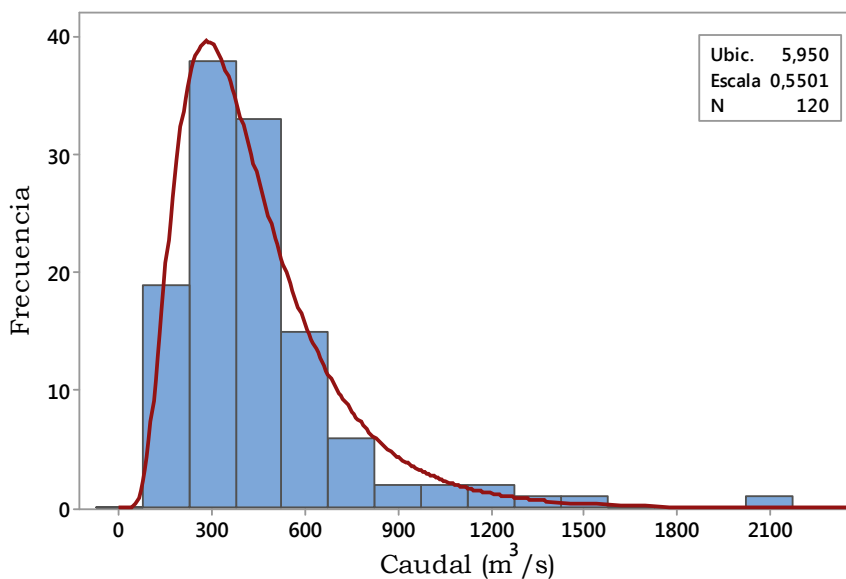


Figura A-3.: Distribución Lognormal.

Los parámetros de la distribución lognormal pueden ser estimados por MOM, MLE, y LM.

Distribución Gamma

Antes de enunciar la distribución gamma, es necesario presentar algunos detalles de la función gamma.

Función Gamma $\Gamma(\beta)$

Es una función que extiende el concepto de factorial a los números complejos. Fue presentada, en primera instancia, por Leonard Euler entre los años 1730 y 1731. (Arroyo et al., 2014)

La *función gamma* $\Gamma(\beta)$ se define como:

sea $\Gamma : (0, \infty) \rightarrow R$, donde

$$\Gamma(\beta) = \int_0^{\infty} x^{\beta-1} e^{-x} dx, \quad \text{Para } \beta > 0 \quad (\text{A-9})$$

Al integrar por partes tomando $u = x^{\beta-1}$ y $dv = e^{-x} dx$ se obtiene:

$$\Gamma(\beta) = (\beta - 1) \int_0^{\infty} x^{\beta-2} e^{-x} dx, \quad (\text{A-10})$$

En donde $\Gamma(\beta) = (\beta - 1)\Gamma(\beta - 1)$, con lo cual: $\Gamma(1) = \int_0^{\infty} e^{-x} dx = 1$, y de aquí $\Gamma(n) = (n - 1)!$

Algunas propiedades adicionales de $\Gamma(\beta)$ son:

- $\Gamma(n + 1) = n!$ si n es un entero positivo
- $\Gamma(n + 1) = n\Gamma(n)$, $n > 0$
- $\Gamma(1/2) = \sqrt{\pi}$

Distribución Gamma

La distribución gamma se ha utilizado para modelar muchos fenómenos naturales, incluidos los caudales diarios, mensuales y anuales, así como los flujos de inundación (Bobée and Ashkar, 1991). Se le conoce, también, como una generalización de la distribución exponencial. Es una distribución de probabilidad continua adecuada para modelar el comportamiento de variables aleatorias con asimetría positiva y/o los experimentos en donde está involucrado el tiempo (Arroyo et al., 2014).

Una variable aleatoria X tiene una distribución gamma si su función de densidad está dada por:

$$f(x, \beta, \alpha) = \begin{cases} \frac{1}{\alpha^\beta \Gamma(\beta)} x^{\beta-1} e^{-x/\alpha}, & \text{para } x > 0; \beta, \alpha > 0; \\ 0, & \text{de otra manera.} \end{cases} \quad (\text{A-11})$$

La PDF de la distribución gamma es expresada como:

$$f(x) = \frac{1}{\alpha^\beta \Gamma(\beta)} x^{\beta-1} e^{-x/\alpha} \quad (\text{A-12})$$

Si $\alpha = 1/\lambda$ entonces la PDF de la distribución gamma sería:

$$f(x) = \frac{\lambda^\beta x^{\beta-1} e^{-\lambda x}}{\Gamma(\beta)} \quad (\text{A-13})$$

donde $\Gamma(\beta)$ es la función gamma para $x \geq 0$.

Los parámetros de la distribución gamma pueden ser estimados por MOM, MLE, y LM. (Chow et al., 1994).

$$\lambda = \frac{\bar{x}}{s_x^2} \quad y \quad \beta = \frac{\bar{x}^2}{s_x^2} = \frac{1}{CV^2}$$

La CDF de la distribución gamma está dada por:

$$F(x) = \frac{1}{\alpha^\beta \Gamma(\beta)} \int_0^x x^{\beta-1} e^{-x/\alpha} dx \quad (\text{A-14})$$

De igual forma que la distribución Lognormal, la distribución gamma también es sesgada a la derecha como se muestra en la figura **A-4**, además se puede aplicar para el análisis de frecuencia de datos de precipitación y caudal.

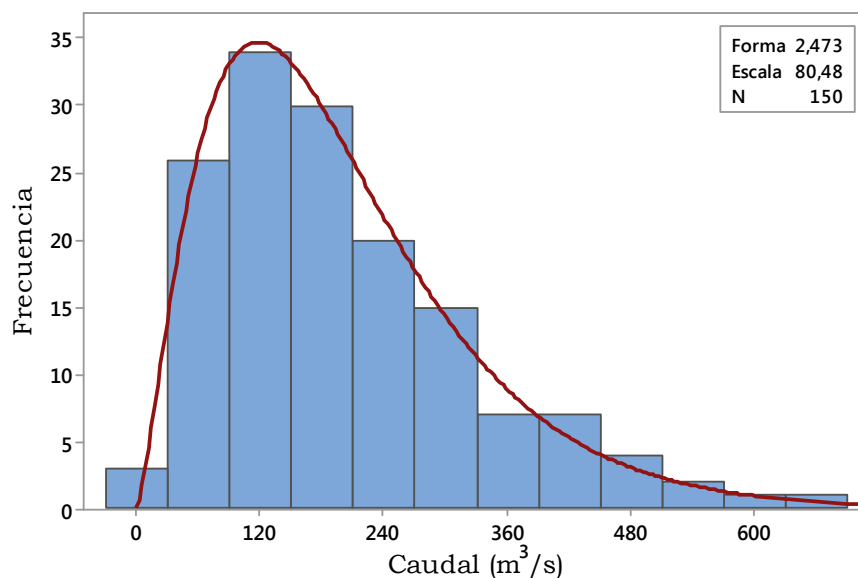


Figura A-4.: Distribución Gamma.

Distribución Pearson tipo III o Gamma de tres parámetros

La distribución Pearson tipo III, también llamada la *distribución gamma de tres parámetros*, introduce un tercer parámetro, el límite inferior γ . Ésta es una distribución muy flexible que puede asumir diferentes formas a medida que λ , β y γ varían. (Chow et al., 1994).

La PDF de la distribución Pearson tipo III es expresada como:

$$f(x) = \frac{1}{\alpha\Gamma(\beta)} \left(\frac{x - \gamma}{\alpha} \right)^{\beta-1} \exp\left(-\frac{x - \gamma}{\alpha}\right) \quad (\text{A-15})$$

Si $\alpha = 1/\lambda$ entonces la PDF de la distribución Pearson tipo III sería:

$$f(x) = \frac{\lambda^\beta (x - \gamma)^{\beta-1} e^{-\lambda(x-\gamma)}}{\Gamma(\beta)} \quad (\text{A-16})$$

donde $\Gamma(\beta)$ es la función gamma para $x \geq \gamma$.

Los parámetros de la distribución Pearson tipo III pueden ser estimados por MOM, MLE, y LM. (Chow et al., 1994).

$$\lambda = \frac{s_x}{\sqrt{\beta}} \quad , \quad \beta = \left(\frac{2}{C_s}\right)^2 \quad y \quad \gamma = \bar{x} - s_x\sqrt{\beta}$$

La CDF de la distribución Pearson tipo III está dada por:

$$F(x) = \frac{1}{\alpha\Gamma(\beta)} \int_{\gamma}^x \left(\frac{x-\gamma}{\alpha}\right)^{\beta-1} \exp\left(-\frac{x-\gamma}{\alpha}\right) dx \quad (\text{A-17})$$

Al igual que las distribuciones exponencial y gamma, la distribución pearson tipo III también es sesgada a la derecha como se muestra en la figura **A-5** , que por lo general se aplica en el análisis de frecuencia de crecidas.

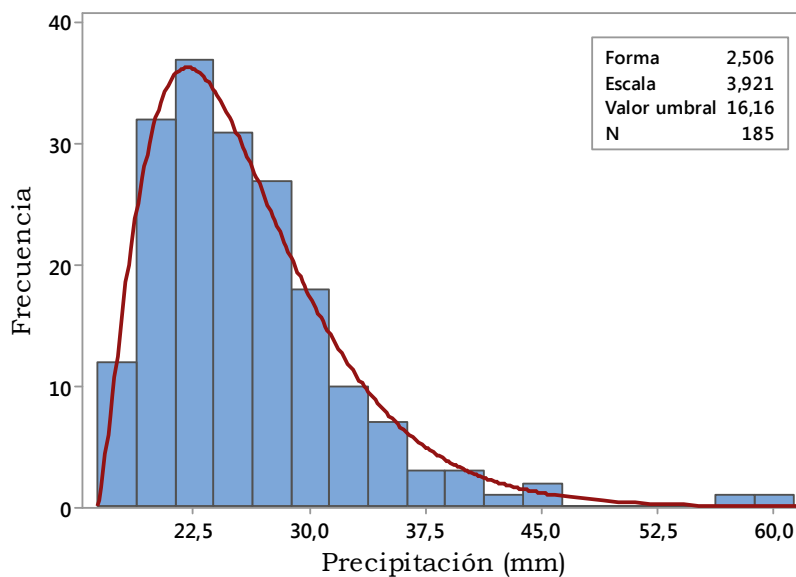


Figura A-5.: Distribución Pearson Tipo III.

Distribución Log-Pearson tipo III

Si $\log x$ sigue una distribución Pearson tipo III, entonces se dice que X sigue una distribución log-pearson tipo III. Ésta es la distribución estandar para análisis de frecuencia de crecientes máximas anuales en los Estados Unidos. Chow et al. (1994).

La PDF de la distribución Log pearson tipo III es expresada como:

$$f(x) = \frac{1}{\alpha x \Gamma(\beta)} \left(\frac{\ln x - \gamma}{\alpha} \right)^{\beta-1} \exp \left(-\frac{\ln x - \gamma}{\alpha} \right) \quad (\text{A-18})$$

Si $\alpha = 1/\lambda$ y además $y = \ln x$ entonces la PDF de la distribución Log pearson tipo III sería:

$$f(x) = \frac{\lambda^\beta (y - \gamma)^{\beta-1} e^{-\lambda(y-\gamma)}}{x \Gamma(\beta)} \quad (\text{A-19})$$

donde $\Gamma(\beta)$ es la función gamma para $\ln x \geq \gamma$.

Los parámetros de la distribución Log pearson tipo III pueden ser estimados por MOM, MLE, y LM. (Chow et al., 1994).

$$\lambda = \frac{s_y}{\sqrt{\beta}} \quad , \quad \beta = \left(\frac{2}{C_s(y)} \right)^2 \quad y \quad \gamma = \bar{y} - s_y \sqrt{\beta} \quad \text{suponiendo que } C_s(y) \text{ es positivo.}$$

La CDF de la distribución Log pearson tipo III está dada por:

$$F(x) = \frac{1}{\alpha \Gamma(\beta)} \int_0^x \frac{1}{x} \left(\frac{\ln x - \gamma}{\alpha} \right)^{\beta-1} \exp \left(-\frac{\ln x - \gamma}{\alpha} \right) dx \quad (\text{A-20})$$

Una vez más, la distribución log-Pearson tipo III, como se muestra en la figura **A-6**, se aplica generalmente para el análisis de frecuencia de crecidas. (Zhang, 2005).

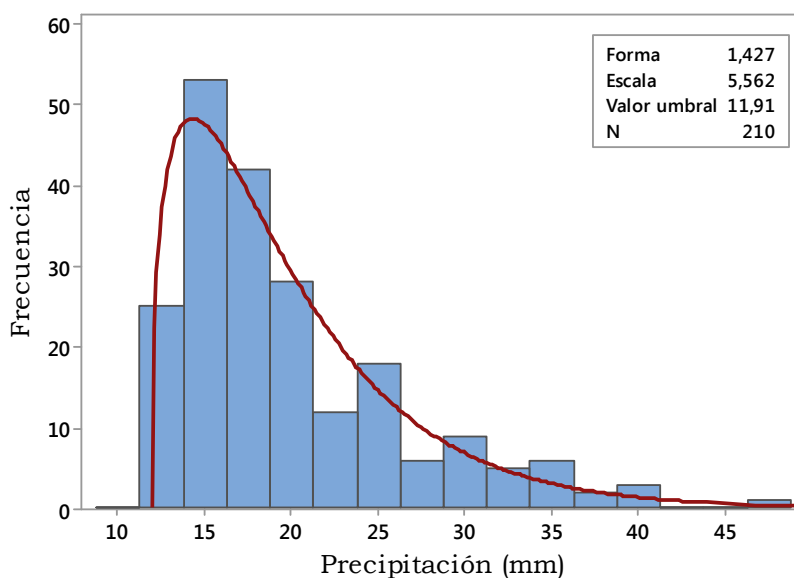


Figura A-6.: Distribución Log Pearson Tipo III.

Distribución Gumbel o valor extremo tipo I

Fisher and Tippett (1928) han demostrado que las distribuciones de valores extremos seleccionados de conjuntos de muestras de cualquier distribución de probabilidad convergen a una de las tres formas de distribuciones de valor extremo, llamadas tipo I, II y III respectivamente, cuando el número de valores extremos seleccionados es grande. (Chow et al., 1994). La distribución de Valor Extremo (EV) I, o la distribución Gumbel, se ha utilizado a menudo para describir los flujos de inundación. Se podría esperar que la distribución de los flujos de inundación máximos anuales pertenecería al conjunto de distribuciones de valores extremos. (Gumbel, 1958; Kottegoda and Rosso, 1997).

La PDF de la distribución Gumbel es expresada como:

$$f(x) = \frac{1}{\alpha} \exp \left[- \left(\frac{x - \beta}{\alpha} \right) - \exp \left(- \frac{x - \beta}{\alpha} \right) \right] \quad - \infty < x < \infty \quad (\text{A-21})$$

Los parámetros de la distribución Gumbel pueden ser estimados por MOM, MLE, y LM. (Chow et al., 1994).

$$\alpha = \frac{\sqrt{6}s_x}{\pi} \quad , \quad \beta = \bar{x} - 0,5772\alpha$$

La CDF de la distribución Gumbel está dada por:

$$F(x) = \exp \left[- \exp \left(- \frac{x - \beta}{\alpha} \right) \right] \quad (\text{A-22})$$

La distribución Gumbel es otra distribución sesgada a la derecha, como se muestra en la Figura A-7 y ha sido utilizada en el análisis de frecuencia de crecidas.

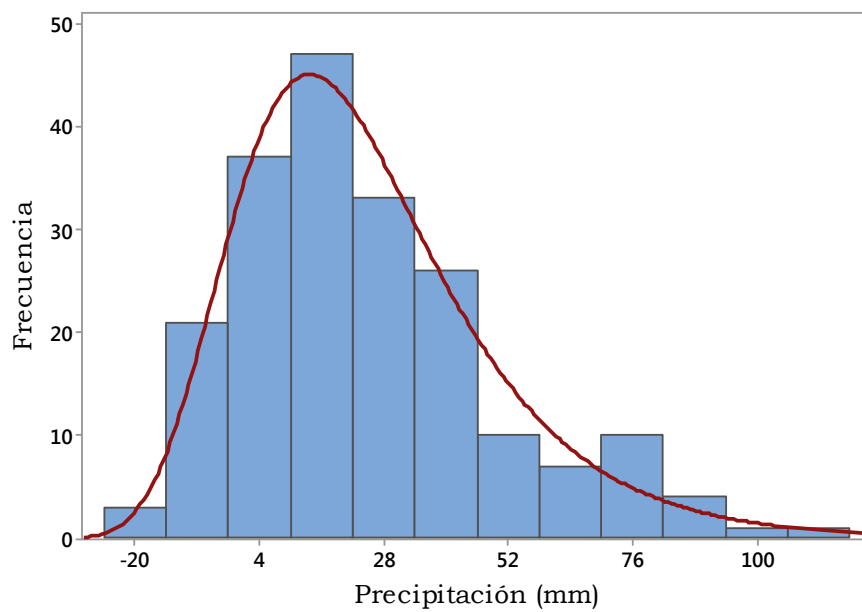
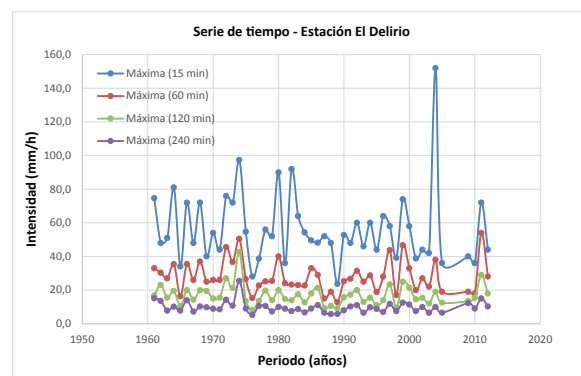
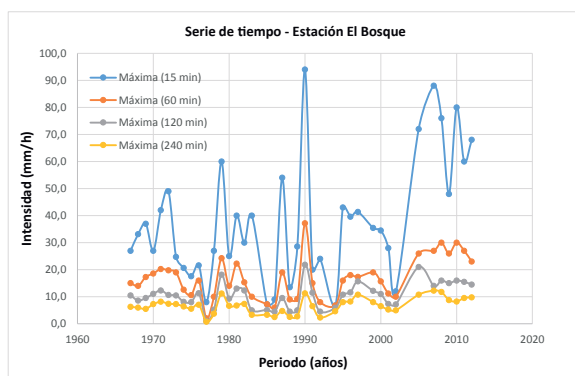


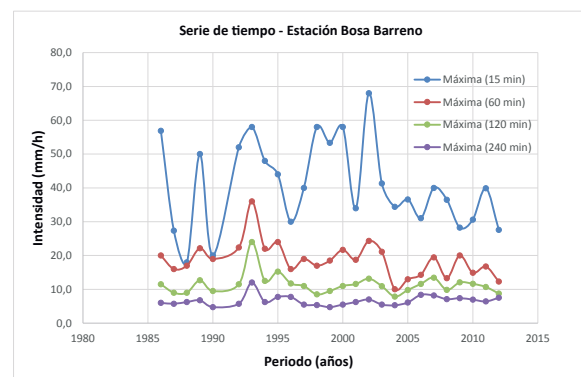
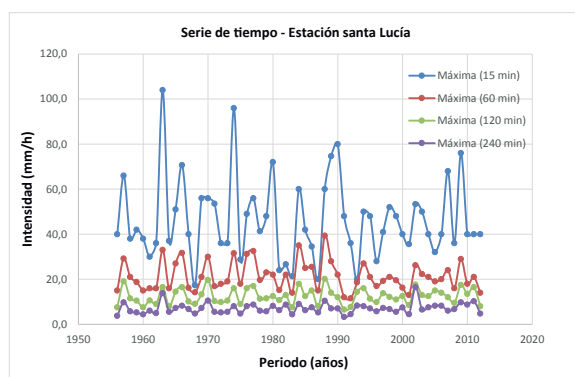
Figura A-7.: Distribución Gumbel.

2. Series de tiempo de las estaciones pluviográficas



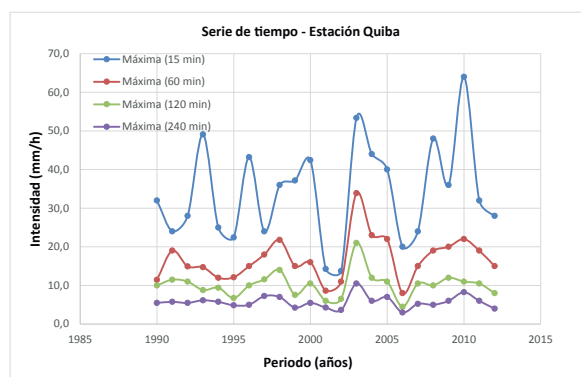
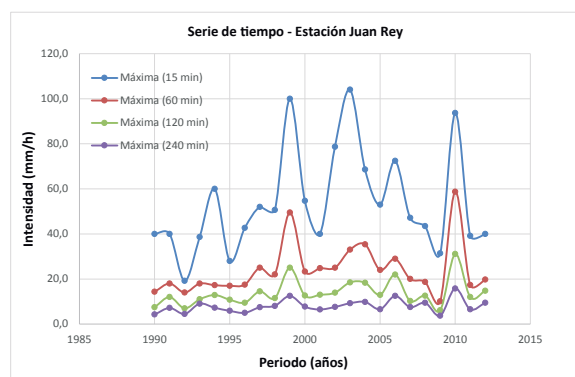
(a) Serie de tiempo para diferentes duraciones. Estación El Bosque.

(b) Serie de tiempo para diferentes duraciones. Estación El Delirio.



(c) Serie de tiempo para diferentes duraciones. Estación Santa Lucía.

(d) Serie de tiempo para diferentes duraciones. Estación Bosa Barreno.



(e) Serie de tiempo para diferentes duraciones. Estación Juan Rey.

(f) Serie de tiempo para diferentes duraciones. Estación Quiba.

Figura A-8.: Series de tiempo de intensidad para diferentes duraciones en cada una de las estaciones analizadas.

Tabla A-1.: Datos nativos de intensidad (mm/h) para la estación El Bosque

| Duración de la tormenta (min) - Estación El Bosque (2120085) | | | | | | | | | | | |
|--|------|------|------|------|------|------|------|------|------|------|-----|
| AÑO | 15 | 30 | 45 | 60 | 90 | 120 | 150 | 180 | 240 | 300 | 360 |
| 1967 | 27,0 | 18,0 | 16,0 | 15,0 | 12,2 | 10,5 | 9,0 | 8,0 | 6,3 | 5,0 | 4,3 |
| 1968 | 33,1 | 18,3 | 15,2 | 14,0 | 10,7 | 8,6 | 7,7 | 7,0 | 6,0 | 5,3 | 4,8 |
| 1969 | 37,0 | 32,0 | 22,2 | 17,3 | 12,4 | 9,5 | 7,7 | 7,0 | 5,5 | 4,8 | 4,3 |
| 1970 | 27,0 | 24,0 | 22,0 | 18,6 | 13,6 | 11,1 | 9,6 | 8,5 | 7,3 | 6,4 | 5,9 |
| 1971 | 42,0 | 36,0 | 25,5 | 20,3 | 15,0 | 12,3 | 10,7 | 9,6 | 8,2 | 7,2 | 6,2 |
| 1972 | 49,0 | 38,0 | 25,8 | 19,8 | 13,7 | 10,6 | 8,8 | 8,5 | 7,4 | 6,2 | 5,3 |
| 1973 | 24,7 | 22,0 | 20,0 | 19,0 | 13,3 | 10,5 | 8,8 | 7,6 | 7,3 | 6,2 | 5,2 |
| 1974 | 20,7 | 17,3 | 15,2 | 12,6 | 9,6 | 8,1 | 7,2 | 6,3 | 6,4 | 5,8 | 5,3 |
| 1975 | 17,6 | 14,0 | 13,3 | 10,6 | 9,2 | 8,0 | 7,3 | 6,8 | 5,6 | 4,6 | 3,8 |
| 1976 | 21,6 | 20,0 | 17,3 | 16,0 | 13,5 | 11,3 | 9,7 | 8,6 | 7,1 | 5,8 | 4,8 |
| 1977 | 8,0 | 4,0 | 2,7 | 2,0 | 1,3 | 1,0 | 1,2 | 1,0 | 0,8 | 0,6 | 0,5 |
| 1978 | 27,0 | 20,0 | 13,3 | 10,0 | 6,7 | 5,2 | 5,3 | 4,7 | 3,8 | 3,0 | 2,5 |
| 1979 | 60,0 | 44,6 | 31,0 | 24,2 | 21,0 | 18,1 | 14,8 | 13,0 | 11,2 | 9,4 | 8,3 |
| 1980 | 25,0 | 17,1 | 15,0 | 14,0 | 10,3 | 9,3 | 8,4 | 7,8 | 6,6 | 5,4 | 5,0 |
| 1981 | 40,0 | 38,0 | 27,5 | 22,2 | 17,0 | 13,0 | 10,4 | 8,7 | 6,8 | 5,4 | 4,5 |
| 1982 | 30,0 | 22,7 | 17,8 | 15,3 | 13,8 | 12,3 | 11,2 | 9,5 | 7,3 | 6,4 | 5,7 |
| 1983 | 40,0 | 20,0 | 13,3 | 10,0 | 6,7 | 5,0 | 4,2 | 3,8 | 3,3 | 3,1 | 2,9 |
| 1985 | 7,2 | 7,2 | 7,2 | 7,2 | 6,3 | 5,1 | 4,4 | 3,9 | 3,3 | 3,0 | 2,5 |
| 1986 | 9,0 | 7,4 | 6,4 | 6,0 | 5,3 | 4,5 | 3,6 | 3,0 | 2,5 | 2,0 | 1,7 |
| 1987 | 54,0 | 38,0 | 25,3 | 19,0 | 12,7 | 9,5 | 7,6 | 6,3 | 4,8 | 3,8 | 3,2 |
| 1988 | 13,5 | 13,5 | 12,0 | 9,0 | 6,0 | 4,5 | 4,0 | 3,3 | 2,5 | 2,1 | 1,8 |
| 1989 | 28,6 | 14,7 | 10,1 | 9,2 | 6,7 | 5,0 | 4,0 | 3,7 | 2,8 | 2,7 | 2,8 |
| 1990 | 94,0 | 62,0 | 46,3 | 37,2 | 27,1 | 21,8 | 18,0 | 15,0 | 11,3 | 9,0 | 7,5 |
| 1991 | 20,0 | 20,0 | 16,7 | 15,0 | 13,3 | 11,5 | 9,8 | 8,7 | 6,5 | 5,8 | 4,8 |
| 1992 | 24,0 | 13,3 | 9,8 | 8,0 | 5,7 | 4,5 | 3,6 | 3,0 | 2,3 | 2,1 | 1,9 |
| 1994 | 6,8 | 6,8 | 6,8 | 6,8 | 6,5 | 5,9 | 5,6 | 5,3 | 4,6 | 4,3 | 3,7 |
| 1995 | 43,0 | 26,0 | 19,3 | 16,0 | 12,7 | 10,8 | 9,6 | 9,0 | 8,0 | 7,4 | 6,3 |
| 1996 | 39,6 | 25,2 | 20,4 | 18,0 | 13,8 | 11,6 | 10,9 | 10,0 | 8,3 | 8,1 | 8,0 |
| 1997 | 41,3 | 24,7 | 19,8 | 17,3 | 16,0 | 15,7 | 16,2 | 14,3 | 10,8 | 8,6 | 7,2 |
| 1999 | 35,4 | 26,3 | 22,0 | 19,0 | 14,9 | 12,2 | 10,5 | 9,3 | 8,0 | 6,8 | 6,0 |
| 2000 | 34,5 | 21,3 | 17,6 | 15,7 | 13,2 | 11,0 | 9,4 | 8,3 | 6,5 | 5,2 | 4,3 |
| 2001 | 28,0 | 16,0 | 11,3 | 11,2 | 8,7 | 7,3 | 7,5 | 6,9 | 5,3 | 4,2 | 3,5 |
| 2002 | 12,0 | 12,0 | 11,3 | 10,0 | 8,2 | 7,2 | 6,7 | 6,3 | 5,0 | 4,4 | 3,7 |
| 2005 | 72,0 | 44,0 | 30,7 | 26,0 | 24,0 | 21,0 | 17,2 | 14,3 | 10,8 | 8,6 | 7,8 |
| 2007 | 88,0 | 54,0 | 36,0 | 27,0 | 18,7 | 14,0 | 12,8 | 15,0 | 12,3 | 11,2 | 9,3 |
| 2008 | 76,0 | 48,0 | 38,7 | 30,0 | 21,3 | 16,0 | 14,4 | 14,0 | 11,8 | 9,8 | 9,0 |
| 2009 | 48,0 | 34,0 | 30,7 | 26,0 | 18,7 | 15,0 | 12,0 | 10,3 | 8,8 | 7,6 | 6,7 |
| 2010 | 80,0 | 48,0 | 38,7 | 30,0 | 21,3 | 16,0 | 12,8 | 10,7 | 8,3 | 6,6 | 6,2 |
| 2011 | 60,0 | 46,0 | 32,0 | 27,0 | 19,3 | 15,5 | 13,2 | 11,0 | 9,5 | 8,8 | 7,5 |
| 2012 | 68,0 | 40,0 | 28,0 | 23,0 | 17,3 | 14,5 | 14,0 | 13,0 | 9,8 | 7,8 | 7,3 |

Tabla A-2.: Datos nativos de intensidad (mm/h) para la estación El Delirio

| Duración de la tormenta (min) - Estación El Delirio (2120013) | | | | | | | | | | | |
|---|-------|------|------|------|------|------|------|------|------|------|------|
| AÑO | 15 | 30 | 45 | 60 | 90 | 120 | 150 | 180 | 240 | 300 | 360 |
| 1961 | 74,7 | 56,0 | 43,3 | 33,0 | 22,0 | 16,8 | 15,1 | 16,1 | 15,1 | 13,0 | 11,3 |
| 1962 | 48,0 | 38,0 | 34,7 | 30,3 | 26,7 | 23,1 | 20,0 | 17,4 | 13,6 | 11,8 | 9,8 |
| 1963 | 51,0 | 44,0 | 34,3 | 27,0 | 19,3 | 15,5 | 12,4 | 10,3 | 7,8 | 7,4 | 6,2 |
| 1964 | 81,0 | 61,6 | 44,3 | 35,3 | 24,4 | 19,5 | 16,0 | 13,3 | 10,2 | 8,6 | 7,2 |
| 1965 | 34,0 | 22,8 | 21,3 | 16,4 | 12,3 | 10,3 | 9,0 | 9,3 | 7,8 | 6,2 | 5,2 |
| 1966 | 72,0 | 50,0 | 41,3 | 35,5 | 24,7 | 20,0 | 19,2 | 17,0 | 14,0 | 13,1 | 12,0 |
| 1967 | 48,0 | 38,0 | 32,0 | 26,0 | 17,3 | 14,3 | 11,6 | 9,7 | 7,3 | 5,8 | 5,2 |
| 1968 | 72,0 | 56,0 | 46,3 | 37,0 | 26,0 | 19,8 | 16,4 | 13,7 | 10,3 | 8,2 | 6,8 |
| 1969 | 40,0 | 32,0 | 26,0 | 25,0 | 25,0 | 19,5 | 15,6 | 13,0 | 9,8 | 7,8 | 6,5 |
| 1970 | 54,0 | 41,6 | 34,7 | 26,0 | 19,8 | 15,0 | 12,0 | 10,7 | 8,9 | 7,9 | 6,8 |
| 1971 | 44,0 | 30,0 | 24,3 | 26,0 | 18,7 | 15,5 | 12,7 | 10,7 | 8,6 | 7,7 | 6,8 |
| 1972 | 76,0 | 59,5 | 50,5 | 45,5 | 36,0 | 27,0 | 22,0 | 18,8 | 14,3 | 11,4 | 9,5 |
| 1973 | 72,0 | 46,0 | 39,0 | 36,7 | 27,7 | 21,3 | 17,2 | 14,3 | 10,8 | 8,6 | 7,2 |
| 1974 | 97,3 | 67,3 | 57,3 | 50,5 | 40,8 | 42,7 | 40,4 | 34,0 | 25,5 | 20,4 | 17,0 |
| 1975 | 54,7 | 38,7 | 30,2 | 26,5 | 18,0 | 13,5 | 12,4 | 11,4 | 9,0 | 7,2 | 6,0 |
| 1976 | 28,0 | 22,0 | 17,3 | 15,3 | 10,7 | 8,2 | 7,4 | 6,5 | 5,3 | 4,2 | 3,5 |
| 1977 | 38,7 | 32,7 | 25,1 | 22,7 | 16,3 | 13,5 | 14,3 | 12,8 | 10,5 | 8,4 | 7,0 |
| 1978 | 56,0 | 40,0 | 30,1 | 25,2 | 22,7 | 19,6 | 16,3 | 14,0 | 10,5 | 8,4 | 7,0 |
| 1979 | 52,0 | 39,4 | 32,4 | 25,5 | 18,0 | 14,0 | 11,3 | 9,5 | 7,3 | 6,1 | 5,5 |
| 1980 | 90,0 | 70,0 | 52,0 | 40,0 | 26,7 | 20,0 | 16,0 | 13,3 | 10,0 | 8,0 | 6,7 |
| 1981 | 36,0 | 28,5 | 26,0 | 24,0 | 18,7 | 14,7 | 12,3 | 10,8 | 8,9 | 7,6 | 6,3 |
| 1982 | 92,0 | 46,0 | 30,7 | 23,3 | 17,7 | 14,0 | 11,6 | 9,7 | 7,5 | 6,4 | 5,7 |
| 1983 | 64,0 | 35,5 | 26,7 | 23,0 | 19,3 | 17,5 | 14,0 | 11,7 | 8,8 | 7,0 | 5,8 |
| 1984 | 54,3 | 34,0 | 29,6 | 22,7 | 15,7 | 12,8 | 10,4 | 8,7 | 6,8 | 5,4 | 4,5 |
| 1985 | 49,5 | 49,5 | 44,0 | 33,0 | 24,0 | 18,0 | 14,4 | 12,0 | 9,0 | 7,2 | 6,0 |
| 1986 | 48,2 | 39,1 | 32,4 | 29,0 | 26,6 | 21,3 | 17,5 | 14,7 | 11,0 | 8,8 | 7,3 |
| 1987 | 52,0 | 26,0 | 17,9 | 15,0 | 11,0 | 9,4 | 9,2 | 7,9 | 6,5 | 5,4 | 4,5 |
| 1988 | 48,0 | 34,0 | 24,7 | 19,0 | 12,7 | 10,6 | 8,9 | 7,6 | 5,8 | 4,6 | 3,8 |
| 1989 | 23,6 | 17,1 | 14,7 | 12,8 | 10,5 | 9,2 | 8,4 | 7,7 | 5,8 | 4,7 | 4,0 |
| 1990 | 52,8 | 42,0 | 31,4 | 25,3 | 20,0 | 15,7 | 12,8 | 10,7 | 8,0 | 6,4 | 5,3 |
| 1991 | 48,0 | 37,7 | 30,9 | 26,7 | 20,4 | 17,3 | 14,9 | 12,9 | 10,3 | 8,2 | 6,8 |
| 1992 | 60,0 | 46,0 | 36,7 | 31,5 | 26,0 | 20,0 | 16,9 | 14,4 | 11,0 | 8,8 | 7,3 |
| 1993 | 46,0 | 38,0 | 29,3 | 25,0 | 17,3 | 13,0 | 10,4 | 8,7 | 6,5 | 5,2 | 4,3 |
| 1994 | 60,0 | 54,0 | 37,1 | 28,7 | 20,3 | 15,5 | 12,4 | 10,3 | 9,8 | 8,0 | 7,0 |
| 1995 | 44,0 | 28,0 | 21,9 | 18,8 | 14,7 | 11,0 | 8,8 | 7,3 | 8,8 | 7,4 | 6,2 |
| 1996 | 64,0 | 46,7 | 35,8 | 28,0 | 18,7 | 14,0 | 11,2 | 9,3 | 7,0 | 5,6 | 4,7 |
| 1997 | 58,0 | 56,0 | 53,3 | 43,8 | 30,7 | 23,4 | 18,8 | 15,7 | 11,8 | 9,4 | 7,8 |
| 1998 | 39,2 | 28,0 | 22,7 | 17,1 | 11,7 | 9,5 | 8,7 | 8,2 | 7,5 | 6,0 | 5,0 |
| 1999 | 74,0 | 66,0 | 57,6 | 46,7 | 33,0 | 25,0 | 20,0 | 16,7 | 12,5 | 10,0 | 8,3 |
| 2000 | 58,0 | 48,0 | 40,0 | 33,0 | 26,0 | 21,5 | 17,4 | 14,7 | 11,4 | 9,2 | 7,8 |
| 2001 | 38,8 | 26,0 | 22,3 | 20,0 | 18,2 | 14,5 | 11,6 | 9,7 | 7,5 | 6,4 | 6,0 |
| 2002 | 44,0 | 34,0 | 29,3 | 27,0 | 19,3 | 15,5 | 13,4 | 12,7 | 10,0 | 8,2 | 7,3 |
| 2003 | 42,0 | 32,8 | 24,3 | 22,0 | 15,3 | 12,0 | 10,0 | 8,7 | 6,5 | 7,0 | 6,6 |
| 2004 | 152,0 | 76,0 | 50,7 | 38,0 | 25,3 | 19,0 | 15,2 | 13,0 | 10,0 | 8,0 | 6,7 |
| 2005 | 36,0 | 20,0 | 20,0 | 19,0 | 16,0 | 12,5 | 10,4 | 8,7 | 6,5 | 5,4 | 5,0 |
| 2009 | 40,0 | 20,0 | 18,7 | 19,0 | 14,7 | 13,5 | 12,0 | 13,3 | 12,3 | 10,8 | 9,3 |
| 2010 | 36,0 | 24,0 | 17,3 | 18,0 | 14,7 | 15,5 | 13,6 | 11,7 | 9,0 | 7,2 | 6,0 |
| 2011 | 72,0 | 64,0 | 61,3 | 54,0 | 38,0 | 29,0 | 23,2 | 19,7 | 15,0 | 12,0 | 10,0 |
| 2012 | 44,0 | 42,0 | 34,7 | 28,0 | 19,3 | 18,0 | 15,6 | 13,3 | 10,3 | 8,2 | 6,8 |

Tabla A-3.: Datos nativos de intensidad (mm/h) para la estación Santa Lucía

| Duración de la tormenta (min) - Estación Santa Lucía (2120052) | | | | | | | | | | | |
|--|-------|------|------|------|------|------|------|------|------|------|------|
| AÑO | 15 | 30 | 45 | 60 | 90 | 120 | 150 | 180 | 240 | 300 | 360 |
| 1956 | 40,0 | 24,7 | 18,0 | 15,0 | 10,0 | 7,5 | 6,0 | 5,0 | 3,8 | 3,0 | 2,5 |
| 1957 | 66,0 | 43,3 | 34,0 | 29,2 | 23,7 | 19,0 | 15,4 | 13,0 | 9,8 | 7,8 | 6,5 |
| 1958 | 38,0 | 32,0 | 25,0 | 21,0 | 14,7 | 11,5 | 9,2 | 7,7 | 5,8 | 4,6 | 3,8 |
| 1959 | 42,0 | 36,0 | 24,0 | 18,8 | 14,0 | 10,5 | 8,4 | 7,0 | 5,3 | 4,2 | 3,5 |
| 1960 | 38,0 | 30,0 | 20,0 | 15,0 | 10,0 | 7,5 | 6,0 | 5,1 | 4,4 | 3,6 | 3,0 |
| 1961 | 30,0 | 23,3 | 18,9 | 16,0 | 12,7 | 10,6 | 9,0 | 8,0 | 6,0 | 4,8 | 4,0 |
| 1962 | 36,0 | 26,0 | 19,3 | 16,0 | 12,0 | 9,0 | 7,2 | 6,3 | 5,0 | 4,0 | 3,3 |
| 1963 | 104,0 | 64,0 | 44,0 | 33,0 | 22,0 | 16,5 | 13,2 | 11,0 | 13,8 | 11,2 | 9,3 |
| 1964 | 37,0 | 32,0 | 21,3 | 16,0 | 10,7 | 8,0 | 7,6 | 7,3 | 5,5 | 4,4 | 3,7 |
| 1965 | 51,0 | 46,0 | 34,0 | 27,0 | 19,3 | 14,5 | 11,6 | 9,7 | 7,3 | 5,8 | 4,8 |
| 1966 | 70,7 | 51,0 | 40,0 | 31,7 | 22,0 | 16,5 | 13,2 | 11,0 | 8,3 | 6,6 | 5,5 |
| 1967 | 40,0 | 27,6 | 20,0 | 16,2 | 12,2 | 10,2 | 9,5 | 8,6 | 6,8 | 5,6 | 5,0 |
| 1968 | 17,3 | 16,7 | 16,0 | 14,2 | 10,5 | 9,1 | 7,6 | 6,3 | 4,8 | 3,8 | 3,2 |
| 1969 | 56,0 | 32,7 | 24,9 | 21,0 | 15,8 | 13,3 | 11,5 | 9,7 | 7,3 | 6,0 | 5,0 |
| 1970 | 56,0 | 46,7 | 32,0 | 30,0 | 21,3 | 19,7 | 16,5 | 14,0 | 10,5 | 8,4 | 7,0 |
| 1971 | 53,6 | 29,1 | 21,0 | 16,9 | 12,8 | 10,3 | 8,5 | 7,3 | 5,5 | 4,6 | 4,4 |
| 1972 | 36,0 | 32,0 | 22,7 | 18,0 | 12,6 | 9,9 | 8,1 | 6,9 | 5,3 | 4,4 | 4,2 |
| 1973 | 36,0 | 25,0 | 23,2 | 19,0 | 13,6 | 10,5 | 8,5 | 7,2 | 5,5 | 4,4 | 3,7 |
| 1974 | 96,0 | 56,4 | 41,5 | 31,5 | 21,3 | 16,0 | 12,8 | 10,7 | 8,0 | 6,4 | 5,3 |
| 1975 | 28,8 | 28,0 | 24,0 | 18,0 | 12,0 | 9,0 | 7,6 | 6,3 | 4,8 | 4,4 | 4,3 |
| 1976 | 49,0 | 44,0 | 38,3 | 31,3 | 21,3 | 16,0 | 12,8 | 10,7 | 8,0 | 6,4 | 5,3 |
| 1977 | 56,0 | 40,0 | 42,0 | 32,5 | 22,0 | 17,0 | 14,0 | 11,7 | 8,8 | 7,0 | 5,8 |
| 1978 | 41,3 | 31,6 | 24,3 | 19,7 | 13,9 | 11,3 | 9,3 | 7,9 | 6,0 | 4,8 | 4,0 |
| 1979 | 48,0 | 42,0 | 29,3 | 23,0 | 15,3 | 11,5 | 9,2 | 7,7 | 5,8 | 5,0 | 4,6 |
| 1980 | 72,0 | 41,0 | 29,3 | 22,0 | 15,7 | 12,5 | 10,8 | 10,0 | 8,2 | 6,9 | 6,3 |
| 1981 | 24,0 | 16,0 | 15,2 | 15,2 | 13,3 | 10,7 | 8,8 | 7,6 | 6,3 | 5,2 | 4,3 |
| 1982 | 26,7 | 26,7 | 26,7 | 22,0 | 16,0 | 13,0 | 12,0 | 11,0 | 8,8 | 7,0 | 5,8 |
| 1983 | 21,3 | 18,0 | 14,7 | 14,0 | 10,0 | 7,5 | 6,0 | 5,4 | 4,4 | 4,0 | 3,8 |
| 1984 | 60,0 | 60,0 | 44,7 | 35,0 | 24,0 | 18,0 | 14,4 | 12,0 | 9,0 | 7,2 | 6,0 |
| 1985 | 42,0 | 38,0 | 32,3 | 25,0 | 16,7 | 12,5 | 10,0 | 8,3 | 6,3 | 5,3 | 4,5 |
| 1986 | 34,5 | 34,5 | 31,5 | 25,5 | 19,3 | 15,0 | 12,0 | 10,0 | 7,5 | 6,0 | 5,0 |
| 1987 | 20,0 | 20,0 | 20,0 | 15,0 | 10,0 | 8,0 | 6,8 | 6,3 | 5,3 | 4,2 | 3,5 |
| 1988 | 60,0 | 52,0 | 52,0 | 39,3 | 26,7 | 20,2 | 16,3 | 13,7 | 10,4 | 8,4 | 7,0 |
| 1989 | 74,7 | 56,0 | 37,3 | 28,0 | 18,7 | 14,0 | 11,2 | 9,3 | 7,0 | 5,6 | 4,7 |
| 1990 | 80,0 | 40,0 | 27,0 | 22,0 | 15,3 | 12,0 | 10,0 | 8,7 | 7,0 | 6,3 | 5,7 |
| 1991 | 48,0 | 24,0 | 16,0 | 12,0 | 8,0 | 6,5 | 5,2 | 4,3 | 3,3 | 2,6 | 2,2 |
| 1992 | 36,0 | 20,8 | 14,6 | 11,5 | 8,7 | 7,6 | 6,6 | 5,8 | 4,5 | 3,6 | 3,0 |
| 1993 | 20,0 | 20,0 | 19,2 | 18,6 | 18,0 | 14,4 | 12,0 | 10,4 | 8,3 | 6,9 | 5,9 |
| 1994 | 50,0 | 46,0 | 34,7 | 27,0 | 21,0 | 16,0 | 12,8 | 10,7 | 8,0 | 6,4 | 5,3 |
| 1995 | 48,0 | 30,8 | 25,0 | 21,0 | 14,0 | 11,4 | 9,9 | 9,0 | 7,0 | 5,6 | 4,7 |
| 1996 | 28,0 | 24,0 | 20,7 | 17,0 | 12,4 | 9,8 | 8,4 | 7,5 | 5,8 | 4,6 | 3,8 |
| 1997 | 41,0 | 28,0 | 22,0 | 19,3 | 16,3 | 13,8 | 11,2 | 9,3 | 7,3 | 5,8 | 4,8 |
| 1998 | 52,0 | 32,0 | 25,3 | 21,0 | 14,7 | 12,2 | 10,4 | 9,0 | 6,8 | 5,4 | 4,9 |
| 1999 | 48,0 | 29,0 | 25,5 | 19,5 | 13,3 | 11,0 | 8,8 | 7,3 | 5,5 | 4,4 | 3,7 |
| 2000 | 40,0 | 32,0 | 21,3 | 16,3 | 14,0 | 12,5 | 10,5 | 9,1 | 7,4 | 6,3 | 5,8 |
| 2001 | 35,6 | 20,4 | 17,3 | 13,0 | 10,1 | 8,5 | 6,9 | 6,0 | 4,5 | 3,6 | 3,0 |
| 2002 | 53,3 | 41,6 | 30,8 | 26,2 | 21,2 | 17,6 | 18,6 | 19,3 | 16,5 | 13,7 | 11,5 |
| 2003 | 50,0 | 40,0 | 29,4 | 22,3 | 15,3 | 13,0 | 10,4 | 8,7 | 6,5 | 5,2 | 4,3 |
| 2004 | 40,0 | 38,0 | 26,7 | 21,0 | 15,3 | 12,5 | 10,8 | 9,7 | 7,5 | 6,2 | 5,3 |
| 2005 | 32,0 | 20,0 | 16,0 | 19,0 | 15,3 | 15,0 | 12,8 | 11,0 | 8,3 | 7,0 | 6,2 |
| 2006 | 40,0 | 30,0 | 24,0 | 20,0 | 18,0 | 14,0 | 12,0 | 10,0 | 8,3 | 7,0 | 5,8 |
| 2007 | 68,0 | 46,0 | 32,0 | 24,0 | 16,0 | 12,0 | 9,6 | 8,0 | 6,0 | 4,8 | 4,0 |
| 2008 | 36,0 | 22,0 | 14,7 | 16,0 | 11,3 | 9,5 | 7,6 | 6,7 | 6,8 | 6,8 | 6,2 |
| 2009 | 76,0 | 46,0 | 36,0 | 29,0 | 21,3 | 17,5 | 14,4 | 12,3 | 9,8 | 8,2 | 7,0 |
| 2010 | 40,0 | 24,0 | 18,7 | 18,0 | 13,3 | 13,5 | 11,2 | 9,7 | 8,8 | 8,0 | 6,8 |
| 2011 | 40,0 | 26,0 | 24,0 | 21,0 | 18,7 | 16,5 | 15,6 | 13,3 | 10,3 | 8,4 | 7,0 |
| 2012 | 40,0 | 20,0 | 17,3 | 14,0 | 10,7 | 8,0 | 7,2 | 6,3 | 4,8 | 4,2 | 3,7 |

Tabla A-4.: Datos nativos de intensidad (mm/h) para la estación Bosa Barreno

| Duración de la tormenta (min) - Estación Bosa Barreno (2120154) | | | | | | | | | | | |
|--|-----------|-----------|-----------|-----------|-----------|------------|------------|------------|------------|------------|------------|
| AÑO | 15 | 30 | 45 | 60 | 90 | 120 | 150 | 180 | 240 | 300 | 360 |
| 1986 | 56,9 | 30,0 | 23,3 | 20,0 | 13,3 | 11,5 | 9,2 | 7,7 | 6,0 | 5,6 | 5,3 |
| 1987 | 27,3 | 24,7 | 21,3 | 16,0 | 11,3 | 9,0 | 7,6 | 6,8 | 5,8 | 5,1 | 4,5 |
| 1988 | 18,0 | 18,0 | 17,3 | 17,0 | 11,3 | 9,0 | 7,9 | 7,2 | 6,3 | 5,4 | 4,7 |
| 1989 | 50,0 | 40,0 | 28,1 | 22,1 | 16,2 | 12,7 | 10,3 | 8,8 | 6,8 | 5,4 | 4,5 |
| 1990 | 20,0 | 20,0 | 19,3 | 19,0 | 12,7 | 9,5 | 7,6 | 6,3 | 4,8 | 3,8 | 3,2 |
| 1992 | 52,0 | 35,0 | 29,3 | 22,4 | 15,3 | 11,5 | 9,2 | 7,7 | 5,8 | 4,6 | 3,8 |
| 1993 | 58,0 | 52,0 | 41,3 | 36,0 | 30,7 | 24,0 | 19,2 | 16,0 | 12,0 | 9,6 | 8,0 |
| 1994 | 48,0 | 32,0 | 25,3 | 22,0 | 16,4 | 12,5 | 10,0 | 8,3 | 6,3 | 5,0 | 4,2 |
| 1995 | 44,0 | 40,0 | 29,3 | 24,0 | 18,3 | 15,3 | 12,4 | 10,3 | 7,8 | 6,2 | 5,2 |
| 1996 | 30,0 | 23,0 | 18,5 | 16,0 | 13,4 | 11,7 | 10,7 | 9,8 | 7,8 | 6,2 | 5,2 |
| 1997 | 40,0 | 26,0 | 21,3 | 19,0 | 14,5 | 11,0 | 8,8 | 7,3 | 5,5 | 4,4 | 3,7 |
| 1998 | 58,0 | 32,0 | 22,7 | 17,0 | 11,3 | 8,5 | 6,8 | 6,2 | 5,3 | 4,9 | 4,2 |
| 1999 | 53,3 | 32,7 | 24,3 | 18,5 | 12,7 | 9,5 | 7,6 | 6,3 | 4,8 | 3,8 | 3,2 |
| 2000 | 58,0 | 40,0 | 28,2 | 21,7 | 14,7 | 11,0 | 8,8 | 7,3 | 5,5 | 4,4 | 3,7 |
| 2001 | 34,0 | 27,6 | 21,7 | 18,7 | 14,9 | 11,6 | 9,6 | 8,3 | 6,3 | 5,0 | 4,2 |
| 2002 | 68,0 | 40,0 | 28,7 | 24,3 | 16,9 | 13,2 | 10,9 | 9,3 | 7,0 | 5,6 | 4,7 |
| 2003 | 41,3 | 28,7 | 25,3 | 21,1 | 14,3 | 10,9 | 8,8 | 7,3 | 5,5 | 4,4 | 3,7 |
| 2004 | 34,4 | 20,0 | 13,3 | 10,0 | 9,7 | 7,9 | 7,5 | 6,4 | 5,3 | 5,3 | 4,5 |
| 2005 | 36,6 | 20,0 | 14,1 | 13,0 | 11,8 | 9,8 | 9,7 | 8,2 | 6,1 | 4,9 | 4,1 |
| 2006 | 31,1 | 20,0 | 15,9 | 14,3 | 13,4 | 11,6 | 10,4 | 10,4 | 8,4 | 6,7 | 5,6 |
| 2007 | 40,0 | 26,1 | 23,3 | 19,5 | 15,2 | 13,4 | 12,0 | 10,4 | 8,2 | 6,8 | 5,7 |
| 2008 | 36,5 | 20,9 | 15,5 | 13,3 | 11,3 | 9,8 | 8,7 | 8,5 | 7,1 | 6,0 | 5,0 |
| 2009 | 28,3 | 20,0 | 19,2 | 20,0 | 14,4 | 12,1 | 10,1 | 8,6 | 7,4 | 7,3 | 6,7 |
| 2010 | 30,6 | 20,0 | 15,4 | 14,9 | 12,8 | 11,6 | 9,5 | 7,9 | 7,0 | 6,0 | 5,0 |
| 2011 | 39,9 | 20,0 | 14,5 | 16,7 | 13,3 | 10,7 | 9,0 | 7,8 | 6,4 | 5,5 | 4,7 |
| 2012 | 27,6 | 20,0 | 13,3 | 12,3 | 11,0 | 8,8 | 8,0 | 7,5 | 7,5 | 6,8 | 5,9 |

3. Gráfica de los datos nativos para la estación El Bosque

Nube de puntos

September 12, 2016

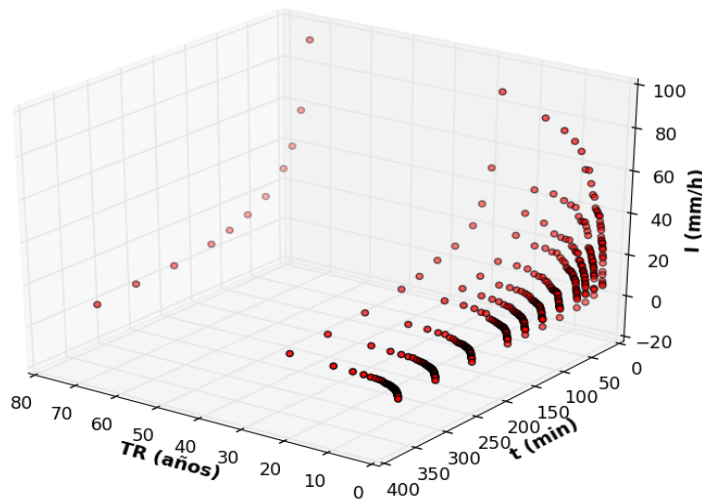
```
In [ ]: from mpl_toolkits.mplot3d import Axes3D
import matplotlib.pyplot as plt

fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')

XT = [loadtxt("Periodo_Returno.dat")] # Periodo de Retorno (años)
yD = [loadtxt("Datos_Duración.dat")] # Duración (minutos)
zI = [loadtxt("Datos_Intensidad.dat")] # Intensidad (mm/h)
ax.scatter(xT, yD, zI, c='r', marker='o')

ax.set_xlabel('TR (años)', fontsize = 13, fontweight='bold')
ax.set_ylabel('t (min)', fontsize = 13, fontweight='bold')
ax.set_zlabel('I (mm/h)', fontsize = 13, fontweight='bold')

plt.show()
```



4. Superficie de los datos nativos para la estación El Bosque

Superficie IDF

September 12, 2016

```
In [ ]: from mpl_toolkits.mplot3d import Axes3D
        from scipy.misc import imread
        import numpy as np
        import matplotlib.pyplot as plt

fig = plt.figure()
ax = Axes3D(fig)
X = np.arange(0, 120, 5)
Y = np.arange(0, 400, 5)
X, Y = np.meshgrid(X, Y)
c = 285.369
e = 0.755
f = 3.024
m = 0.339

Z = (c*(X**m))/((Y**e) + f)

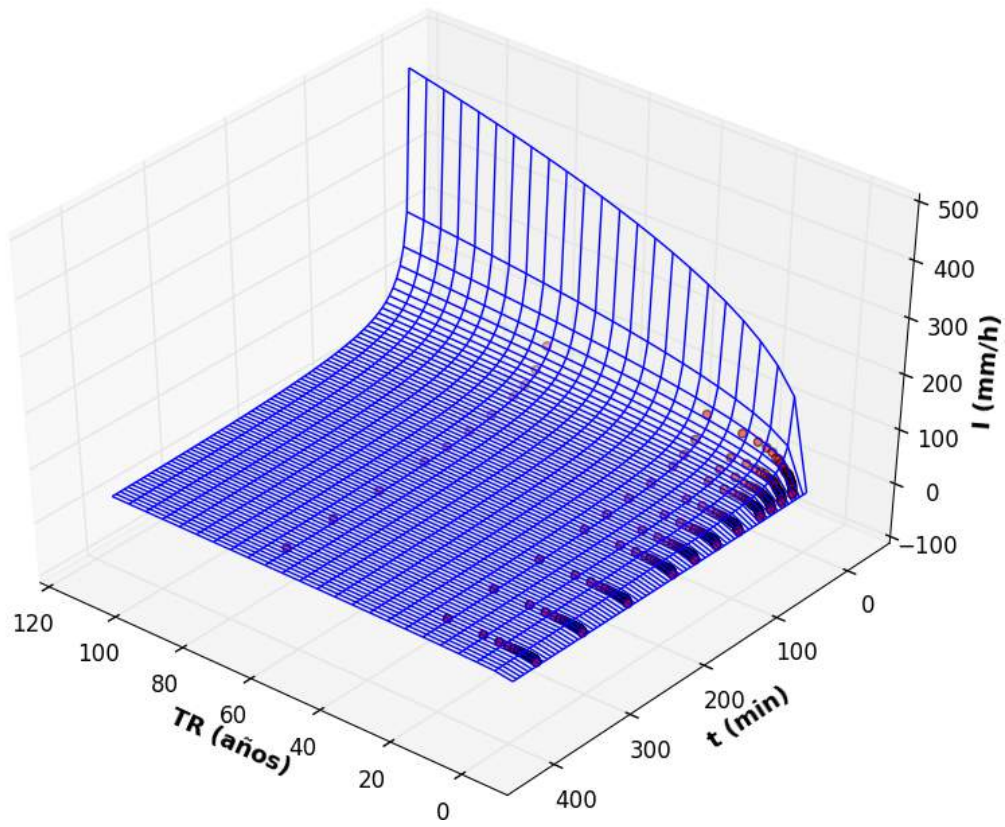
ax.plot_wireframe(X, Y, Z, rstride=1, cstride=1)

T = [loadtxt("Periodo_Returno.dat")] # Periodo de Retorno (años)
D = [loadtxt("Datos_Duración.dat")] # Duración (minutos)
I = [loadtxt("Datos_Intensidad.dat")] # Intensidad (mm/h)
#surf = ax.plot_wireframe(X, Y, Z, rstride=1, cstride=1)

ax.scatter(T, D, I, c='r', marker='o') #color rojo

ax.set_xlabel('TR (años)', fontsize = 13, fontweight='bold')
ax.set_ylabel('t (min)', fontsize = 13, fontweight='bold')
ax.set_zlabel('I (mm/h)', fontsize = 13, fontweight='bold')

plt.show()
```



B. Anexo: Capítulo 3. Cópulas Bivariadas

1. Coeficiente de correlación τ de Kendall para la estación el Bosque

Tau Kendall (Esta. El Bosque)

September 12, 2016

```
In [1]: %pylab inline
import scipy.misc as sp
import scipy.stats

In [2]: X = [loadtxt("Datos_Intensidad.dat")] # Intensidad (mm/h)
Y = [loadtxt("Datos_Duración.dat")] # Duración (minutos)

N = len(X) # filas, numero de datos en total
M = 2 # columnas, X e Y
Signo = np.zeros([N,N])

for i in range(N):
    for j in range(i+1, N):
        aux = ( X[i]-X[j] )*( Y[i]-Y[j])
        Signo[i,j] = np.sign(aux) # Función Signo

Tau_a = sum(Signo)*(sp.comb(N,M))**-1 # Definicion Tau_a
print "Tau_a = ", Tau_a

Tau_a = -0.556429902671
```


2. Método gráfico para la selección de la cópula de mejor ajuste

Graphic Method

September 21, 2016

```
In [ ]: import scipy.misc as sp
import scipy.stats
from sympy import init_session
init_session (use_latex = True)

In [ ]: X = np.array([loadtxt("Datos_Intensidad.dat")])
Y = np.array([loadtxt("Datos_Intensidad.dat")])

# X = Intensidad (mm/h) El Delirio
# Y = Intensidad (mm/h) santa Lucía

N = len(X)      # filas, numero de datos en total
M = len(Y)

Z = np.zeros([N])
K_z = np.zeros([N])

for i in range(N):
    lista = list(range(N))
    lista.remove(i)

    Z[i]=len(np.where((X[i] > X[lista]) & (Y[i] > Y[lista]))[0])/(N-1.)

for i in range(M):
    lista_z = list(range(M))

    K_z[i] = len(np.where(Z[i] >= Z[lista_z])[0])*1./len(Z)

In [ ]: ### Cópula Gumbel ###

L = len(K_z)
theta = 2.56
K_t_Gumbel = np.zeros([L])
```

```

Phi = (- log (t))**theta
Phi_prima = Phi.diff(t)

for i in range(L):
    G = t - (Phi / Phi_prima )
    H = G.evalf(subs={t:K_z[i]})
    K_t_Gumbel[i] = np.sum([H])
#K_t_Gumbel

```

```
In [ ]: ### Cópula Frank ###
```

```

Fr = len(K_z)
theta_Fr = 8.18
K_t_Frank = np.zeros([Fr])
t_Fr = Symbol('t_Fr')

Phi_Fr = -log((exp(-theta_Fr*t_Fr)-1)/(exp(-theta_Fr)-1))
Phi_prima_Fr = Phi_Fr.diff(t_Fr)

for i in range(Fr):
    G_Fr = t_Fr - (Phi_Fr / Phi_prima_Fr )
    H_Fr = G_Fr.evalf(subs={t_Fr:K_z[i]})
    K_t_Frank[i] = np.sum([H_Fr])
#K_t_Frank

```

```
In [ ]: ### Cópula Clayton ###
```

```

Cl = len(K_z)
theta_Cl = 3.116
K_t_Clayton = np.zeros([Cl])
t_Cl = Symbol('t_Cl')

Phi_Cl = (1/theta_Cl)*(t_Cl**(-theta_Cl)-1)
Phi_prima_Cl = Phi_Cl.diff(t_Cl)

for i in range(Cl):
    G_Cl = t_Cl - (Phi_Cl / Phi_prima_Cl )
    H_Cl = G_Cl.evalf(subs={t_Cl:K_z[i]})
    K_t_Clayton[i] = np.sum([H_Cl])
#K_t_Clayton

```

```
In [ ]: ### Cópula Joe ###
```

```

Joe = len(K_z)
theta_Joe = 3.92
K_t_Joe = np.zeros([Joe])
t_Joe = Symbol('t_Joe')

```

```
Phi_Joe = -log(1-(1-t_Joe)**(theta_Joe))
Phi_prima_Joe = Phi_Joe.diff(t_Joe)

for i in range(Joe):
    G_Joe = t_Joe - (Phi_Joe / Phi_prima_Joe )
    H_Joe = G_Joe.evalf(subs={t_Joe:K_z[i]})
    K_t_Joe[i] = np.sum([H_Joe])
#K_t_Joe
```

```
In [ ]: t = np.linspace(0.01,0.99,100)
fig = plt.figure(figsize=(6,6))
fig.patch.set_facecolor('w')
plt.plot(t,t)
plt.scatter(K_z , K_t_Joe)
plt.xlim(0,1)
plt.ylim(0,1)
plt.title("Cópula")
plt.xlabel("$K_c(t)$", fontsize = 15)
plt.ylabel("$K_n(z)$", fontsize = 15)
```

3. Código en R: Prueba de aleatoriedad y prueba de independencia mutua biviada

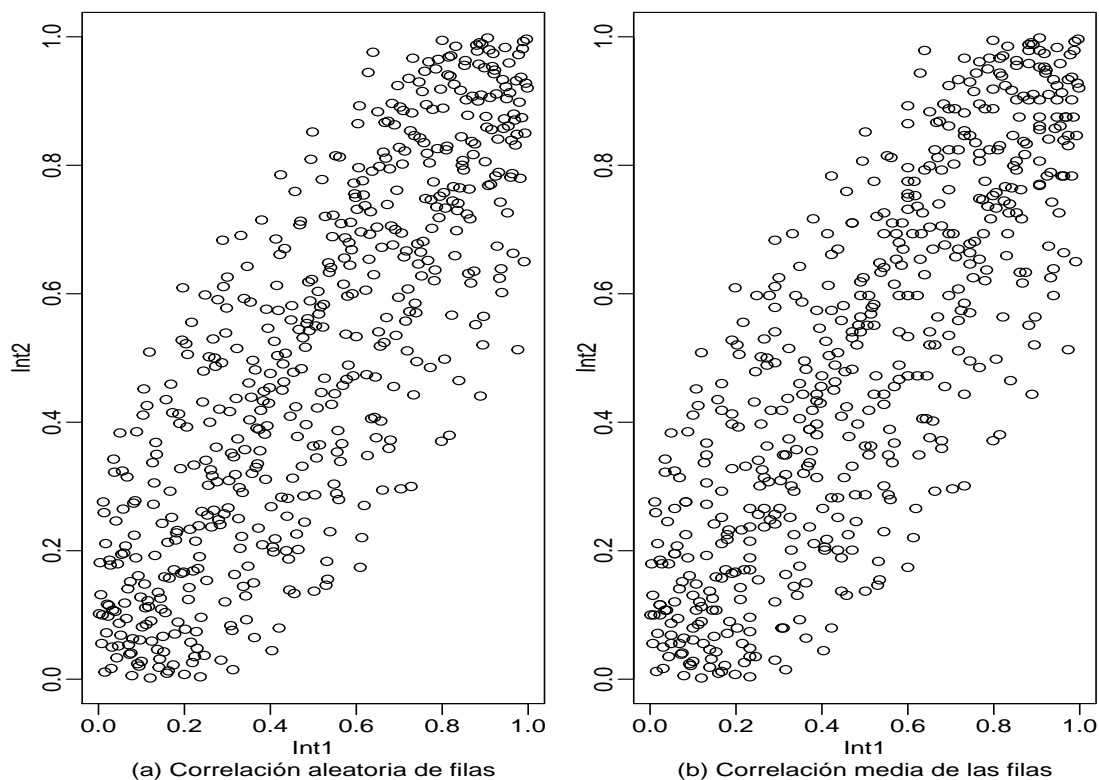
```

setwd("~/Cpulas/RStudio/Conjunta")
library("copula")
Conjunta<-read.table("Datos_Delirio_Lucia.txt", header=T)
myConjunta <- subset(Conjunta, select = c("Int1", "Int2"))
nrow(myConjunta)

## [1] 539

set.seed(123)
pseudoConjunta<-sapply(myConjunta,rank,ties.method="random")/(nrow(myConjunta)+ 1)
pseudoConjunta.ave <- sapply(myConjunta, rank)/(nrow(myConjunta) + 1)
par(mfrow = c(1, 2), mgp = c(1.5, 0.5, 0), mar = c(3.5, 2.5, 0, 0))
plot(pseudoConjunta, sub = "(a) Correlacin aleatoria de filas")
plot(pseudoConjunta.ave, sub = "(b) Correlacin media de las filas")

```



```

system.time(empsamp <- indepTestSim(nrow(pseudoConjunta),
                                     p = 2, N = 1000, verbose = 0))

## user system elapsed
## 20.52 0.00 20.64

indepTest(pseudoConjunta, empsamp)

```

4. Código en R: Prueba de bondad y ajuste para la selección de la cópula bivariada

```
##
## Global Cramer-von Mises statistic: 3.811837 with p-value 0.0004995005
## Combined p-values from the Mobius decomposition:
##   0.0004995005 from Fisher's rule,
##   0.0004995005 from Tippett's rule.

system.time(Conjunta.Gof.Gumbel.mult <- gofCopula(gumbelCopula(),
pseudoConjunta,estim.method="itau",simulation="mult", N = 1000))

##   user  system elapsed
##   2.01   0.00   2.01

Conjunta.Gof.Gumbel.mult

##
## Multiplier bootstrap goodness-of-fit test with 'method'="Sn",
## 'estim.method'="itau"
##
## data:  x
## statistic = 0.062262, parameter = 2.5639, p-value = 0.0004995

system.time(Conjunta.Gof.Frank.mult <- gofCopula(francCopula(),
pseudoConjunta,estim.method="itau",simulation="mult", N = 1000))

##   user  system elapsed
##   3.34   0.00   3.34

Conjunta.Gof.Frank.mult

##
## Multiplier bootstrap goodness-of-fit test with 'method'="Sn",
## 'estim.method'="itau"
##
## data:  x
## statistic = 0.013805, parameter = 8.2018, p-value = 0.469

system.time(Conjunta.Gof.Clayton.mult<-gofCopula(claytonCopula(),
pseudoConjunta,estim.method="itau",simulation = "mult",N = 1000))

##   user  system elapsed
##   1.92   0.00   1.92
```

```
Conjunta.Gof.Clayton.mult

##
## Multiplier bootstrap goodness-of-fit test with 'method'="Sn",
## 'estim.method'="itau"
##
## data: x
## statistic = 0.19378, parameter = 3.1277, p-value = 0.0004995

system.time(Conjunta.Gof.Joe.pb <- gofCopula(joeCopula(),
      pseudoConjunta,estim.method="itau", simulation = "pb",
      N = 1000, verbose = 0))

## user system elapsed
## 6.07 0.00 6.07

Conjunta.Gof.Joe.pb

##
## Parametric bootstrap goodness-of-fit test with 'method'="Sn",
## 'estim.method'="itau"
##
## data: x
## statistic = 0.12979, parameter = 3.9514, p-value = 0.0004995

fitCopula(frunkCopula(), pseudoConjunta, method = "itau")

## fitCopula() estimation based on 'inversion of Kendall's tau'
## and a sample of size 539.
## Estimate Std. Error
## param 8.202 0.222
```

C. Anexo: Capítulo 4. Cópulas Multivariadas

1. Determinación del parámetro θ para la cópula Ali-Mikhail-Haq mediante el método semiparamétrico

Copula Ali-Mikhail-Haq (Log-Likelihood)

September 21, 2016

```
In [16]: import scipy.misc as sp
import scipy.stats
from sympy import init_session
init_session (use_latex = True)
```

IPython console for SymPy 1.0 (Python 3.5.2-64-bit) (ground types: python)

These commands were executed:

```
>>> from __future__ import division
>>> from sympy import *
>>> x, y, z, t = symbols('x y z t')
>>> k, m, n = symbols('k m n', integer=True)
>>> f, g, h = symbols('f g h', cls=Function)
>>> init_printing()
```

```
In [17]: C = (x*y*z)/(1 - t*(1-x)*(1-y)*(1-z))
C
```

Out[17]:

$$\frac{xyz}{-t(-x+1)(-y+1)(-z+1)+1}$$

```
In [18]: c = C.diff(x,y,z)
l_Theta = C.diff(x,y,z,t)
```

```
In [29]: X = np.array([loadtxt("Datos_Intensidad.dat")])
Y = np.array([loadtxt("Datos_Intensidad.dat")])
Z = np.array([loadtxt("Datos_Duración.dat")])

# X = Intensidad El delirio (mm/h)
# Y = Intensidad santa Lucía (mm/h)
# Z = Duración Conjunta (min)
```

```
N = len(X)      # filas, número de datos en total
M = 3           # columnas, X, Y, Z.

theta = 0.08035

Vector = np.zeros([N])

for i in range(N):

    A = l_Theta / c
    B = A.evalf(subs={x:X[i], y:Y[i], z:Z[i], t:theta})

    Vector[i] = np.sum([B])

suma = sum(Vector)

print ("suma = ")
print (suma)

suma =
-0.000162670130386
```


2. Determinación del parámetro θ para la cópula Frank mediante el método semiparamétrico

Copula Frank Trivariada (Log-Likelihood)

September 21, 2016

```
In [1]: import scipy.misc as sp
import scipy.stats
from sympy import init_session
init_session (use_latex = True)
```

IPython console for SymPy 1.0 (Python 3.5.2-64-bit) (ground types: python)

These commands were executed:

```
>>> from __future__ import division
>>> from sympy import *
>>> x, y, z, t = symbols('x y z t')
>>> k, m, n = symbols('k m n', integer=True)
>>> f, g, h = symbols('f g h', cls=Function)
>>> init_printing()
```

Documentation can be found at <http://docs.sympy.org/1.0/>

```
In [2]: C = -(1/t)*log((1+((exp(-t*x) - 1)*(exp(-t*y) - 1)*(exp(-t*z) - 1))
/((exp(-t) - 1)**2))))
C
```

Out[2]:

$$-\frac{1}{t} \log \left(1 + \frac{1}{(-1 + e^{-t})^2} (-1 + e^{-tx}) (-1 + e^{-ty}) (-1 + e^{-tz}) \right)$$

```
In [ ]: c = C.diff(x,y,z)
l_Theta = C.diff(x,y,z,t)
```

```
In [ ]: X = np.array([loadtxt("Datos_Intensidad.dat")])
Y = np.array([loadtxt("Datos_Intensidad.dat")])
Z = np.array([loadtxt("Datos_Intensidad.dat")])

# X = Intensidad El delirio (mm/h)
# Y = Intensidad santa Lucía (mm/h)
# Z = Intensidad Conjunta (mm/h)
```

```
N = len(X)      # filas, número de datos en total
M = 3           # columnas, X, Y, Z.

theta = 11.2

Vector = np.zeros([N])

for i in range(N):

    A = l_Theta / c
    B = A.evalf(subs={x:X[i], y:Y[i], z:Z[i], t:theta})

    Vector[i] = np.sum([B])

suma = sum(Vector)

print ("suma = ")
print (suma)
```

3. Determinación de la función de distribución condicional del caso 1 multivariado para una duración de 15 minutos

Caso1. Juan Rey y Quiba Trivariado

September 26, 2016

```
In [1]: from mpl_toolkits.mplot3d import Axes3D
        from scipy.misc import imread
        import numpy as np
        import matplotlib.pyplot as plt
        from scipy.stats import lognorm

In [3]: fig = plt.figure()
        ax = Axes3D(fig)

        X = np.arange(0, 100, 5)
        Y = np.arange(0, 100, 5)
        X, Y = np.meshgrid(X, Y)

        mu1 = np.exp (2.6684553)
        s1 = 0.7872742
        A = lognorm.cdf(X, s1, loc=0, scale=mu1)

        mu2 = np.exp (2.333994)
        s2 = 0.717206
        B = lognorm.cdf(Y, s2, loc=0, scale=mu2)

        t = 0.15
        d = 0.1066 # Duración 15 min
        W = -t*(-A + 1)*(-B + 1)*(-d + 1) + 1
        C = -((t*A*B*d*(-A+1)*(-B+1))/(W**2)) + ((A*B)/(W))
        Z = 1/(1-C)

        ax.plot_wireframe (X, Y, Z, rstride=1, cstride=1)
        ax.set_xlabel('I (mm/h) Juan Rey ', fontsize = 13, fontweight='bold')
        ax.set_ylabel('I (mm/h) Quiba', fontsize = 13, fontweight='bold')
        ax.set_zlabel('TR (años) ' , fontsize = 13, fontweight='bold')

        plt.show()
```

4. Código en R: Prueba de aleatoriedad y prueba de independencia mutua multivariada

```
setwd("~/Cpulas/RStudio/Conjunta/Trivariado")
library("copula")
Trivariado<-read.table("Datos_Delirio_Lucia_Trivariado.txt", header=T)
nrow(Trivariado)

## [1] 539

apply(Trivariado[, 2:4], 2, function(x) length(unique(x)))

## Int1 Int2 Int3
## 266 310 495

pseudoTrivariado <- apply(Trivariado [, 2:4], 2, rank)/(nrow(Trivariado) + 1)

set.seed(123)
system.time(TrivariadoMultSerialIndepTest<-multSerialIndepTest(Trivariado[,2:4]^2,
  lag.max = 4, verbose = 0))

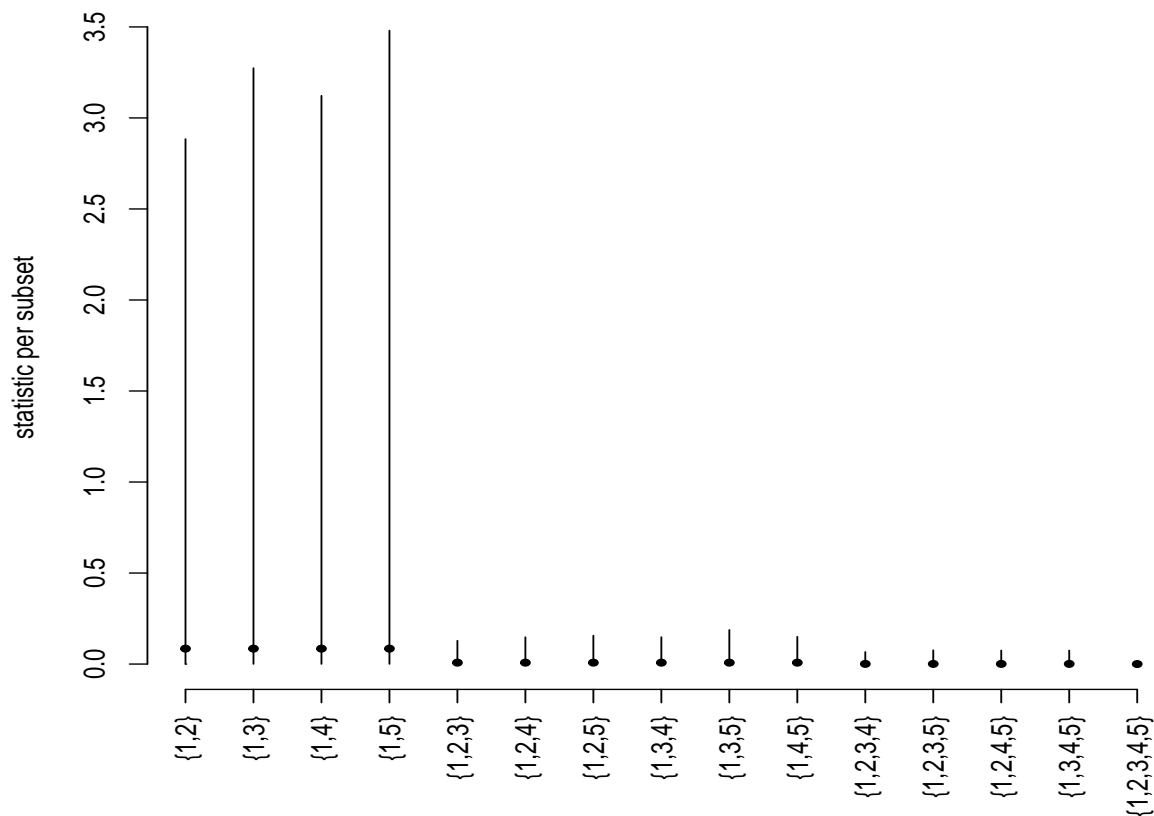
## user system elapsed
## 296.50 0.05 298.29

TrivariadoMultSerialIndepTest

##
## Global Cramer-von Mises statistic: 1.118285 with p-value 0.0004995005
## Combined p-values from the Mobius decomposition:
## 0.0004995005 from Fisher's rule,
## 0.0004995005 from Tippett's rule.

dependogram(TrivariadoMultSerialIndepTest)
dependogram(TrivariadoMultSerialIndepTest, print = TRUE)
```

Dependogram



```
## The subset statistics, p-values and critical values are:
```

```
##      subset  statistic      pvalue  critvalue
## 1      {1,2} 2.88449248 0.0004995005 0.0848224107
## 2      {1,3} 3.27421923 0.0004995005 0.0848224107
## 3      {1,4} 3.12240224 0.0004995005 0.0848224107
## 4      {1,5} 3.48064635 0.0004995005 0.0848224107
## 5     {1,2,3} 0.12879420 0.0004995005 0.0077452476
## 6     {1,2,4} 0.14798656 0.0004995005 0.0077452476
## 7     {1,2,5} 0.15703089 0.0004995005 0.0077452476
## 8     {1,3,4} 0.14851534 0.0004995005 0.0077452476
## 9     {1,3,5} 0.18813387 0.0004995005 0.0077452476
## 10    {1,4,5} 0.15053955 0.0004995005 0.0077452476
## 11    {1,2,3,4} 0.06692381 0.0004995005 0.0008893659
## 12    {1,2,3,5} 0.07636261 0.0004995005 0.0008893659
## 13    {1,2,4,5} 0.07543701 0.0004995005 0.0008893659
## 14    {1,3,4,5} 0.07554736 0.0004995005 0.0008893659
## 15   {1,2,3,4,5} 0.01188203 0.0004995005 0.0001340035
```

```
## The critical values are such that the simultaneous acceptance region
## has probability 1 - 0.05 under the null.
```

```
## The individual rejection probability for any statistic obtained from the Mobius
## decomposition is 1 - 0.9965863 under the null.
```

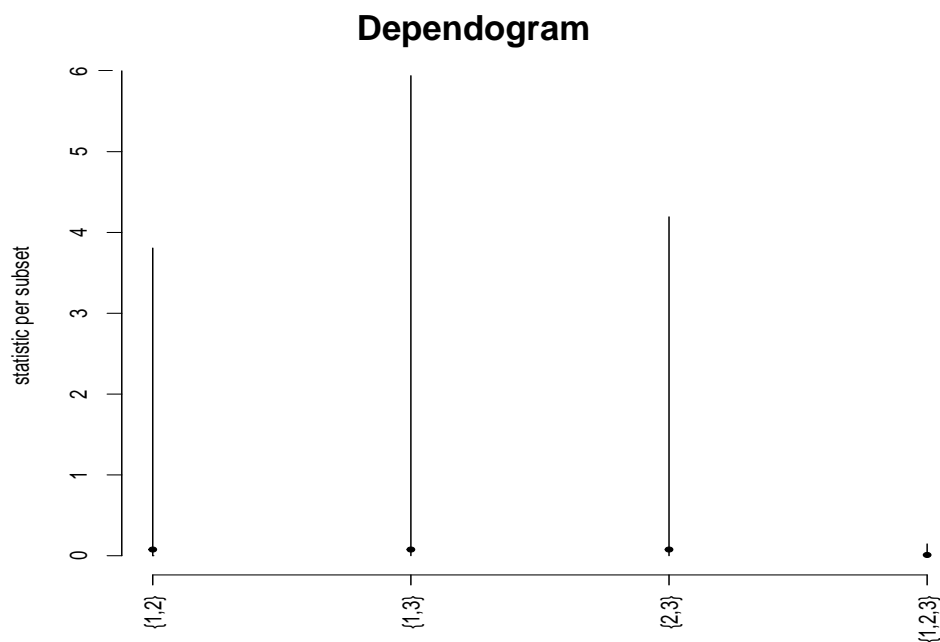
```
system.time(empsamp <- indepTestSim(nrow(pseudoTrivariado), p = 3, N = 1000,
  verbose = 0))

##    user  system elapsed
##  92.34    0.01   92.81

TrivariadoMultIndepTest <- indepTest(pseudoTrivariado, empsamp)
TrivariadoMultIndepTest

##
## Global Cramer-von Mises statistic: 9.637795 with p-value 0.0004995005
## Combined p-values from the Mobius decomposition:
##    0.0004995005 from Fisher's rule,
##    0.0004995005 from Tippett's rule.

dependogram(TrivariadoMultIndepTest)
```



5. Código en R: Prueba de bondad y ajuste para la selección de la cópula Trivariada

```
## Copula Gumbel

system.time(g.copula.mult <- gofCopula(gumbelCopula(dim = 3), pseudoTrivariado,
  estim.method = "mpl", simulation = "mult" ))

##   user  system elapsed
##   3.19   0.00   3.18

g.copula.mult

##
## Multiplier bootstrap goodness-of-fit test with 'method'="Sn",
## 'estim.method'="mpl"
##
## data:  x
## statistic = 0.9969, parameter = 2.7549, p-value = 0.0004995

fitCopula(gumbelCopula(dim = 3), pseudoTrivariado, method = "mpl")

## fitCopula() estimation based on 'maximum pseudo-likelihood'
## and a sample of size 539.
##      Estimate Std. Error
## param  2.755      0.126
## The maximized loglikelihood is 739.1
## Optimization converged
## Number of loglikelihood evaluations:
## function gradient
##      27      9

## Copula Clayton

system.time(Clay.copula.mult <- gofCopula(claytonCopula(dim = 3),
  pseudoTrivariado, estim.method = "mpl", simulation = "mult" ))

##   user  system elapsed
##   2.50   0.01   2.51

Clay.copula.mult

##
## Multiplier bootstrap goodness-of-fit test with 'method'="Sn",
## 'estim.method'="mpl"
##
## data:  x
## statistic = 2.7779, parameter = 2.3092, p-value = 0.0004995
```

```
fitCopula(claytonCopula(dim = 3), pseudoTrivariado, method = "mpl")

## fitCopula() estimation based on 'maximum pseudo-likelihood'
## and a sample of size 539.
##      Estimate Std. Error
## param    2.309      0.137
## The maximized loglikelihood is 597.9
## Optimization converged
## Number of loglikelihood evaluations:
## function gradient
##      25          5

##Copula Frank

system.time(Frank.copula.mult <- gofCopula(frunkCopula(dim = 3),
      pseudoTrivariado, estim.method = "mpl", simulation = "mult" ))

##   user  system elapsed
##   3.07   0.00   3.07

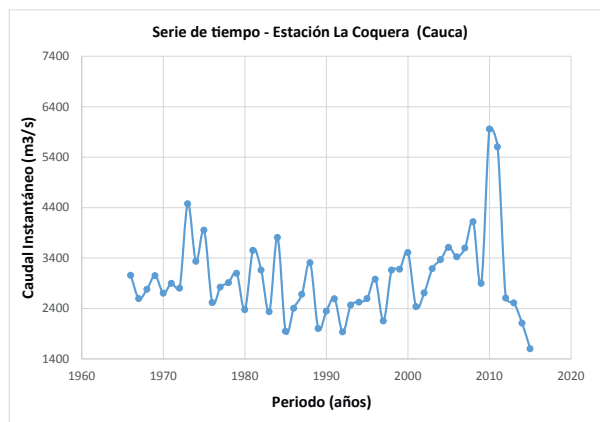
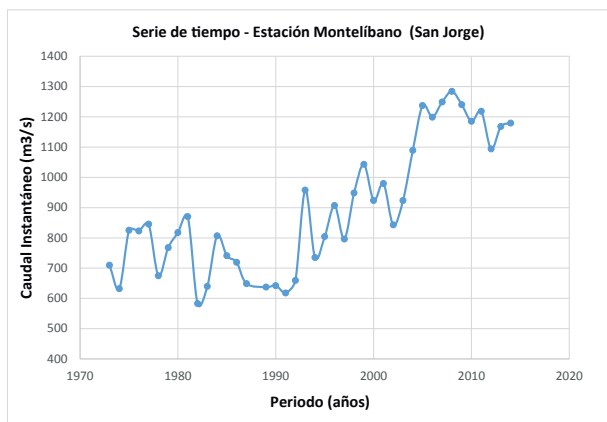
Frank.copula.mult

##
## Multiplier bootstrap goodness-of-fit test with 'method'="Sn",
## 'estim.method'="mpl"
##
## data:  x
## statistic = 0.35519, parameter = 11.18, p-value = 0.0004995

fitCopula(frunkCopula(dim = 3), pseudoTrivariado, method = "mpl")

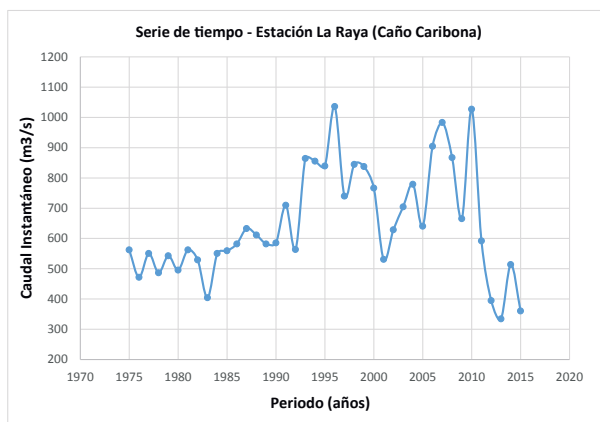
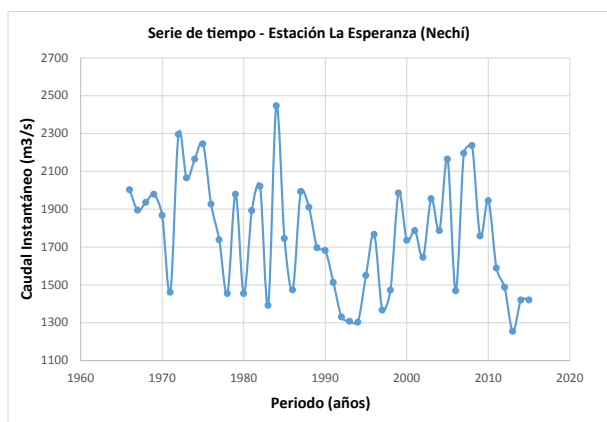
## fitCopula() estimation based on 'maximum pseudo-likelihood'
## and a sample of size 539.
##      Estimate Std. Error
## param    11.18      0.359
## The maximized loglikelihood is 840.8
## Optimization converged
## Number of loglikelihood evaluations:
## function gradient
##      30          9
```


6. Series de tiempo para las estaciones pertenecientes a la región de la Mojana



(a) Serie de tiempo de caudal instantáneo (m^3/s) para la estación Montelíbano.

(b) Serie de tiempo de caudal instantáneo (m^3/s) para la estación La Coquera.



(c) Serie de tiempo de caudal instantáneo (m^3/s) para la estación La esperanza.

(d) Serie de tiempo de caudal instantáneo (m^3/s) para la estación La Raya.

Figura C-1.: Series de tiempo de caudal instantáneo (m^3/s) para las estaciones pertenecientes a la región de la Mojana.

Bibliografía

- Acevedo Aristizábal, L. A. et al. (2009). *Estimación hidrológica bajo escenarios de cambio climático en Colombia*. PhD thesis, Universidad Nacional de Colombia.
- Adaptación, F. (2016). La mojana. [urlhttp://sitio.fondoadaptacion.gov.co/index.php/el-fondo/macroyectos/la-mojana](http://sitio.fondoadaptacion.gov.co/index.php/el-fondo/macroyectos/la-mojana).
- Aguilera, M. (2004). La mojana: riqueza natural y potencial económico. Technical report, BANCO DE LA REPÚBLICA-ECONOMÍA REGIONAL.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.
- Ali, M. M., Mikhail, N., and Haq, M. S. (1978). A class of bivariate distributions including the bivariate logistic. *Journal of multivariate analysis*, 8(3):405–412.
- Ariff, N. M., Jemain, A. A., Ibrahim, K., and Zin, W. W. (2012). Idf relationships using bivariate copula for storm events in peninsular malaysia. *Journal of hydrology*, 470:158–171.
- Aron, G., Wall, D. J., White, E. L., and Dunn, C. N. (1987). Regional rainfall intensity-duration-frequency curves for pennsylvania1. *JAWRA Journal of the American Water Resources Association*, 23(3):479–485.
- Arroyo, I., Bravo, L. C., Llinás, H., and Muñoz, F. L. (2014). Distribuciones poisson y gamma: Una discreta y continua relación. *Prospectiva*, 12(1):99–107.
- Ashkar, F. (1980). *Partial duration series models for flood analysis*. PhD thesis, École Polytechnique de Montréal.
- Bacchi, B., Becciu, G., and Kottegoda, N. T. (1994). Bivariate exponential model applied to intensities and durations of extreme rainfall. *Journal of Hydrology*, 155(1-2):225–236.
- Belzunce, F., Castaño, A., Olvera-Cervantes, A., and Suárez-Llorens, A. (2007). Quantile curves and dependence structure for bivariate distributions. *Computational Statistics & Data Analysis*, 51(10):5112–5129.

- Berg, D. (2009). Copula goodness-of-fit testing: an overview and power comparison. *The European Journal of Finance*, 15(7-8):675–701.
- Bernard, M. M. (1932). Formulas for rainfall intensities of long duration. *Transactions of the American Society of Civil Engineers*, 96(1):592–606.
- Bezak, N., Šraj, M., and Mikoš, M. (2016). Copula-based idf curves and empirical rainfall thresholds for flash floods and rainfall-induced landslides. *Journal of Hydrology*.
- Bhat, C. R. and Eluru, N. (2009). A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. *Transportation Research Part B: Methodological*, 43(7):749–765.
- Blum, J. R., Kiefer, J., and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *The annals of mathematical statistics*, pages 485–498.
- Bobée, B. and Ashkar, F. (1991). The gamma distribution and derived distributions applied in hydrology.
- Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G., and Roncalli, T. (2000). Copulas for finance-a reading guide and some applications. *Available at SSRN 1032533*.
- Briere, F. (1999). Drinking-water distribution, sewage and rainfall collection. *Polytechnique International Press*.
- Chandra, R., Saha, U., and Mujumdar, P. (2015). Model and parameter uncertainty in idf relationships under climate change. *Advances in Water Resources*, 79:127–139.
- Chang, C.-H., Tung, Y.-K., and Yang, J.-C. (1994). Monte carlo simulation for correlated variables with marginal distributions. *Journal of Hydraulic Engineering*, 120(3):313–331.
- Chaouch, M. and Goga, C. (2010). Design-based estimation for geometric quantiles with application to outlier detection. *Computational Statistics & Data Analysis*, 54(10):2214–2229.
- Chebana, F. and Ouarda, T. B. (2009). Index flood-based multivariate regional frequency analysis. *Water Resources Research*, 45(10).
- Chow, V. T., Maidment, D. R. M., Saldarriaga, L. W., Juan, G., et al. (1994). Hidrología aplicada. Technical report.

- Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 377–392.
- Coles, S. G. and Tawn, J. A. (1996). A bayesian analysis of extreme rainfall data. *Applied statistics*, pages 463–478.
- Cook, R. D. and Johnson, M. E. (1981). A family of distributions for modelling non-elliptically symmetric multivariate data. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 210–218.
- Cox, D. R. and Oakes, D. (1984). *Analysis of survival data*, volume 21. CRC Press.
- De Michele, C. and Salvadori, G. (2003). A generalized pareto intensity-duration model of storm rainfall exploiting 2-copulas. *Journal of Geophysical Research: Atmospheres*, 108(D2).
- De Michele, C., Salvadori, G., Canossi, M., Petaccia, A., and Rosso, R. (2005). Bivariate statistical approach to check adequacy of dam spillway. *Journal of Hydrologic Engineering*, 10(1):50–57.
- Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. un test non paramétrique d'indépendance. *Acad. Roy. Belg. Bull. Cl. Sci.(5)*, 65(6):274–292.
- Deheuvels, P. (1980). Non parametric tests of independence. In *Statistique non paramétrique asymptotique*, pages 95–107. Springer.
- Demarée, G. R. (1985). *Intensity-duration-frequency Relationship of Point Precipitation at Uccle: Reference Period 1934-1983*. Institut Royal Météorologique de Belgique.
- Díaz-Granados, M., Valdes, J. B., and Bras, R. (1984). A physically based flood frequency distribution. *Water Resources Research*, 20(7):995–1002.
- Doukhan, P., Fermanian, J.-D., Lang, G., et al. (2005). *Copulas of a vector-valued stationary weakly dependent process*. INSEE.
- EAAB-ESP (2014). *Desarrollo y validación de la evaluación regional del agua en la cuenca hidrográfica del río Tunjuelo*. Empresa de Acueducto, Alcantarillado y Aseo de Bogotá E.S.P. (EAAB-ESP).
- Eagleson, P. S. (1972). Dynamics of flood frequency. *Water Resources Research*, 8(4):878–898.
- Erdely, A. (2009). Cópulas y dependencia de variables aleatorias: Una introducción. *Miscelánea Matemática*, 48:7–28.

- Euscátegui, C. and Hurtado, G. (2011). Análisis del impacto del fenómeno de la "la niña" 2010-2011 en la hidroclimatología del país. *IDEAM, Bogotá, Colombia*.
- Favre, A.-C., El Adlouni, S., Perreault, L., Thiémonge, N., and Bobée, B. (2004). Multivariate hydrological frequency analysis using copulas. *Water resources research*, 40(1).
- Fisher, N. I. (1997). Copulas. *Encyclopedia of statistical sciences*.
- Fisher, R. (1992). Statistical methods for research workers. In *Breakthroughs in Statistics*, pages 66–70. Springer.
- Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge Univ Press.
- Frank, M. J. (1979). On the simultaneous associativity of (x, y) and $x+y-f(x, y)$. *Aequationes mathematicae*, 19(1):194–226.
- Garcia-Bartual, R. and Schneider, M. (2001). Estimating maximum expected short-duration rainfall intensities from extreme convective storms. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, 26(9):675–681.
- Genest, C., Favre, A.-C., Béliveau, J., and Jacques, C. (2007). Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resources Research*, 43(9).
- Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552.
- Genest, C. and MacKay, R. J. (1986). Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *Canadian Journal of Statistics*, 14(2):145–159.
- Genest, C. and Rémillard, B. (2004). Test of independence and randomness based on the empirical copula process. *Test*, 13(2):335–369.
- Genest, C. and Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. In *Annales de l'IHP Probabilités et statistiques*, volume 44, pages 1096–1127.
- Genest, C., Rémillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and economics*, 44(2):199–213.

- Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American statistical Association*, 88(423):1034–1043.
- Ghoudi, K., Kulperger, R. J., and Rémillard, B. (2001). A nonparametric test of serial independence for time series and residuals. *Journal of Multivariate Analysis*, 79(2):191–218.
- Goel, N., Kurothe, R., Mathur, B., and Vogel, R. (2000). A derived flood frequency distribution for correlated rainfall intensity and duration. *Journal of Hydrology*, 228(1):56–67.
- Goel, N., Seth, S., and Chandra, S. (1998). Multivariate modeling of flood flows. *Journal of Hydraulic Engineering*, 124(2):146–155.
- Grégoire, V., Genest, C., and Gendron, M. (2008). Using copulas to model price dependence in energy markets. *Energy risk*, 5(5):58–64.
- Grimaldi, S. and Serinaldi, F. (2006a). Asymmetric copula in multivariate flood frequency analysis. *Advances in Water Resources*, 29(8):1155–1167.
- Grimaldi, S. and Serinaldi, F. (2006b). Design hyetograph analysis with 3-copula function. *Hydrological Sciences Journal*, 51(2):223–238.
- Gumbel, E. (1958). *Statistics of extremes*.
- Gumbel, E. J. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Publ. Inst. Statist. Univ. Paris*, 9:171–173.
- Hollander, M. and Wolfe, D. A. (1973). *Nonparametric statistical methods*.
- Hosking, J. R. (1980). The multivariate portmanteau statistic. *Journal of the American Statistical Association*, 75(371):602–608.
- Huard, D., Mailhot, A., and Duchesne, S. (2010). Bayesian estimation of intensity–duration–frequency curves and of the return period associated to a given rainfall event. *Stochastic Environmental Research and Risk Assessment*, 24(3):337–347.
- Hutchinson, T. and Lai, C. D. (1990). Continuous bivariate distributions emphasising applications. Technical report.
- Jaworski, P., Durante, F., Härdle, W. K., and Rychlik, T. (2010). *Copula theory and its applications: proceedings of the workshop held in Warsaw, 25-26 September 2009*, volume 198. Springer Science & Business Media.

- Joe, H. (1993). Parametric families of multivariate distributions with given margins. *Journal of multivariate analysis*, 46(2):262–282.
- Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.
- Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford University Press on Demand.
- Johnson, N. L., Kotz, S., and Balakrishnan, N. (2002). *Continuous multivariate distributions, volume 1, models and applications*, volume 59. New York: John Wiley & Sons.
- Kao, S.-C. and Govindaraju, R. S. (2008). Trivariate statistical analysis of extreme rainfall events via the plackett family of copulas. *Water Resources Research*, 44(2).
- Katz, R. W., Parlange, M. B., and Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in water resources*, 25(8):1287–1304.
- Kojadinovic, I., Yan, J., et al. (2010). Modeling multivariate distributions with continuous margins using the copula r package. *Journal of Statistical Software*, 34(9):1–20.
- Kojadinovic, I., Yan, J., and Holmes, M. (2011). Fast large-sample goodness-of-fit tests for copulas. *Statistica Sinica*, pages 841–871.
- Kottegoda, N. T. and Rosso, R. (1997). Probability, statistics, and reliability for civil and environmental engineers.
- Koutsoyiannis, D., Kozonis, D., and Manetas, A. (1998). A mathematical framework for studying rainfall intensity-duration-frequency relationships. *Journal of Hydrology*, 206(1):118–135.
- Krstanovic, P. and Singh, V. P. (1987). A multivariate stochastic flood analysis using entropy. In *Hydrologic frequency modeling*, pages 515–539. Springer.
- Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association*, 53(284):814–861.
- Lehmann, E. L. and D’abrera, H. (1975). *Nonparametrics: statistical methods based on ranks*. Holden-Day.
- Ljung, G. M. and Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303.
- Loganathan, G., Kuo, C., and Yannaccone, J. (1987). Joint probability distribution of streamflows and tides in estuaries. *Hydrology Research*, 18(4-5):237–246.

- Mijares, A. (1992). Fundamentos de hidrología de superficie. francisco javier aparicio mijares.
- Minitab (2016). Penn state. <http://www.minitab.com/es-mx/products/minitab/>.
- Mohyomont, B., Demarée, G., and Faka, D. (2004). Establishment of idf-curves for precipitation in the tropical area of central africa-comparison of techniques and results. *Natural Hazards and Earth System Science*, 4(3):375–387.
- Moreno Chavarro, D. C. (2012). *Método para elegir una cópula arquimediana óptima/Method to choose a archimedean copula optimal*. PhD thesis, Universidad Nacional de Colombia.
- Muller, A., Bacro, J.-N., and Lang, M. (2008). Bayesian comparison of different rainfall depth–duration–frequency relationships. *Stochastic Environmental Research and Risk Assessment*, 22(1):33–46.
- Nelsen, R. (2006). An introduction to copulas, 2nd. *New York: SpringerScience Business Media*.
- Nelsen, R. B. (1999). An introduction to copulas, volume 139 of lecture notes in statistics.
- Norusis, M. (1993). Documentation spss for windows, spss. *Inc., Chicago*.
- Overeem, A., Buishand, A., and Holleman, I. (2008). Rainfall depth-duration-frequency curves and their uncertainties. *Journal of Hydrology*, 348(1):124–134.
- Paloma, V. Q. (2006). Teoría de cópulas y control de riesgo financiero.
- Peláez, J. R. T. (1978). *Cálculo hidrometeorológico de caudales máximos en pequeñas cuencas naturales*.
- Pérez, E. G. and Carballo, N. (2008). Modelo idf regional para el estado cojedes, venezuela. *Revista INGENIERÍA UC*, 15(3):53–63.
- Prodanovic, P. and Simonovic, S. P. (2007). *Development of rainfall intensity duration frequency curves for the City of London under the changing climate*. Department of Civil and Environmental Engineering, The University of Western Ontario.
- Python (2015). Van rossum. <https://www.python.org>.
- Rao, A. and Hamed, K. (2000). Flood frequency analysis. crc press. *Boca Raton, FL*.
- Raynal-Villasenor, J. and Salas, J. (1987). Multivariate extreme value distributions in hydrological analyses. *IN: Water for the Future: Hydrology in Perspective. IAHS Publication*, (164).

- Renard, B. and Lang, M. (2007). Use of a gaussian copula for multivariate extreme value analysis: some case studies in hydrology. *Advances in Water Resources*, 30(4):897–912.
- Rota, G.-C. (1964). On the foundations of combinatorial theory i. theory of möbius functions. *Probability theory and related fields*, 2(4):340–368.
- Sackl, B. and Bergmann, H. (1987). A bivariate flood model and its application. In *Hydrologic frequency modeling*, pages 571–582. Springer.
- Salvadori, G. and De Michele, C. (2004). Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water Resources Research*, 40(12).
- Salvadori, G. and De Michele, C. (2010). Multivariate multiparameter extreme value models and return periods: A copula approach. *Water resources research*, 46(10).
- Sarría, F. A. (2006). Sistemas de información geográfica. *Universidad de Murcia*.
- Schliep, E. M., Cooley, D., Sain, S. R., and Hoeting, J. A. (2010). A comparison study of extreme precipitation from six different regional climate models via spatial hierarchical modeling. *Extremes*, 13(2):219–239.
- Schweizer, B. and Wolff, E. F. (1981). On nonparametric measures of dependence for random variables. *The annals of statistics*, pages 879–885.
- Serfling, R. (2002). Quantile functions for multivariate analysis: approaches and applications. *Statistica Neerlandica*, 56(2):214–232.
- Serinaldi, F. and Grimaldi, S. (2007). Fully nested 3-copula: procedure and application on hydrological data. *Journal of Hydrologic Engineering*, 12(4):420–430.
- Sherman, C. W. (1931). Frequency and intensity of excessive rainfalls at boston, massachusetts. *Transactions of the American Society of Civil Engineers*, 95(1):951–960.
- Shiau, J. (2003). Return period of bivariate distributed extreme hydrological events. *Stochastic environmental research and risk assessment*, 17(1-2):42–57.
- Singh, K. and Singh, V. (1991). Derivation of bivariate probability density functions with exponential marginals. *Stochastic Hydrology and Hydraulics*, 5(1):55–68.
- Singh, V. P. and Zhang, L. (2007). Idf curves using the frank archimedean copula. *Journal of Hydrologic Engineering*, 12(6):651–662.
- Sklar, M. (1959). *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8.

- Stewardson, M. J. and McMahon, T. A. (2002). A stochastic model of hydraulic variations within stream channels. *Water Resources Research*, 38(1).
- Tippett, L. (1931). The method of statistics.
- Van de Vyver, H. (2015a). Bayesian estimation of rainfall intensity–duration–frequency relationships. *Journal of Hydrology*, 529:1451–1463.
- Van de Vyver, H. (2015b). On the estimation of continuous 24-h precipitation maxima. *Stochastic Environmental Research and Risk Assessment*, 29(3):653–663.
- Wenzel, H. G. (1982). Rainfall for urban stormwater design. *Urban stormwater hydrology*, pages 35–67.
- Yue, S. (1999). Applying bivariate normal distribution to flood frequency analysis. *Water International*, 24(3):248–254.
- Yue, S. (2000a). The bivariate lognormal distribution to model a multivariate flood episode. *Hydrological Processes*, 14(14):2575–2588.
- Yue, S. (2000b). The gumbel logistic model for representing a multivariate storm event. *Advances in Water Resources*, 24(2):179–185.
- Yue, S. (2000c). Joint probability distribution of annual maximum storm peaks and amounts as represented by daily rainfalls. *Hydrological Sciences Journal*, 45(2):315–326.
- Yue, S., Ouarda, T., Bobée, B., Legendre, P., and Bruneau, P. (1999). The gumbel mixed model for flood frequency analysis. *Journal of hydrology*, 226(1):88–100.
- Yue, S. and Rasmussen, P. (2002). Bivariate frequency analysis: discussion of some useful concepts in hydrological application. *Hydrological Processes*, 16(14):2881–2898.
- Yue, S. and Wang, C. (2004). A comparison of two bivariate extreme value distributions. *Stochastic Environmental Research and Risk Assessment*, 18(2):61–66.
- Zhang, L. (2005). *Multivariate hydrological frequency analysis and risk mapping*. PhD thesis, Beijing Normal University.
- Zhang, L. and Singh, V. P. (2007a). Bivariate rainfall frequency distributions using archimedean copulas. *Journal of Hydrology*, 332(1):93–109.
- Zhang, L. and Singh, V. P. (2007b). Gumbel–hougaard copula for trivariate rainfall frequency analysis. *Journal of Hydrologic Engineering*, 12(4):409–419.