



# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE POSGRADO EN CIENCIAS DE LA TIERRA

## EVENTOS EXTREMOS DE PRECIPITACIÓN: ESTUDIO DE CASO EN APIZACO, TLAXCALA

### TESIS

QUE COMO REQUISITO PARA OBTENER EL GRADO DE:  
MAESTRO EN CIENCIAS DE LA TIERRA

PRESENTA

MÓNICA ADRIANA LOZANO JUÁREZ

JURADO EXAMINADOR

- 1) DRA. ANA CECILIA CONDE ÁLVAREZ
- 2) DR. CARLOS GAY GARCÍA
- 3) DRA. MARÍA ENGRACIA HERNÁNDEZ CERDA
- 4) M. EN C. RENÉ GARDUÑO LÓPEZ
- 5) DR. ERNESTO CAETANO NETO



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

# Agradecimientos Especiales

Agradezco especialmente al Maestro en Ciencias Óscar Sánchez Meneses y al Licenciado Francisco Estrada Porrúa por su asesoría en la elaboración de esta tesis “Eventos extremos de precipitación: estudio de caso en Apizaco, Tlaxcala”

# Índice

<b>Resumen</b>	1
<b>I-Introducción</b>	2
Forzamiento Radiativo	2
Gases de Efecto Invernadero	4
Aerosoles	5
Forzantes Naturales	5
Calentamiento Global	5
Eventos Extremos	7
Definición	8
Metodología	9
Procedimiento	10
Teoría de Valores Extremos	11
Región de Estudio	12
<b>II-Método y Resultados</b>	16
<b>Climatología</b>	16
Distribución de la Precipitación	18
Datos	22
Umbral	22
El Niño Oscilación Sur	22
<b>Índices de Eventos Extremos</b>	24
Índices	24
Resultados	26
Análisis	36

<b>Modelos</b>	37
Modelos Asintóticos	39
Distribución General de Valores Extremos	41
Tamaño de Bloques	42
Estimación de los Parámetros	43
Niveles de Retorno	43
Verificación del Modelo	45
Resultados	46
Modelos de Umbral	63
Distribución Pareto Generalizada	63
Selección del Umbral	65
Estimación de los Parámetros	68
Niveles de Retorno	68
Verificación del Modelo	71
Resultados	72
Procesos Puntuales	79
Teoría	79
Ley de convergencia	81
Resultados	83
Comparación	87
Serie No Estacionaria	88
<b>III-Conclusiones</b>	92
<b>Referencias</b>	94
<b>Apéndice</b>	96
A.1 Términos	96
A.2 Temperatura de Brillo	96
A.3 Absorción y Emisión en la Atmósfera	96
A.4 Cambio de Temperatura por Forzamiento Radiativo	97
A.5 R	97

A.6 Rclimdex	97
A.7 ExtRemes	98
A.8 Diagrama de Caja	98
A.9 Perfil del Logaritmo de la Verosimilitud	98
A.10 Métodos de Optimización	99
A.11 Procesos de Poisson	100
A.12 Método Delta	100

# Resumen

El objetivo de esta tesis es desarrollar una metodología para caracterizar eventos extremos de precipitación. Los eventos climáticos extremos generan grandes pérdidas tanto humanas como económicas. Los eventos extremos podrían aumentar en frecuencia, intensidad y duración debido al calentamiento global, por lo que su estudio es de suma importancia. Se ha tomado el municipio de Apizaco en el estado de Tlaxcala como estudio de caso. La caracterización se llevó a cabo en tres partes: La primera consistió en determinar la climatología normal del lugar y considerar El Niño Oscilación Sur (ENOS) como un factor que podría afectar la ocurrencia e intensidad de los eventos extremos de precipitación. Posteriormente se desarrolló un método para estimar cómo han cambiado en intensidad, frecuencia y duración los eventos extremos de precipitación diarios, haciendo uso de índices de precipitación. Finalmente se usó la Teoría de Valores extremos para estimar cómo se comportarán los eventos extremos de precipitación en el futuro a partir de los datos observados. Para las últimas dos partes del análisis se usaron dos herramientas del software libre R, Rclimindex y extRemes, respectivamente. En conjunto todo el análisis es una herramienta para la adaptación y la toma de decisiones presentes y futuras. Se observó que ENOS no es un factor que afecta los eventos extremos de precipitación en la región. Como resultado del análisis de índices se sabe que la precipitación total ha permanecido sin cambios, mientras que el número de días con tormentas severas ha aumentado al igual que la máxima anual en un día de precipitación; sólo un índice disminuyó, el índice simple de intensidad. Se usaron dos familias de modelos, los asintóticos y los de umbral, ambos modelos se ajustaron exitosamente y arrojaron resultados importantes como niveles de retorno de más de 100 mm para los próximos 100 años. Todos estos resultados muestran que los eventos extremos de precipitación en la región de estudio han cambiado. La metodología usada en este trabajo puede ser usada en otras regiones lo cual será una herramienta útil en la toma de decisiones con el objetivo de reducir la vulnerabilidad ante eventos extremos de precipitación.

# I-Introducción

## **Forzamiento Radiativo**

La energía primordial que alimenta el sistema climático proviene de la radiación emitida por el Sol. La superficie de la Tierra absorbe parte de esta radiación y redistribuye su energía mediante circulaciones atmosféricas y oceánicas. Después de cierto tiempo se alcanza el equilibrio térmico, en consecuencia la energía emitida por el sol que es absorbida por el sistema Tierra-Atmósfera debe ser reemitida al espacio para que se mantenga el equilibrio entre energía recibida y emitida. Así entonces esta energía neta ganada, es irradiada nuevamente al espacio en longitudes de onda más largas, es decir infrarrojas. De la radiación total que llega al sistema Tierra-Atmósfera aproximadamente el 30% es reflejada en el tope de ella por el albedo terrestre, el restante es absorbido y dispersado en la atmósfera. La parte que llega a la superficie de la Tierra es absorbida 70% por el océano y 30% por la tierra.

La radiación térmica infrarroja emitida por la Tierra genera un espectro muy cercano al emitido por un cuerpo negro a 290 K (figura 1-l), que es aproximadamente la temperatura de la superficie, dicho espectro se muestra como una curva no diferenciable entre las curvas suaves de los espectros de cuerpo negro a 275 K y 300 K. Los picos de dicha curva que se alejan de la del espectro de cuerpo negro a 290 K corresponden a longitudes de onda absorbidas por gases, como vapor de agua, bióxido de carbono y metano, presentes en la atmósfera. Estos gases que absorben energía térmica infrarroja dejándola atrapada en la atmósfera, y contribuyendo a aumentar la temperatura de ésta, son conocidos como gases de efecto invernadero o GEI. El efecto invernadero es el resultado del mecanismo mediante el cual los gases antes mencionados evitan que la radiación infrarroja proveniente de la superficie terrestre escape al espacio.

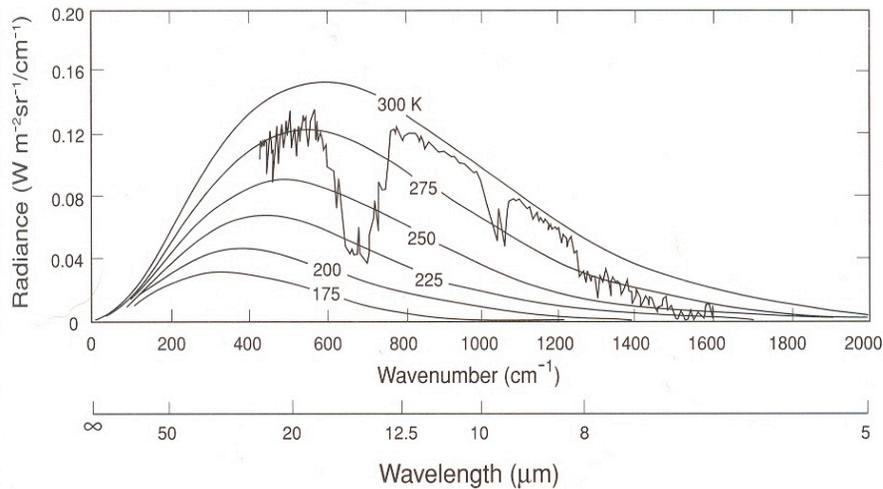


Figura 1-I. Espectro de radiación térmica infrarroja terrestre comparada con espectro electromagnético de cuerpo negro (Salby, 1996). Las curvas suaves corresponden a la radiación de cuerpo negro a diferentes temperaturas. La curva no diferenciable entre las temperaturas de 275 y 300 K corresponde a la radiación infrarroja emitida por la Tierra y liberada al espacio.

La energía proveniente de la radiación solar que ingresa se equilibra aproximadamente con la radiación terrestre saliente. Cualquier factor que altere la radiación recibida del Sol o emitida al espacio, o que perturbe la redistribución de energía dentro de la atmósfera y entre atmósfera, tierra y océano, puede afectar el clima.

Estos factores son los forzamientos radiativos, específicamente se refieren a cambios en la energía radiativa neta disponible para el sistema Tierra-atmósfera. Los positivos tienden a calentar la superficie de la Tierra y la atmósfera inferior, los negativos tienden a enfriarlas. El origen de estos factores puede ser tanto natural como antropogénico.

Normalmente la atmósfera absorbe radiación terrestre y la emite hacia mayores altitudes y temperaturas más bajas, si esto no ocurre se produce un forzamiento radiativo positivo que tiende a calentar la atmósfera inferior y la superficie. Si se emite menos calor hacia el espacio, se refuerza el efecto invernadero.

Cuando cambian los forzamientos radiativos, el sistema climático responde en diversas escalas temporales. Las más prolongadas se deben a la gran capacidad de almacenamiento de calor de las profundidades de los océanos y al ajuste dinámico de los mantos de hielo. Esto significa que la respuesta a un cambio, ya sea positivo o negativo puede durar miles de años. Todo cambio en el equilibrio radiativo de

la Tierra, incluso los debidos a un incremento en los gases de efecto invernadero o en los aerosoles, alterará el ciclo hidrológico, así como la circulación atmosférica y oceánica, afectando por tanto los patrones meteorológicos, las temperaturas y las precipitaciones (IPCC 2007).

El cambio en el clima provocado por los seres humanos se añadirá a las variaciones climáticas naturales que se producen en toda una gama de escalas temporales y espaciales. La variabilidad climática puede ser resultado de cambios naturales en el forzamiento del sistema climático, como variaciones de intensidad de la radiación solar entrante y cambios en las concentraciones de aerosoles producidos por erupciones volcánicas. También pueden producirse variaciones climáticas naturales sin que exista un cambio en el forzamiento externo, como resultado de complejas interacciones entre los componentes del sistema climático, como en la interfase entre la atmósfera y los océanos. El Niño-Oscilación Sur (ENOS) es un ejemplo de esa variabilidad natural en escalas temporales interanuales (IPCC 2001).

### **Gases de Efecto Invernadero**

El efecto invernadero ha ocurrido en la atmósfera durante millones de años, debido a la presencia de gases que se producen naturalmente. Sin embargo la eficiencia con la que la superficie de la Tierra irradia energía al espacio se ve reducida por el aumento en las concentraciones de dichos gases, entre otros factores como cambios en el uso del suelo que modifican el *albedo*<sup>1</sup> de la superficie.

Como se observa en la figura 1-1, parte del espectro de longitud de onda larga emitido por la Tierra es absorbido en ciertas bandas por algunos gases presentes en la atmósfera, estos son los gases de efecto invernadero (GEI). Los principales absorbentes son vapor de agua, dióxido de carbono y ozono, así como óxido nitroso, metano y óxido nítrico, que lo hacen en menor proporción dada su baja concentración en la atmósfera.

La cantidad de forzamiento radiativo depende de la magnitud del aumento en la concentración de cada GEI, de las propiedades radiativas de los gases y de las concentraciones de otros GEI ya presentes en la atmósfera. Además, muchos GEI permanecen en la atmósfera durante decenas de años después de haber sido emitidos.

La mayoría de los cambios en las temperaturas extremas, están relacionados con un incremento en la temperatura media global, esta señal se observa más claramente en las temperaturas mínimas. Este aumento es producto, al menos en parte, del incremento en los gases invernadero producidos por actividades humanas (IPCC 2007), lo que hace pensar que el cambio en las temperaturas extremas

---

<sup>1</sup> Albedo: La razón entre el flujo de radiación solar reflejado y el flujo incidente.

también debería, al menos en parte, estar generado por actividades humanas. El efecto invernadero antropogénico está relacionado con cambios en la variabilidad de la temperatura y en el ciclo hidrológico.

### **Aerosoles**

Los aerosoles son partículas microscópicas que pueden ser tanto sólidas como líquidas, de origen antropogénico o natural, y se encuentran en la troposfera. Los de origen antropogénico se derivan de combustibles fósiles y de la combustión de biomasa, mientras que los de origen natural se generan de sal de mar, polen y erupciones volcánicas. Algunos de estos aerosoles pueden reflejar la radiación solar, lo que tiende a enfriar el sistema climático. Sin embargo, los aerosoles de hollín calientan el sistema. Además, los cambios en las concentraciones de aerosoles pueden alterar la cantidad y el albedo de las nubes, por su efecto sobre las propiedades y duración de éstas.

En la mayoría de los casos, los aerosoles troposféricos tienden a producir un forzamiento radiativo negativo, que enfría el sistema climático. Tienen una duración breve de días a semanas, mientras que la mayor parte de los gases invernadero duran decenas de años e incluso siglos, como resultado, sus concentraciones responden mucho más rápidamente a los cambios en las emisiones.

La actividad volcánica puede inyectar en la estratosfera grandes cantidades de gases de azufre, sobre todo, anhídrido sulfuroso, que se transforman en aerosoles de sulfatos. Las erupciones individuales pueden producir un gran forzamiento radiativo negativo, aunque momentáneo, tendiente a enfriar la superficie de la Tierra y la atmósfera inferior por algunos años.

### **Forzantes Naturales**

La generación de energía del sol varía en pequeñas cantidades (0,1%) en un ciclo de actividad de once años, aunque también pueden producirse variaciones durante períodos más prolongados. En escalas temporales de decenas a miles de años, las lentas variaciones en la órbita de la Tierra han ocasionado cambios en la distribución estacional y latitudinal de la radiación solar. Esos cambios han desempeñado un importante papel al controlar las variaciones del clima, como en los ciclos glaciales e interglaciales.

### **Calentamiento Global**

Cambios en los factores anteriormente mencionados, especialmente el aumento en las concentraciones de gases de efecto invernadero, han dado como resultado un calentamiento global. Como consecuencia de éste, se tienen diferentes cambios en el clima que se han documentado (IPCC 2007) en particular en las últimas décadas del siglo XX. En conjunto, estas tendencias ilustran la imagen completa de un período de calentamiento, y se muestran a continuación:

- Se han medido y ajustado independientemente registros de la temperatura en la superficie de la tierra y el océano. Todos los conjuntos de datos muestran tendencias mundiales en ascenso, con dos períodos principales de calentamiento en todo el mundo: de 1910 a 1945 y desde 1976. Existe una creciente tendencia a que las temperaturas de la atmósfera en la superficie terrestre, sobre todo el mundo, aumenten más rápido que las temperaturas en la superficie oceánica.

- Las mediciones con globos meteorológicos muestran que las temperaturas en la troposfera inferior han estado aumentando desde 1958, aunque sólo levemente desde 1979. A partir de 1979, se cuenta con datos satelitales, que muestran tendencias similares a los datos recogidos por globos.

- La reducción del margen de variación de la temperatura diurna en la superficie de los continentes coincide con los aumentos en la nubosidad, las precipitaciones y el aumento en el contenido de vapor de agua total en la atmósfera.

- La disminución casi mundial en la extensión de los glaciares de montaña y de la masa de hielo coincide con los aumentos de la temperatura en la superficie. Unas pocas excepciones recientes en las regiones costeras son coherentes con las variaciones en la circulación atmosférica y los correspondientes aumentos en las precipitaciones.

- Las reducciones en la capa de nieve y el acortamiento de las temporadas de congelación en lagos y ríos se relacionan bien con los aumentos de temperatura en la superficie terrestre, en el hemisferio norte.

- La reducción sistemática de la extensión del hielo marino en primavera y verano y de su espesor en el Ártico es coherente con los aumentos de temperatura en la mayoría de las tierras y océanos adyacentes.

- El contenido de calor de los océanos se ha incrementado y ha subido el nivel medio del mar en todo el mundo, debido a la expansión térmica.

- Los aumentos en el vapor de agua total en la troposfera en los últimos 25 años son cualitativamente coherentes con los aumentos de las temperaturas en la troposfera y con un ciclo hidrológico aumentado, que provoca precipitaciones más intensas y frecuentes en las latitudes medias y altas del hemisferio norte.

Sin embargo también hay algunos aspectos importantes del clima que parecen no haber cambiado (IPCC 2001):

- Unas cuantas zonas del globo no se han calentado en los últimos decenios, principalmente en algunas partes de los océanos del hemisferio sur y en partes de la Antártida.

- No hay tendencias significativas claras en la extensión del hielo marino en la Antártida desde 1978, período en que se han registrado mediciones sistemáticas por satélite.

## Eventos Extremos

La Tierra se está calentando y el clima está cambiando, debido a las actividades humanas. Los efectos pueden ser muy variados, pero uno de los que genera más pérdidas económicas y humanas son los eventos extremos. Mucho se ha especulado al respecto, pero es posible que al añadir más energía térmica al sistema, la frecuencia e intensidad de los eventos extremos se incrementen, esto es potencialmente probable en sistemas con una atmósfera más cálida (Houghton 1994).

La mayoría de los impactos generados por variaciones climáticas son producto de la ocurrencia y la frecuencia de eventos extremos, como grandes precipitaciones, tormentas de viento y temperaturas extremas durante algunas horas o algunos días.

La percepción general es que el número de eventos extremos ha aumentado a través de los años. Así como el costo de los daños, sin embargo las pérdidas económicas han crecido más que las pérdidas humanas (Karl T. R. 1999). Las aseguradoras han reportado un incremento en los costos resultantes de eventos extremos desde 1970, en parte debido a los cambios recientes en la frecuencia e intensidad de algunos tipos de eventos extremos, por lo que analizar cómo se desarrollan estos eventos es de suma importancia (Beniston 2004).

En la tabla 1 se muestra un resumen de eventos extremos y su probabilidad de ocurrencia en el presente y en el futuro (IPCC 2007)

<b>Tabla 1: Eventos Extremos, Probabilidad de Ocurrencia</b>			
<b>Fenómeno y Tendencia</b>	<b>Probabilidad de que la tendencia ocurrió en el siglo XX (después de 1960)</b>	<b>Probabilidad de contri-bución humana a la ten-dencia observada</b>	<b>Probabilidad de tenden-cias futuras basadas en proyecciones</b>
Días cálidos, menos días y noches fríos	90%	66%	99%
Días y noches cálidos más frecuentes	90%	66% (Noches)	99%
Incremento en la frecuencia de ondas de calor	66%	50%	90%
Eventos de precipitación fuerte. Incremento de la frecuencia en todas las áreas	66%	50%	90%
Áreas afectadas por incremento en sequías	66% desde 1970	50%	66%
Incremento en la intensidad y actividad de ciclones tropicales	66% desde 1970	50%	66%

## Definición

Existen muchas definiciones de lo que es un evento extremo, por ejemplo el Panel Intergubernamental de Cambio Climático -IPCC, por sus siglas en inglés- lo define como un evento tan raro como el décimo o el nonagésimo percentil de una distribución particular de una variable atmosférica, como temperatura, precipitación o velocidad del viento (IPCC 2001). En otras palabras los eventos extremos son aquellos acontecimientos que están más alejados de la media o la mediana.

Sin embargo, no existe una definición universal de eventos extremos. En algunos casos los eventos extremos pueden ser definidos como la máxima o la mínima distribución de una variable durante un cierto periodo, también pueden ser definidos como aquellos eventos que exceden en magnitud un cierto umbral. O simplemente se puede definir como el valor más grande o más chico de una variable atmosférica en una serie de datos.

Los eventos extremos se pueden cuantificar como (Beniston 2004):

- Qué tan raros son: Se analiza la frecuencia con que ocurren.
- Qué tan intensos son: Se establece un umbral y se estudian los eventos que lo sobrepasan.
- Cómo impactan: Se analizan los efectos que producen sobre el ambiente o en sectores económicos, en términos de costos o daños.
- Cuánto duran: Se analizan días consecutivos de precipitación.

Necesitamos detectar cambios en el clima y en los eventos extremos, luego saber si estos cambios son normales dentro de la variabilidad natural del clima y además detectar evidencia que relacione los cambios observados en extremos con los efectos antropogénicos, ya que lo que sí sabemos es que la concentración de gases de efecto invernadero aumenta y con ello la cantidad de vapor de agua disponible. Entender el potencial del cambio climático en términos de tendencias y cambios en los eventos extremos es sumamente importante.

Debe quedar claro que cambios en una sola variable, como la temperatura global anual, no necesariamente implican que el cambio climático está ocurriendo. Cambios en la variabilidad son más importantes que cambios en la media (Katz 1992).

La precipitación en general se ha incrementado sobre casi todas las áreas en latitudes altas (IPCC 2001) El incremento se ha dado tanto en frecuencia como en intensidad en EUA (Karl et al 1996) .

Nuevos análisis muestran que en las regiones en que la precipitación total ha aumentado, es muy probable que haya habido aumentos más pronunciados aún en episodios de precipitaciones intensas y extremas. También ocurre lo contrario, algunas regiones como Sudáfrica, Siberia, Japón, Noreste de los Esta-

dos Unidos y centro de México están experimentando un aumento en precipitaciones fuertes, mientras que la precipitación total o el número de días con precipitación se han mantenido igual o incluso ha decrecido (IPCC 2007). Esto se atribuye a una disminución en la frecuencia de los eventos de precipitación.

En general, es probable que para muchas zonas en las latitudes medias y altas, principalmente en el hemisferio norte, se hayan producido aumentos estadísticamente significativos en la proporción de precipitaciones anuales totales que corresponde a episodios de precipitaciones intensas y extremas; es probable que haya habido un aumento del 2 al 4% en la frecuencia de los episodios de precipitaciones intensas en la última mitad del siglo XX.

A futuro se prevé que las precipitaciones extremas aumentarán tanto en frecuencia como en intensidad, alcanzando valores superiores al promedio. Se proyecta que la frecuencia de las precipitaciones extremas se incrementará en casi todo el mundo (IPCC 2007).

## **Metodología**

Existen varias maneras de abordar el tema de eventos extremos dependiendo de la información que queremos obtener de ellos.

1. Método estadístico, donde se hace una exploración rigurosa de los cambios en extremos y se determinan los índices estadísticamente.
2. Usando Modelos de Circulación General para construir escenarios de cambio climático y con ello determinar la distribución de eventos extremos.

En este estudio el método a usar es primordialmente estadístico, utilizando datos diarios.

El análisis se llevará a cabo en tres partes:

- Manejo básico de los datos.
- Cálculo de índices de eventos extremos.
- Ajuste de modelos.

El método consiste en hacer una exploración rigurosa de los cambios en eventos extremos y determinar estadísticamente los umbrales a partir de los cuales se encontrarán dichos eventos. Por tratarse de precipitación, y de datos diarios, el umbral que nos interesa es el del percentil 90 de la distribución, que corresponde a lluvias torrenciales.

Posteriormente se calcularán índices que nos ayudarán a determinar el cambio en frecuencia, intensidad y duración de dichos eventos; para ello se utilizará la herramienta *RClimdex*.<sup>2</sup>

Así mismo se obtendrán modelos que nos permitirán estimar el comportamiento futuro de los eventos extremos extrapolando a partir de los datos actuales, para ello se usará la *Teoría General de Valores Extremos* y la herramienta *ExtRemes*.<sup>3</sup>

## **Procedimiento**

Primeramente se debe verificar que la serie de tiempo sea suficientemente larga, al menos 30 años, y no contenga periodos muy extensos sin datos, como años completos a la mitad de la serie de tiempo.

Posteriormente se establece la climatología normal del lugar (figura 3-II), esto es, el promedio mensual de la precipitación, temperatura máxima y mínima del periodo 1961 a 1990. Observando la figura 3-II se puede tener una idea del ciclo estacional que presenta la precipitación. Gracias a este ciclo se tienen diferentes distribuciones de la precipitación para cada mes, lo cual es evidente al observar la figura 5-II. En algunos casos, como el *método de bloques*, será necesario separar los meses húmedos de los secos al hacer el análisis. Esto debido a que dicho método usa valores máximos en un cierto periodo de tiempo y dada la diferencia entre las distribuciones mensuales no sería muy adecuado comparar los máximos de un mes húmedo como julio con los de un mes seco como enero ya que la diferencia entre esos valores es de casi un orden de magnitud. En cuanto al ajuste de los *modelos de umbral*, únicamente será necesario introducir la tendencia del ciclo en los parámetros.

La segunda parte consiste en llevar a cabo el cálculo de una serie de índices de eventos extremos, usando como base las funciones descritas por Coles (2001).

Finalmente se ajustarán dos modelos que permiten saber el comportamiento que los eventos extremos han tenido y que probablemente tendrán en el futuro, esto sin incluir cambios en los forzantes climáticos actuales. La base de estos modelos se encuentra en la *Teoría de Valores Extremos* y cada modelo corresponde a las dos diferentes formas que existen de abordar el problema bajo esta teoría.

La formulación matemática en la que se fundamenta la *Teoría de Valores Extremos*, así como sus métodos y las fórmulas en las que se basaron los programas *RClimdex* y *ExtRemes* se puede encontrar en el libro de Coles (2001).

---

<sup>2</sup> Ver Apéndice A.6.

<sup>3</sup> Ver Apéndice A.7.

## Teoría de Valores Extremos

El objetivo de esta teoría es cuantificar el comportamiento estocástico de procesos inusualmente grandes, o pequeños (Coles 2001). La teoría de valores extremos da la posibilidad de estimar la probabilidad de ocurrencia de eventos extremos futuros usando un periodo relativamente corto de datos. Incluso obteniendo la estimación de eventos más inusuales o extremos que los que se hayan observado hasta ese momento. Debido a que no existe una regla física que determine el comportamiento de las variables bajo estudio, extrapolar valores futuros no es algo tan sencillo.

La teoría de valores extremos se basa en el hecho de que bajo suposiciones convenientes, el comportamiento aproximado de  $M_n$ , donde éste es el máximo de una secuencia de  $n$  observaciones, puede ser extrapolado para valores grandes de  $n$ , en el límite cuando  $n \rightarrow \infty$  se tiene una familia de modelos que pueden ser calibrados mediante valores observados de  $M_n$ , esto se conoce como el paradigma del valor extremo.

Para que esto funcione se tiene que hacer la suposición de que el mecanismo estocástico del proceso modelado es suficientemente suave para permitir la extrapolación hacia valores no observados.

La teoría tiene limitaciones que surgen de usar el paradigma del valor extremo. Dichas limitaciones radican primeramente en el hecho de que los modelos se desarrollan bajo suposiciones asintóticas; por tanto no se deben tomar como exactos los resultados provenientes de muestras finitas. Segundo, el modelo se deriva de circunstancias ideales, como suponer que los forzantes climáticos se mantienen constantes en el futuro; este tipo de suposiciones pueden no ser exactas o adecuadas para un proceso en particular. Tercero, algunos métodos desperdician mucha información ya que sólo toman el máximo o el mínimo de los datos en un periodo de tiempo; tal es el caso del *Método de Bloques* que ajusta una función de distribución a valores máximos o mínimos (anuales, mensuales o estacionales) que pueden dejar fuera eventos extremos de algunos años que incluso pudieran ser mayores o menores a los de otros años.

Por estas limitaciones la teoría de eventos extremos debe ir acompañada de herramientas estadísticas a manera de complemento.

El análisis bajo la Teoría de Valores Extremos tiene las siguientes características:

- Estimación del modelo: Se usan técnicas probabilísticas para inferir los parámetros del modelo con base en las observaciones históricas. En particular se usa la técnica de *Máxima Verosimilitud*<sup>4</sup> o MLE (Maximum Likelihood Estimation) por sus siglas en inglés.
- Cuantificación de la Incertidumbre: Esto cobra especial importancia al modelar valores extremos, ya que pequeños cambios en el modelo pueden ser magnificados en la extrapolación. Dicha incertidumbre se puede obtener mediante la función de probabilidad.
- Diagnóstico del Modelo: El modelo debe ser capaz de representar adecuadamente los valores extremos observados.
- Máximo uso de información: Es posible disminuir la incertidumbre al trabajar con métodos que usan más datos, y no sólo unos cuantos, como es el caso del método de máximos de bloques.

Existen dos familias básicas de modelos en la *Teoría de Valores Extremos*. Una está formada por los *Modelos Asintóticos* como la *Distribución General de Valores Extremos*, en la que se usa el *Método de Bloques* para tomar valores máximos o mínimos en un cierto periodo de tiempo y a estos valores analizarlos ajustándoles la *Distribución General de Valores Extremos*. Otra familia son los *Modelos de Umbral*, donde el procedimiento consiste en establecer un umbral usando todos los datos disponibles y posteriormente analizar todos los valores que se encuentren por encima de este umbral bajo distintos métodos como el ajuste de la *Distribución Pareto Generalizada* y *Procesos Puntuales*.

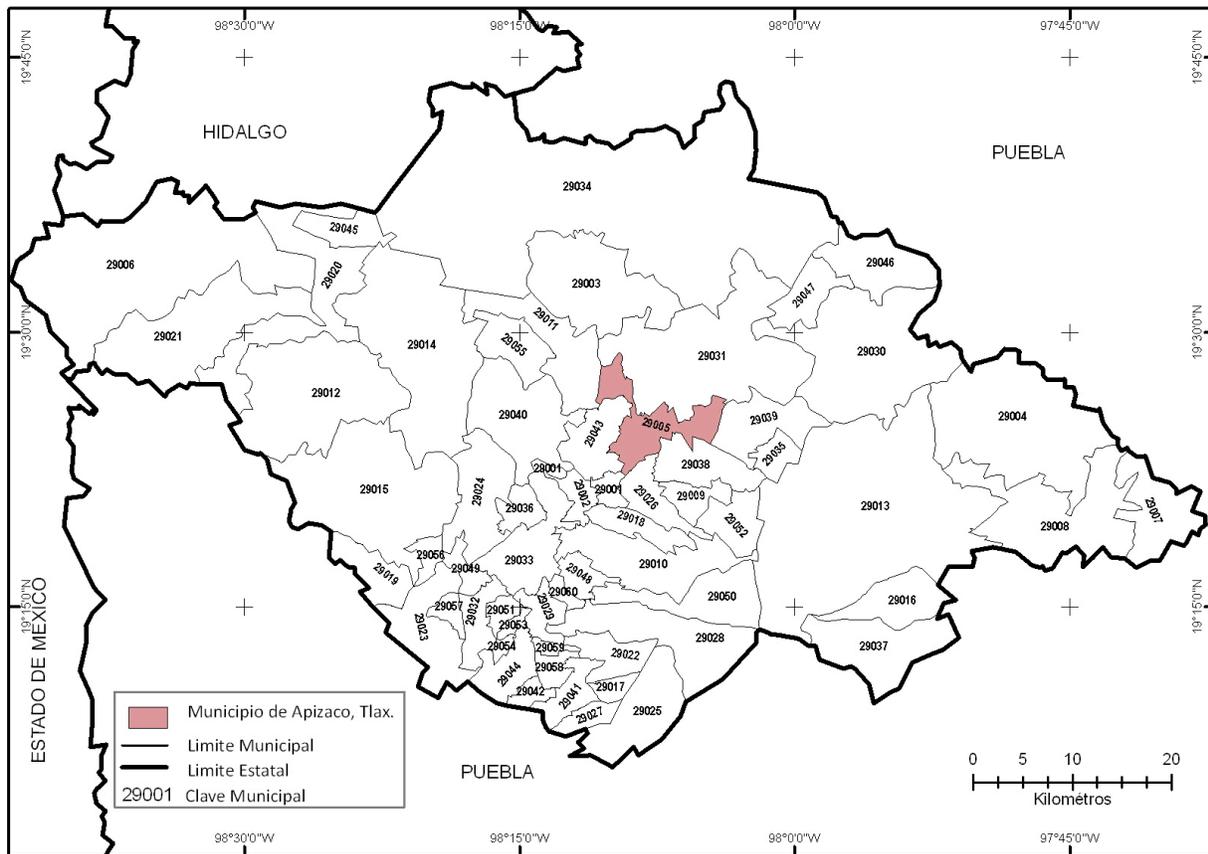
## Región de Estudio

El estado de Tlaxcala (Figura 2-1) se localiza en la región central oriental del país sobre el eje Neovolcánico por encima de los 2000 metros de altitud; entre los 97° 37' 07" y los 98° 42' 51" de longitud oeste, y entre los 19° 05' 43" y los 19° 44' 07" de latitud norte. El estado de Tlaxcala tiene una superficie de 4060.92 Km<sup>2</sup> y representa el 0.2% de la superficie del país. Tlaxcala colinda al norte con los estados de Hidalgo y Puebla, al este y sur con Puebla; al oeste con Puebla, México e Hidalgo. Tlaxcala se divide en 60 municipios entre ellos Apizaco. (INEGI 2005)

El municipio de Apizaco se localiza en la zona central del estado de Tlaxcala, en 19° 25' latitud norte y 98° 08' longitud oeste, a una altitud de 2408 metros sobre el nivel del mar. Ocupa una superficie de 56.83 kilómetros cuadrados que corresponden al 1.4% de la superficie del estado.

---

<sup>4</sup> Máxima Verosimilitud: Es un método usado para ajustar modelos estadísticos a datos, determinando los estimadores de los parámetros del modelo que mejor se ajusten a éste.



Clave □ Municipio		Clave □ Municipio		Clave □ Municipio	
29001	Amamax de Guerrero	29021	Nanacamilpa de Mariano Arista	29041	Papalotla de Xicohtencatl
29001	Amamax de Guerrero	29022	Acuamanala de Miguel Hidalgo	29042	Xicohtzinco
29002	Apetatitlan de Antonio Carvajal	29023	Nativitas	29043	Yauhquemecan
29003	Atlangatepec	29024	Panotla	29044	Zacatelco
29004	Altzayanca	29025	San Pablo del Monte	29045	Benito Juarez
29005	Apizaco	29026	Santa Cruz Tlaxcala	29046	Emiliano Zapata
29006	Calpulalpan	29027	Tenancingo	29047	Lazaro Cardenas
29007	El Carmen Tequexquitta	29028	Teolochoico	29048	La Magdalena Tlaltelulco
29008	Cuapiaxtla	29029	Tepeyanco	29049	San Damian Texoloc
29009	Cuaxomulco	29030	Terrenate	29050	San Francisco Tetlanohcan
29010	Chiautempan	29031	Tetla de la Solidaridad	29051	San Jeronimo Zacualpan
29011	MuCoz de Domingo Arenas	29032	Tetlatlahuca	29052	San Jose Teacalco
29012	EspaCita	29033	Tlaxcala	29053	San Juan Huactzinco
29013	Huamantla	29034	Tlaxco	29054	San Lorenzo Axocomanitla
29014	Hueyotlipan	29035	Tocatlan	29055	San Lucas Tecopilco
29015	Ixtacuixtla de Mariano Matamoros	29036	Totolac	29056	Santa Ana Nopalucan
29016	Ixtenco	29037	Zitlaltepec de Trinidad Sanchez Santos	29057	Santa Apolonia Teacalco
29017	Mazatecochco de Jose Maria Morelos	29038	Tzompantepec	29058	Santa Catarina Ayometla
29018	Contla de Juan Cuamatzi	29039	Xaloztoc	29059	Santa Cruz Quilehltla
29019	Tepetitla de Lardizabal	29040	Xaltocan	29060	Santa Isabel Xiloxxtla
29020	Sanctorum de Lazaro Cardenas				

Figura 2-I. Mapa del estado de Tlaxcala. Fuente, División municipal digitalizada, (INEGI 2007).

Limita al norte con los municipios de Tetla de la Solidaridad y Domingo Arenas, al sur con Tzompantepec y Santa Cruz Tlaxcala, al oriente con Xalostoc y Tzompantepec y al poniente con Yauhquemecan y Amaxac de Guerrero. Se divide en 11 localidades, de las cuales las más importantes son Apizaco, que es la cabecera municipal, Santa María Texcala y San Luis Apizaquito.

Su clima es templado subhúmedo con lluvias en los meses de mayo a septiembre y parte de octubre, los meses más calurosos son mayo y junio; la dirección de los vientos en general es de norte a sur. La temperatura media anual es de 13.9°C, la temperatura mínima promedio es de 4.7°C y la máxima promedio es 22.6°C.

La precipitación es una variable importante en México y en esta región, dada la gran cantidad de cultivos de temporal que dependen de ésta. En Tlaxcala, 92% de las unidades de producción agrícola son cultivos de temporal (INEGI 1996) y 89% de la superficie cultivable del estado se dedica a cultivos de temporal, esto hace que la producción agrícola sea muy sensible a los eventos climáticos especialmente a los el cual se agrava al presentarse eventos extremos de precipitación. En trabajos previos (Conde, et al 2006) se menciona que el grupo más vulnerable dentro del sector agrícola es precisamente el de los productores de maíz de temporal. El maíz necesita ciertas condiciones de precipitación, temperatura y nutrientes dependiendo de la etapa de crecimiento de la planta; cuando estas condiciones no se dan en la cantidad o momento adecuados la cosecha se puede perder. Un problema importante en Tlaxcala es la degradación de suelos, siendo éste el estado con el suelo menos fértil de todo el país (SEMARNAT 1995). Este problema se agrava al ocurrir eventos de precipitación intensa que promueven el proceso de lixiviación, que “lava” los pocos nutrientes que se encuentran en el suelo, y aumentan la erosión hídrica.

Como hemos mencionado, no sólo la cantidad de precipitación es un factor importante, también lo es el momento en el que se presentan durante el crecimiento de la planta. Tal es el caso de la canícula<sup>5</sup>, donde canículas muy intensas son perjudiciales para la planta ya que éste es un periodo importante en el desarrollo del cultivo.

Los cultivos de temporal dependen en gran medida de los factores climáticos como la precipitación, esto junto con las condiciones de marginalidad en el sector rural, son factores que amplifican el efecto negativo de los eventos extremos (Saldaña 2007). Esto es, los efectos negativos de los eventos extremos no sólo se concentran en pérdida de infraestructura o de cosechas, también tienen efectos socioeconómicos a largo plazo. La pérdida de cosechas o infraestructura, aunado a las malas condiciones del sector agrícola en el país, generan grandes migraciones hacia zonas urbanas o más aún hacia otros países, lo

---

<sup>5</sup> Canícula: Mínimo relativo que se en la temporada de lluvias de verano en México. También conocido como sequía de medio verano y sequía intraestival (Magaña 1999).

que a su vez aumenta las tensiones políticas y disminuye la calidad de vida en las zonas urbanas. Al mismo tiempo la pobreza se puede incrementar en la zona afectada al desacelerarse o incluso hacer negativo el crecimiento económico de la región.

Ya existe una serie de condiciones socioeconómicas que aumentan la vulnerabilidad de los agricultores por tanto es muy importante disminuir la vulnerabilidad debida a factores climáticos, en este caso eventos extremos de precipitación.

Finalmente, aunque estos cambios en la precipitación pueden generar efectos adversos, se pueden también abrir oportunidades hacia un desarrollo sustentable que no sólo disminuya la vulnerabilidad frente a factores climáticos, si no que también lo haga frente a factores socioeconómicos. Esto se puede lograr mediante planes integrales que incluyan a todos los actores y que contemple todos los factores tanto climáticos, como socioeconómicos, tal como se ha hecho en estudios previos (Conde 2006).

# II-Método y Resultados

## Climatología

Por su ubicación el clima en México está determinado tanto por sistemas tropicales como por sistemas de latitudes medias y por las interacciones entre ambos. Cuando una masa de aire del norte se encuentra con una masa de aire tropical se forma un frente y será cálido o frío dependiendo de cuál de las masas de aire tenga una mayor velocidad. A nivel regional el clima en México está determinado por las condiciones de frontera, tales como topografía, uso del suelo y presencia del mar, entre otros (Magaña et al 1999). La precipitación en México depende principalmente de las siguientes estructuras meteorológicas: La Zona Inter-Tropical de Convergencia<sup>1</sup> (ZITC), el Monzón Mexicano, las Ondas del Este y las Tormentas Tropicales o Huracanes en ambos océanos. El comportamiento de la ZITC influye en gran medida sobre la precipitación en verano, de su posición y densidad depende la ocurrencia de lluvias intensas o sequías. El Monzón Mexicano, por su parte, afecta de manera importante las lluvias de verano al presentar entradas de humedad en la alta y media troposfera desde el Golfo de México y entradas de humedad provenientes del Pacífico tropical y del Golfo de California en los niveles bajos de la troposfera. Las Ondas del Este son sistemas de baja presión que se forman en los océanos tropicales y que se encuentran inmersos en los vientos alisios, en las zonas de máxima temperatura del mar (entre las latitudes 5°-20° norte y sur); viajan en dirección oeste y pueden evolucionar hasta transformarse en una depresión tropical o incluso en un huracán (Reyes Coca 2002).

El fenómeno de El Niño Oscilación Sur (ENOS) juega un papel importante al modular gran parte del clima en el país e incluso puede disparar eventos extremos. En el caso de la precipitación, puede hacer que ésta aumente en invierno y disminuya en verano. Tlaxcala es el estado más pequeño del país, sin embargo su topografía es compleja. Las lluvias de verano dependen en gran medida de la temporada de huracanes en el océano Atlántico.

---

<sup>1</sup> Zona Inter-Tropical de Convergencia (ZITC): zona donde convergen los vientos alisios de ambos hemisferios.

En la figura 3-II se muestra la climatología de Apizaco para el periodo 1961-1990.

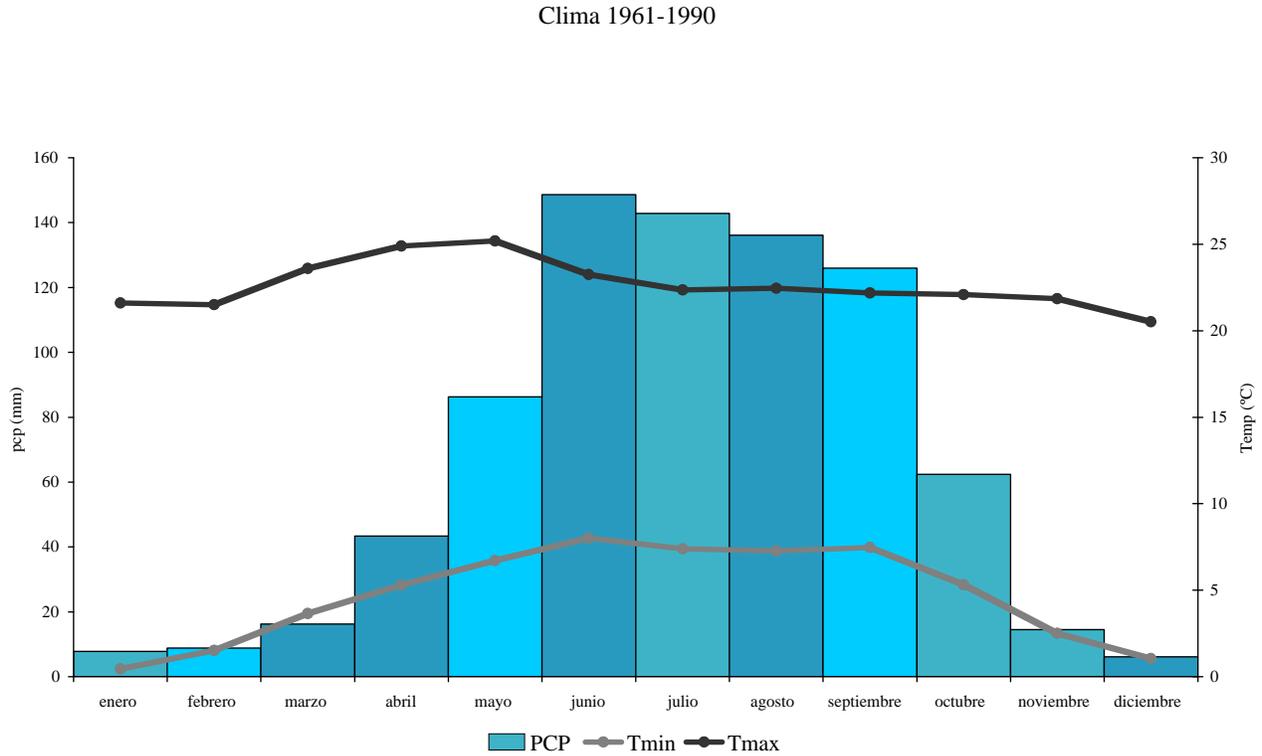


Figura 3-II. Climatología Normal (datos mensuales de precipitación acumulada, temperaturas máximas y mínimas promedio), Apizaco Tlaxcala 1961-1990.

En la figura 3-II se aprecia el ciclo estacional de la precipitación. La cantidad de precipitación (PCP) acumulada en los meses húmedos es mucho mayor, casi en un orden de magnitud, respecto de los meses secos. Lo anterior sugiere que cada mes podría tener una distribución diferente, lo cual se aprecia mejor en la figura 5-II.

Esta combinación de variables, entre otras cosas, permite que la producción agrícola predominante en la región sea el maíz (SEGOB 1988). Si se llega a modificar la distribución de estas variables bajo cambio climático, se podría afectar en algún grado la producción agrícola u otros sectores.

### Distribución de la Precipitación

La precipitación generalmente tiene una distribución tal que presenta muchos días con poca precipitación y muy pocos días con mucha precipitación. Lo anterior se aprecia en la figura 4-II que muestra el histograma de la dis-

tribución de los días con precipitación mayor a cero en todo el periodo, desde 1940 hasta 2004. Se observa que el evento de mayor precipitación en un día corresponde a 90 mm.

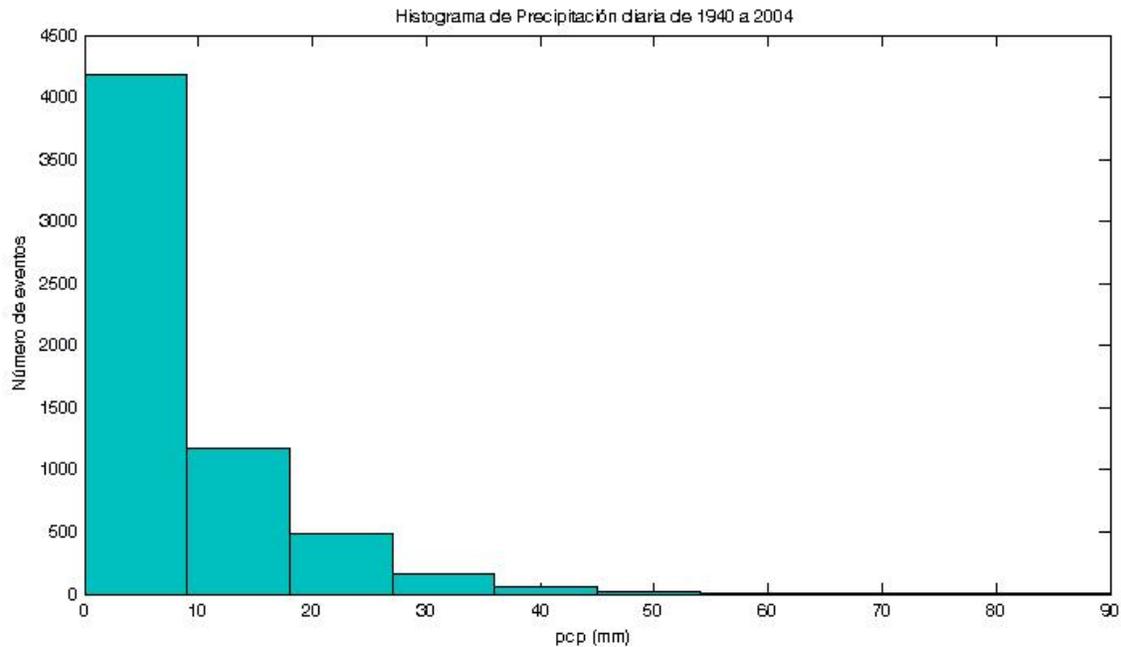


Figura 4-II. Histograma de días con precipitación.

La figura 5-II muestra las distribuciones de precipitación mensual, usando diagramas de caja<sup>2</sup> (Wilks 1995) En ella se observa que las distribuciones de los meses húmedos se encuentran en rangos muy distintos a los de los meses secos. Así mismo se aprecia una disminución en las medianas de las distribuciones de precipitación en los meses de julio y agosto. Mientras que en septiembre la mediana continua disminuyendo, sin embargo el rango intercuartílico se amplía respecto a los meses de julio y agosto, esto significa que la distribución se ha ensanchado, o en otras palabras presenta mayor variabilidad.

<sup>2</sup> Ver apéndice A.8.

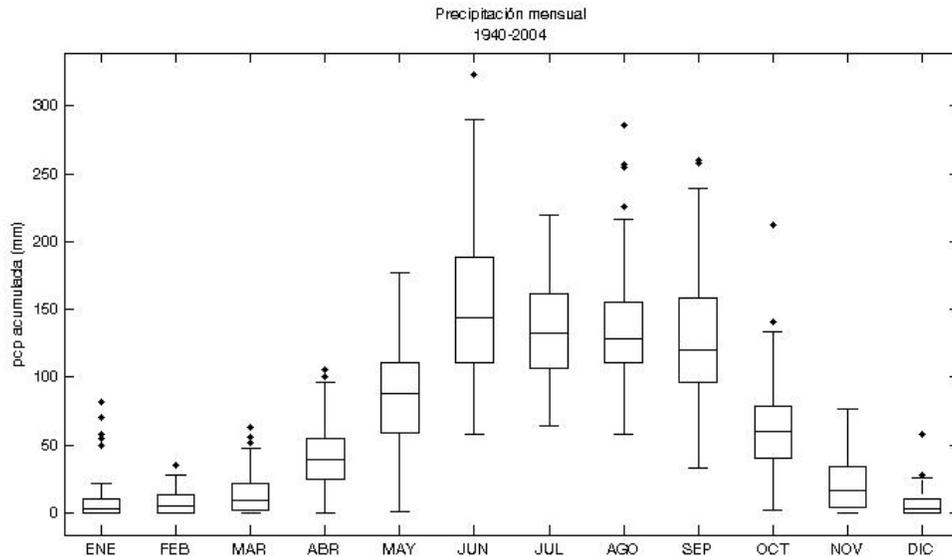


Figura 5-II. Distribuciones de la precipitación mensual acumulada.

Debido a la diferencia en rangos que se observa entre las distribuciones de los meses húmedos y de los secos, el análisis se llevará a cabo tanto separando meses húmedos de secos como introduciendo el ciclo anual de la precipitación. La razón para analizar por separado la serie radica en el hecho de que es necesario que la serie de tiempo sea estacionaria, esto es, que la varianza sea constante en el tiempo, lo cual puede no ser cierto al estar presente un ciclo.

En la figura 6-II se muestran los diagramas de caja de las distribuciones quincenales. Esta gráfica es muy similar a la figura 5-II, sólo que en ésta se aprecia mucho mejor el periodo de canícula o sequía de medio-verano (Magaña et al, 1999), donde la mediana de la primera quincena de julio y hasta la primera de agosto disminuye, luego aumenta ligeramente en la segunda quincena de agosto para volver a disminuir de ahí en adelante. El comportamiento del rango intercuartílico de las distribuciones es ligeramente diferente, disminuye durante julio y agosto, aumenta significativamente en septiembre, y disminuye radicalmente desde octubre hasta diciembre, sin embargo es en estos meses también donde se observan algunos valores atípicos.

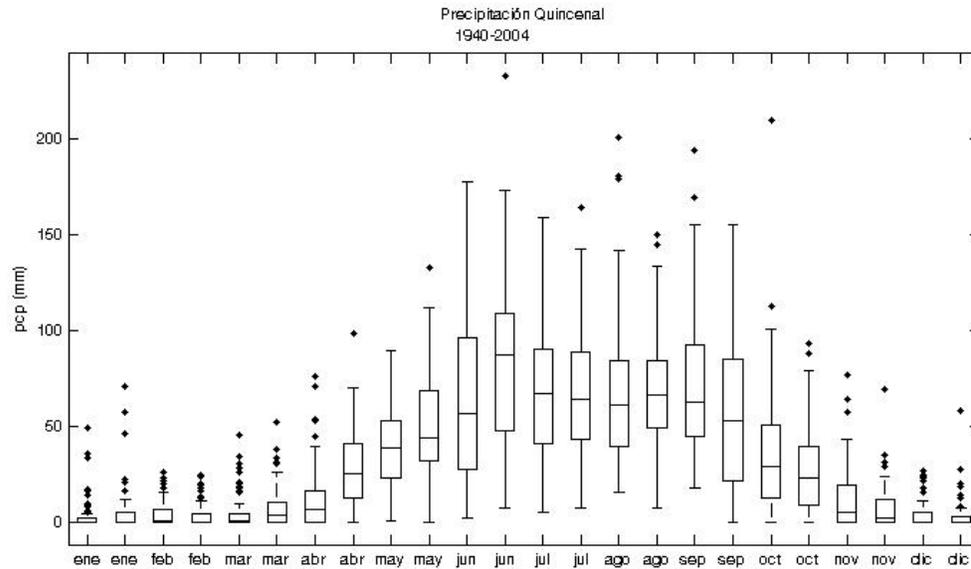
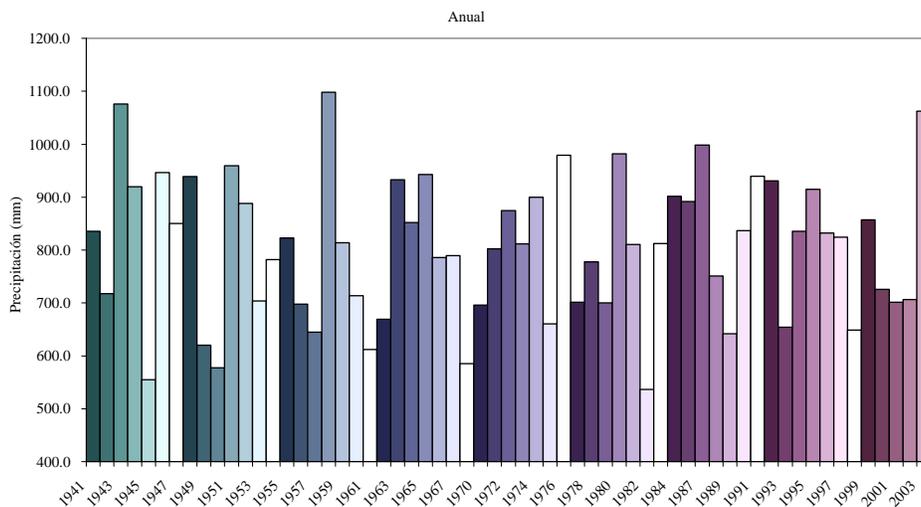


Figura 6-II. Distribuciones de la precipitación quincenal acumulada.

La figura 7-II muestra la precipitación anual desde 1940 hasta 2004. En ésta figura no se aprecia una tendencia



clara en cuanto a la cantidad total de precipitación anual.

Figura 7-II. Precipitación anual acumulada.

Finalmente en la figura 8-II se observa la anomalía quincenal establecida entre la normal del clima 61-90 y el periodo después de 1975, ya que como se menciona en IPCC (2007) es posible observar un cambio en el clima a mediados de esa década.

Lo que se aprecia aquí es un aumento máximo en la precipitación para las quincenas correspondientes a la primera de junio, segunda de agosto, septiembre y primera de octubre, un aumento más ligero se da en la segunda de diciembre, la segunda de enero y la primera de febrero. Es importante notar que también se han dado disminuciones, siendo las más importantes en la segunda mitad de junio, julio y la primera mitad de agosto. Esto último es importante ya que, como se vio en la figura 6-II la disminución en la precipitación se da justo en estas semanas, por tanto si esta tendencia continúa en el futuro se podrían tener episodios de canícula más intensos.

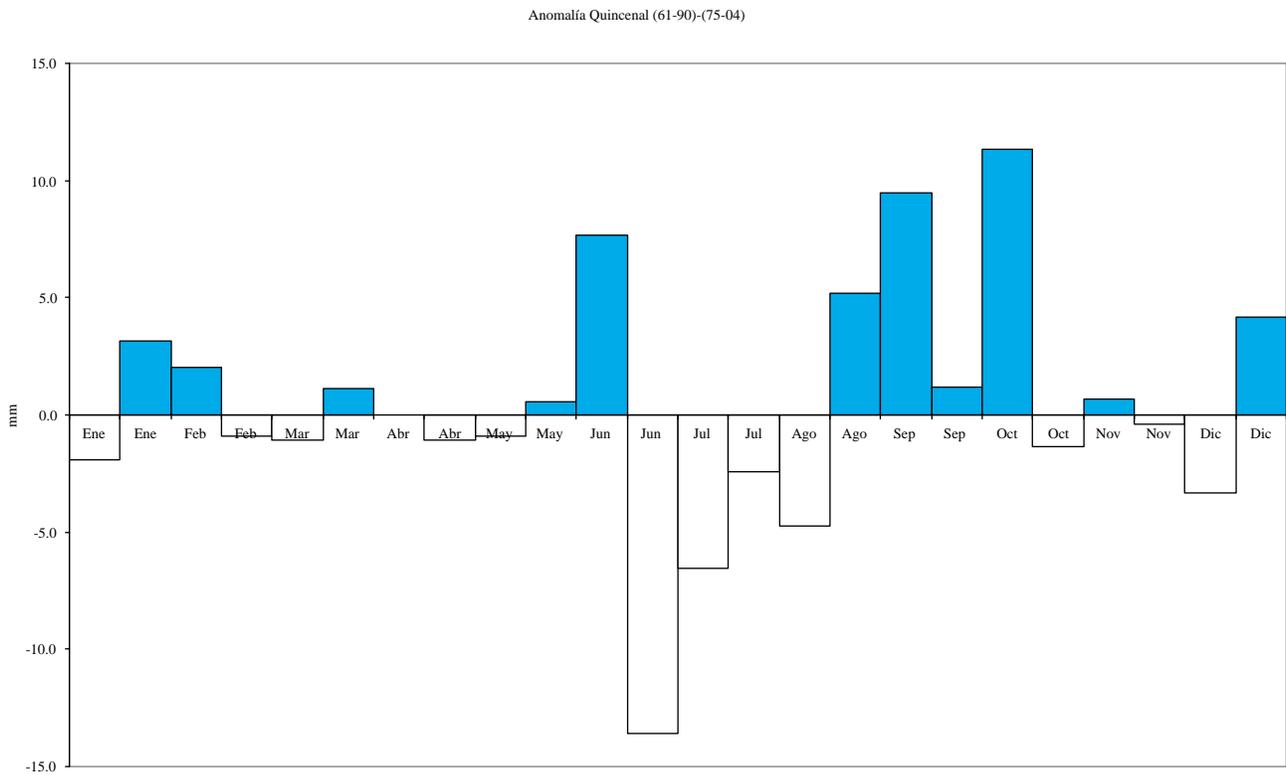


Figura 8-II. Anomalías quincenales para el periodo 1975-2004 con respecto al periodo normal 1961-1990.

## Datos

Se usaron los datos de precipitación diaria de la estación Apizaco (00029002) localizada en 19°25'41" de latitud Norte y 98°07'51" de longitud oeste, a una altura de 2410 m snm. Los datos abarcan desde septiembre de 1940 hasta abril de 2004. Los huecos en la serie se encuentran en: diciembre 1940 a abril de 1941 (122 días); mayo de 1941 (32 días); 8 de enero de 1942; 30 de abril de 1945; 2 de enero de 1948; 1 de febrero de 1951; 1 de noviembre de 1952; enero de 1953 a marzo de 1953 (59 días); 31 julio de 1954; marzo de 1968 a junio de 1968 (92 días); noviembre de 1976 a enero de 1977 (61 días); julio de 1986 a enero de 1987 (184 días); 31 de diciembre de 1994. Hacen un total de 557 días de 23253, es decir sólo un 2.4%, por lo tanto la serie está completa en un 97.6%. Los datos faltantes fueron sustituidos por el valor 0.0 mm.

## **Umbral**

Es necesario establecer un umbral a partir del cual un evento será extremo, la determinación de éste dependerá del tipo de información que deseemos obtener de los datos y del problema en particular con el que estemos trabajando. En ocasiones el problema es tan concreto como la pérdida de una cosecha o una inundación a partir de cierta cantidad de mm de precipitación que depende del problema en sí, si tal es el caso, esta cantidad determinada de mm de precipitación es la que se usa como umbral. En este trabajo se usó un umbral estadístico que corresponde al percentil 90 de la distribución de días húmedos, siguiendo la sugerencia del IPCC. El percentil 90 se estableció para la distribución de todos los días húmedos, es decir los datos mayores o iguales a 0.1 mm de precipitación diaria. El mismo percentil 90 se puede determinar no para toda la distribución, sino para las distribuciones de datos diarios separados por mes o por estación. La decisión de separar o no los datos dependerá, una vez más, de la información que se desee obtener.

## **El Niño Oscilación Sur**

Como se mencionó anteriormente El Niño Oscilación Sur (ENOS) juega un papel importante en la cantidad de precipitación, haciendo que, en general, ésta aumente en invierno y disminuya en verano. En la figura 9-II se muestra la comparación de las máximas mensuales con el índice Multivariado de El Niño Oscilación Sur (MEI, por sus siglas en inglés). Los valores negativos de dicho índice corresponden a La Niña, los positivos a El Niño y los valores cercanos a cero corresponden a años neutros. Claramente se observa que las máximas se acumulan en la zona central, es decir en los años neutros. Lo que significa que los eventos extremos en este lugar, incluso en invierno, se presentan en años neutros; en otras palabras los eventos extremos de precipitación en Apizaco no tienen mucha relación con el ENOS.

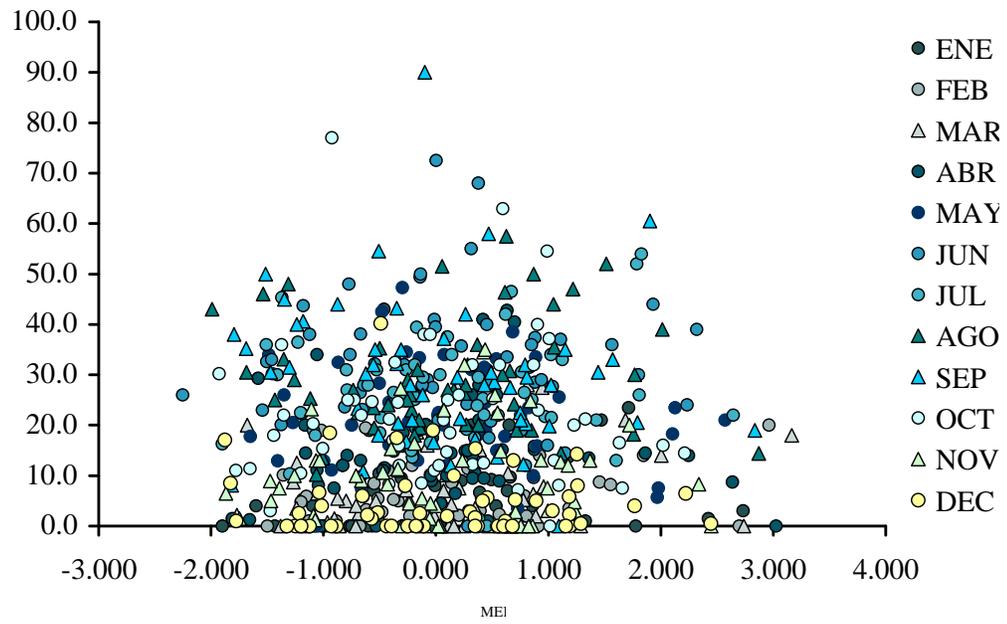


Figura 9-II. Máxima mensual comparada con MEI.

# Índices de Eventos Extremos

## Índices

Es importante tener varios criterios basados en el clima para elegir correctamente los indicadores de eventos extremos. Godess (2001) proponen indicadores tanto para temperatura como para precipitación, en la tabla 2 se muestran los segundos.

**Tabla 2: Índices calculados de datos de lluvias diarias**

- Índice simple de intensidad diaria (precipitación total/número de días con precipitación).
- Magnitud de percentiles 90 ó 95.
- Frecuencia de excedencia de los percentiles 90 ó 95 (número de días).
- Porcentaje de lluvia caída en días con lluvia sobre el percentil 90 ó 95.
- Porcentaje de la región con precipitación menor al percentil 5 ó 10.
- Duración máxima húmedo/seco.
- Número de días con lluvia menor o igual a 10 mm.
- Máximos 5 días de precipitación.
- Magnitud de 2, 5, 10 y 20 periodos de retorno.

En este trabajo los índices a usar de la tabla anterior son:

- Índice simple de intensidad diaria (lluvia total/ número de días de lluvia).

- Magnitud del percentil 90 y frecuencia de excedencia de éste (número de días).
- Duración máxima de húmedo/seco.
- Máximos 5 días consecutivos con precipitación.

Como se mencionó anteriormente para poder determinar cuándo un evento de precipitación se convierte en extremo es necesario establecer características locales de precipitación diaria. El criterio para determinar las precipitaciones extremas en este trabajo es: Los eventos de precipitación en 24 horas (Cavazos 2004) que estén por encima del percentil 90 (IPCC 2001) de la población de días con precipitación diaria mayor o igual a 0.1 mm, esta cantidad corresponde a la mínima resolución del programa.

Para llevar a cabo el cálculo de índices se usó una aplicación del software libre R<sup>3</sup> que se especializa en el cálculo de índices de eventos extremos, dicha herramienta tiene por nombre *RClimdex*<sup>4</sup>. Usa como base teórica las funciones propuestas por Coles (2001).

La aplicación calcula los siguientes índices para precipitación:

- RX1day (Cantidad máxima de precipitación en un día): Máxima mensual de precipitación en un día. [mm]
- RX5day (Cantidad máxima de precipitación en 5 días): Máxima mensual en 5 días consecutivos de precipitación. [mm]
- SDII (Índice simple de Intensidad diaria): Precipitación anual total dividida entre el número de días húmedos, definidos como días con precipitación mayor o igual a 1 mm, en un año. [mm/día]
- R10 (Número de días con precipitación intensa): Número de días en un año en que la precipitación fue mayor o igual a 10 mm. [día]
- R20 (Número de días con precipitación muy intensa): Número de días en un año en que la precipitación fue mayor o igual a 20 mm. [día]
- Rnn (Número de días sobre nn mm): Número de días en un año en que la precipitación fue mayor o igual a nn mm, donde nn es un valor introducido por el usuario. [día]
- CDD (Días secos consecutivos): Número máximo de días consecutivos con precipitación menor a 1 mm. [día]
- CWD (Días húmedos consecutivos): Número máximo de días consecutivos con precipitación mayor o igual a 1 mm. [día]
- R95p (Días muy húmedos): Precipitación anual total en que ésta es mayor o igual al percentil 95. [mm]
- R99p (Días extremadamente húmedos): Precipitación anual total en que ésta es mayor o igual al percentil 99. [mm]

---

<sup>3</sup> Ver apéndice A.2.

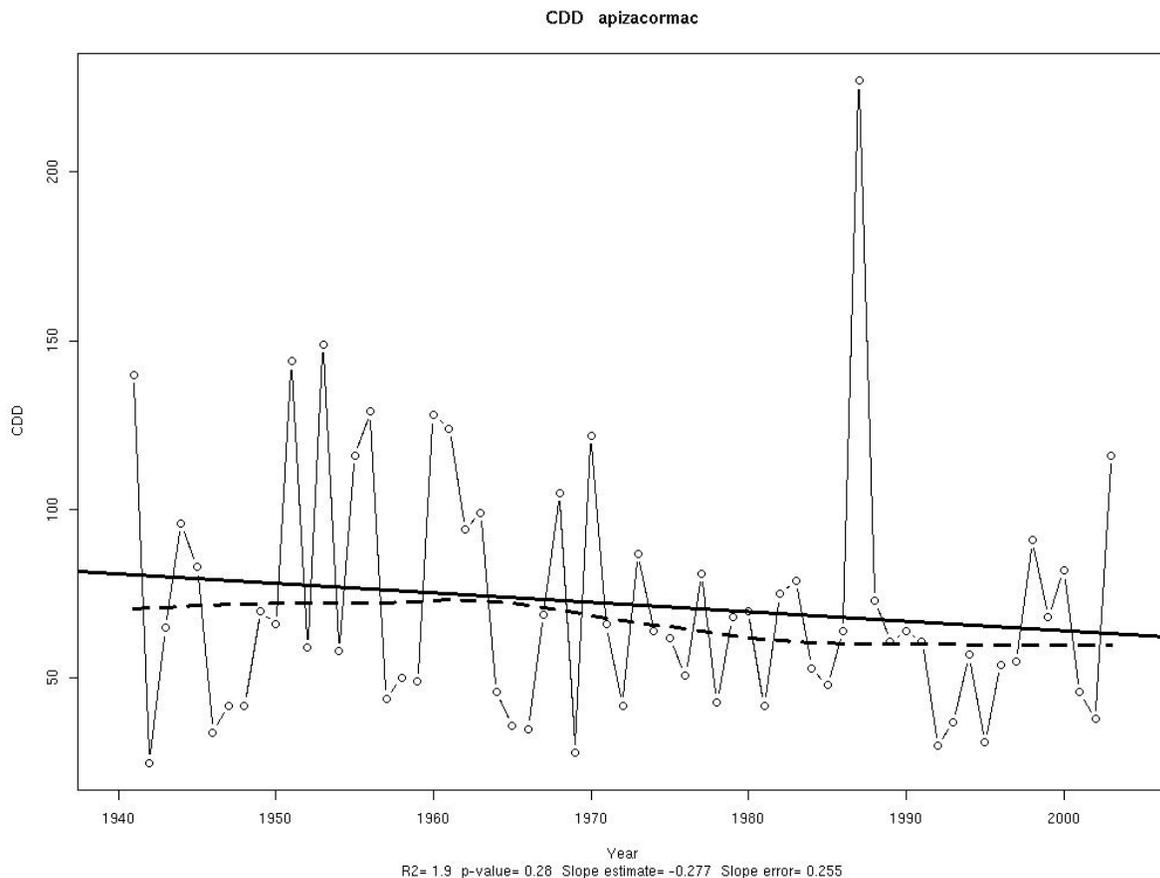
<sup>4</sup> Ver apéndice A.3.

- PRCPTOT (Precipitación total anual en los días húmedos): Precipitación anual total en días donde ésta es mayor o igual a 1mm. [mm]

Como mencionamos anteriormente, para poder determinar cuándo un evento de precipitación se convierte en extremo es necesario establecer características locales de precipitación diaria. El criterio para determinar las precipitaciones extremas en este trabajo es: Los eventos de precipitación en 24 horas (Cavazos 2004) que estén por encima del percentil 90 (IPCC 2001) de la población de días con precipitación diaria mayor a cero.

## Resultados

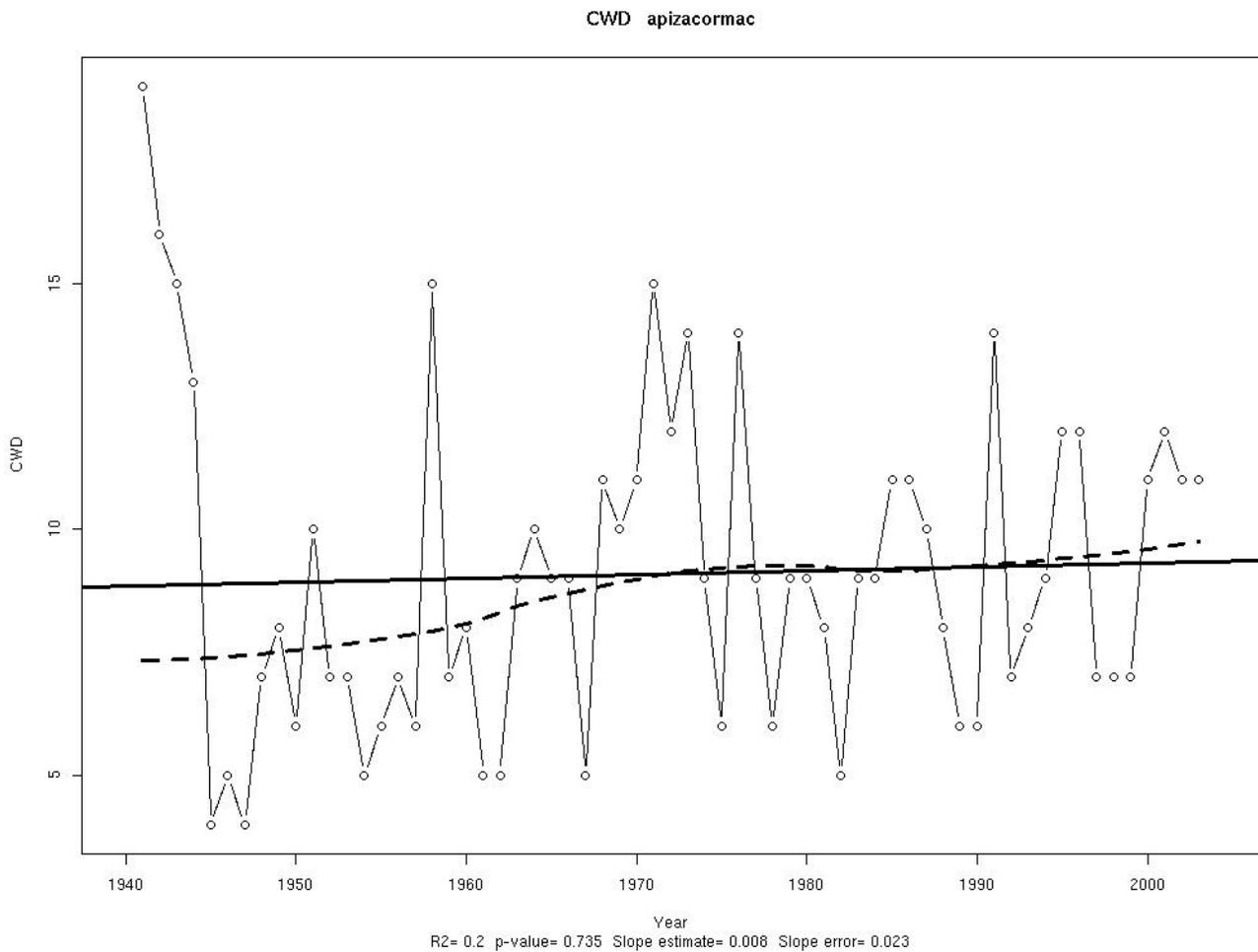
A continuación se muestran las gráficas que se obtienen al calcular los índices en *RClindex*. La línea sólida es la



tendencia lineal y la punteada es una tendencia calculada mediante medias móviles.

Figura 10-II. Número máximo de días consecutivos con precipitación menor a 1 mm.

En la figura 10-II se puede ver que la tendencia lineal, aunque pequeña, es significativa, lo que se traduce en una disminución del número máximo de días consecutivos con muy poca precipitación. Esto podría tener como con-

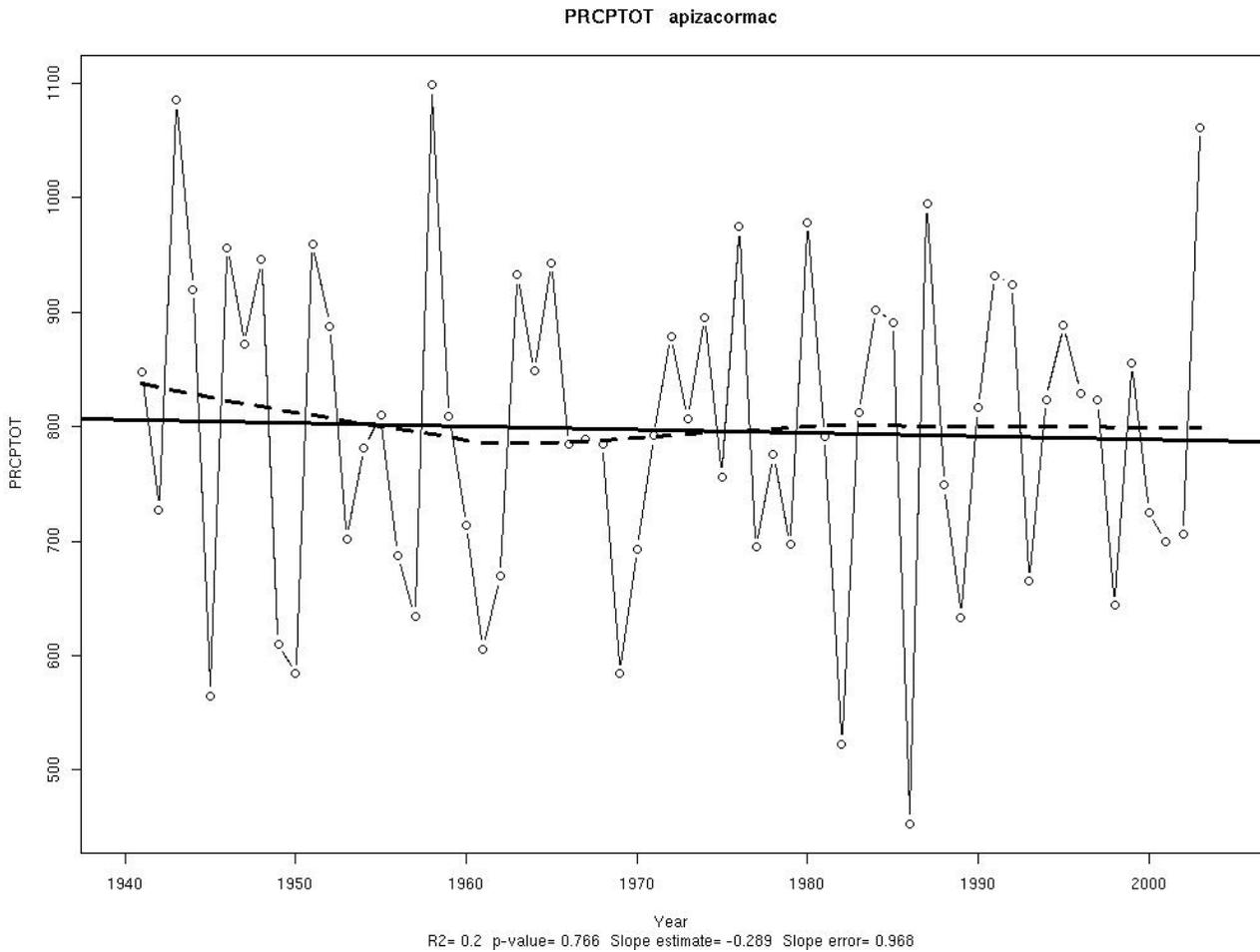


secuencia que ahora llueva más, ya sea en el mismo número de días o bien en menos días.

Figura 11-II. Número máximo de días consecutivos con precipitación mayor o igual a 1 mm.

En la figura 11-II se aprecia que el número máximo de días consecutivos con precipitación mayor a 1 mm ha aumentado, de acuerdo con la figura 10-II. La tendencia lineal no es significativa, sin embargo la tendencia obtenida mediante medias móviles presenta un quiebre a partir de 1951, posterior a este quiebre la tendencia sería

ligeramente positiva. Así entonces ahora podría no haber cambiado la cantidad de precipitación pero sí haber



aumentado el número de días consecutivos con precipitación.

Figura 12-II. Precipitación anual total en días donde ésta es mayor o igual a 1 mm.

La precipitación anual total (figura 12-II) parece haber disminuido, aunque la tendencia no es significativa. Es interesante ver que en la tendencia punteada disminuye con una pendiente mayor hasta inicios de los 60, luego aumentó ligeramente hasta los 70 para mantenerse casi estática desde entonces. Si la tendencia es cierta estaría de acuerdo con los reportes del IPCC de otras regiones donde la precipitación total ha disminuido o se ha mantenido, mientras que el número de episodios con precipitación intensa ha aumentado.

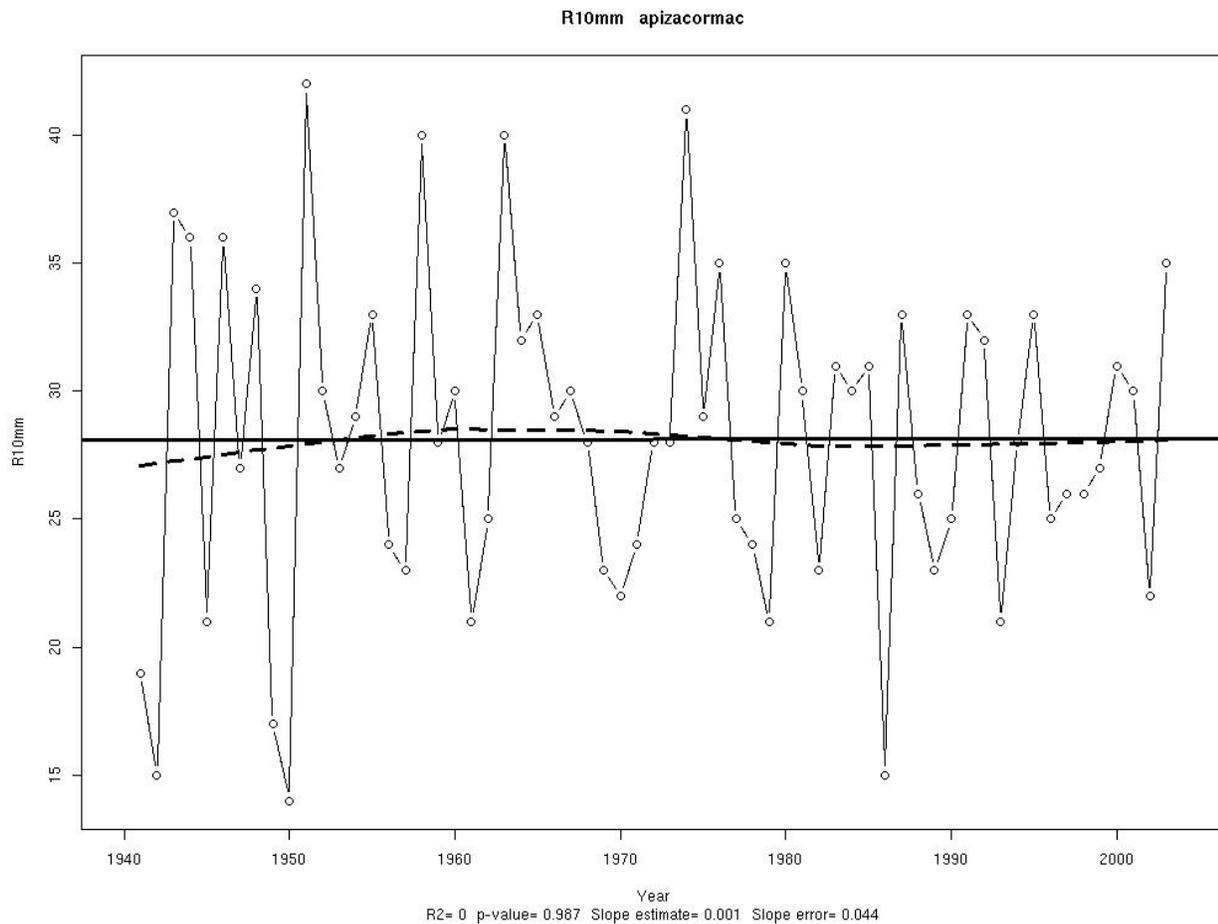


Figura 13-II. Número de días en un año en que la precipitación fue mayor o igual a 10 mm.

En el caso de la figura 13-II ninguna de las tendencias es significativa, lo que quiere decir que los eventos con precipitación mayor a 10 mm han permanecido sin cambios.

R 20 mm apizacormac

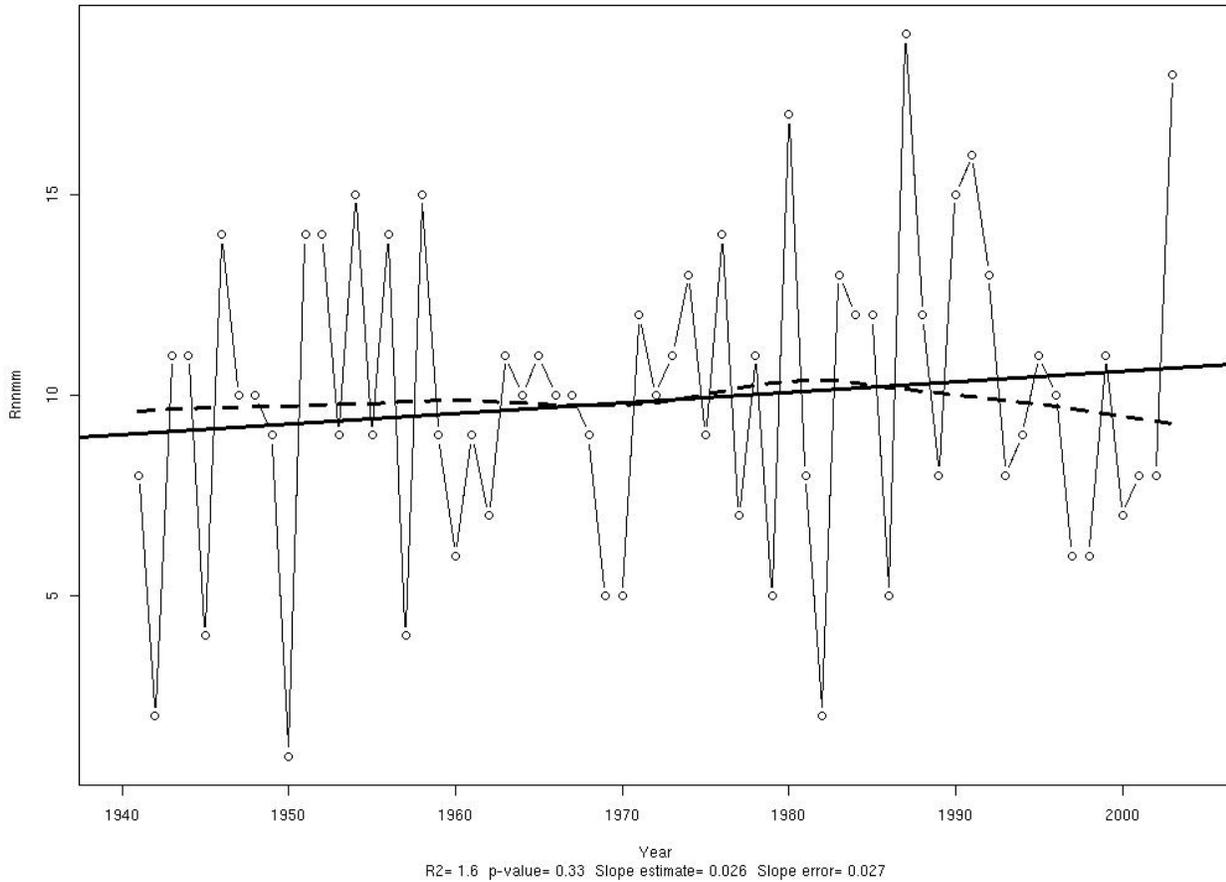


Figura 14-II. Número de días en un año en que la precipitación fue mayor o igual a 20 mm.

El número de días con precipitación mayor a 20 mm (figura 14-II), donde 20 mm corresponde al umbral de eventos extremos de esta distribución, presenta tendencia lineal significativa, así entonces el número de días con precipitación intensa estaría aumentando. Aunque la tendencia obtenida mediante medias móviles no se encontró ningún quiebre significativo en la serie.

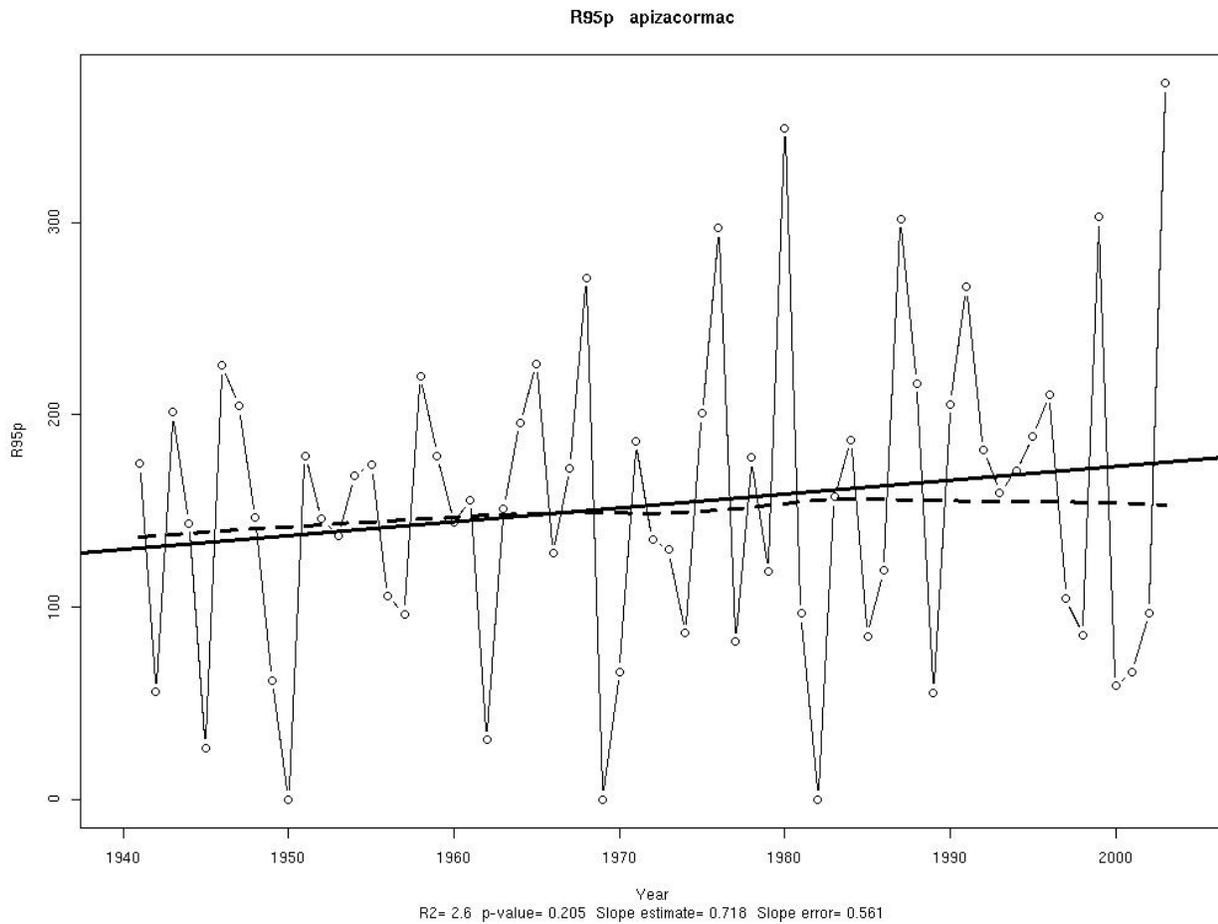


Figura 15-II. Precipitación anual total en que ésta es mayor o igual al percentil 95 de la distribución.

La precipitación anual total mayor al 95% presenta una tendencia positiva significativa. Esto quiere decir que los eventos de precipitación anual total están aumentando, esto puede ser importante en el manejo de presas de la zona de estudio, y en la pérdida de fertilidad de los suelos como se mencionó anteriormente.

R99p apizacormac

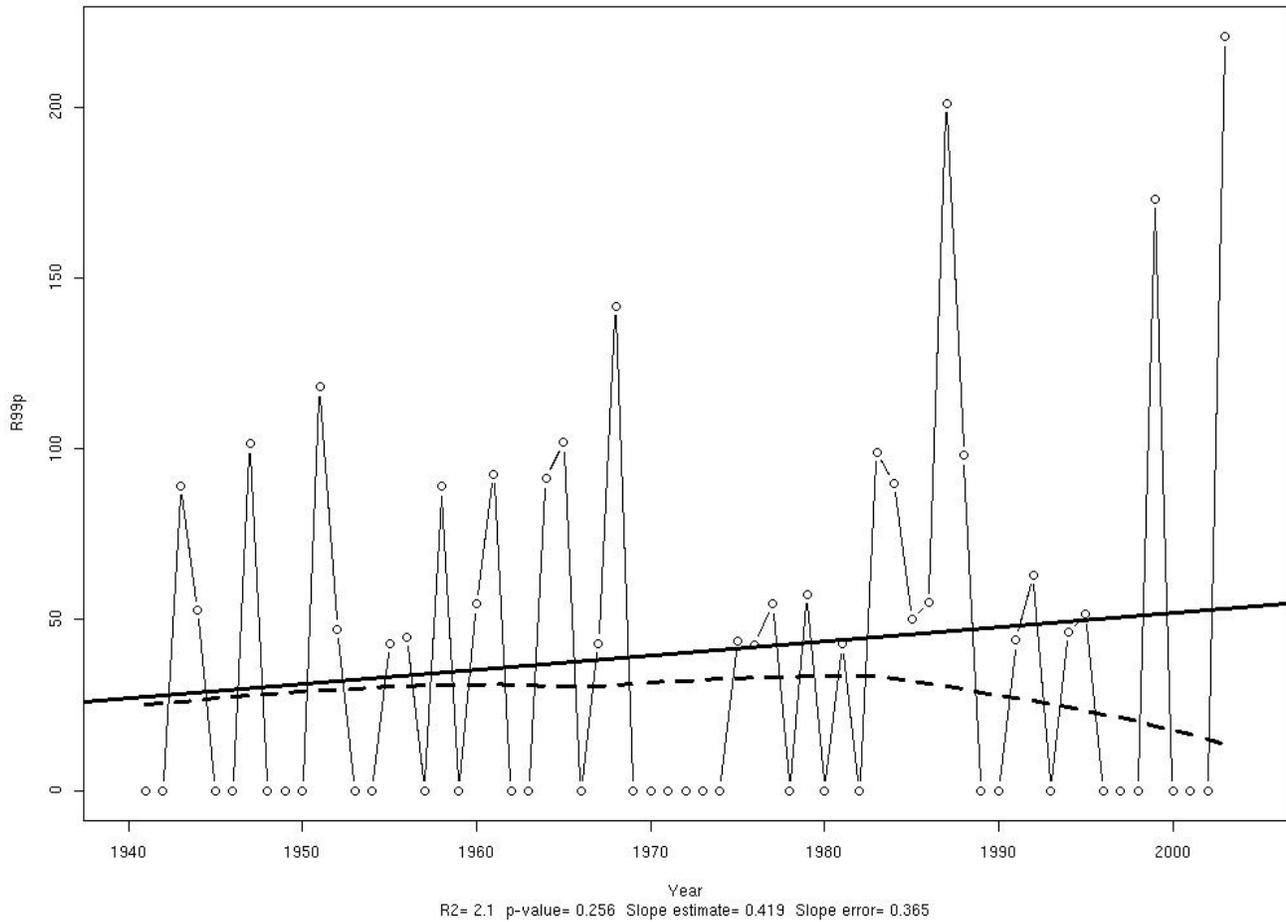


Figura 16-II. Precipitación anual total en que ésta es mayor o igual al percentil 99.

En el caso de la precipitación anual (figura 16-II) la tendencia positiva también es significativa aunque de una mayor pendiente.

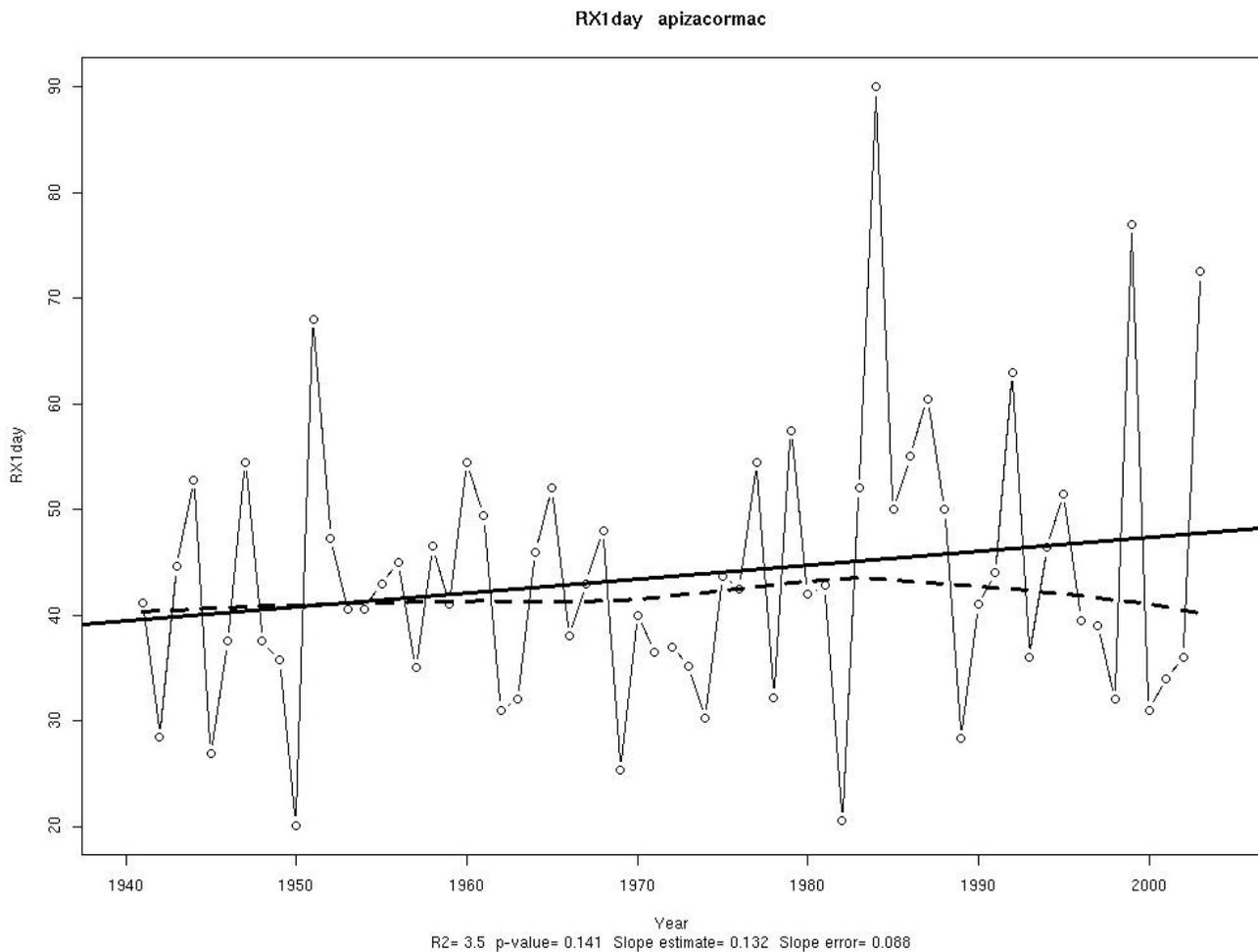


Figura 17-II. Máxima anual de precipitación en un día.

En la figura 17-II se observa que la tendencia lineal muestra un aumento de la máxima anual en un día. Aunque esta tendencia no se reproduce al tomar las máximas mensuales. Sin embargo esto sí podría significar que ahora los eventos de un día de precipitación intensa están aumentando.

RX5day apizacormac

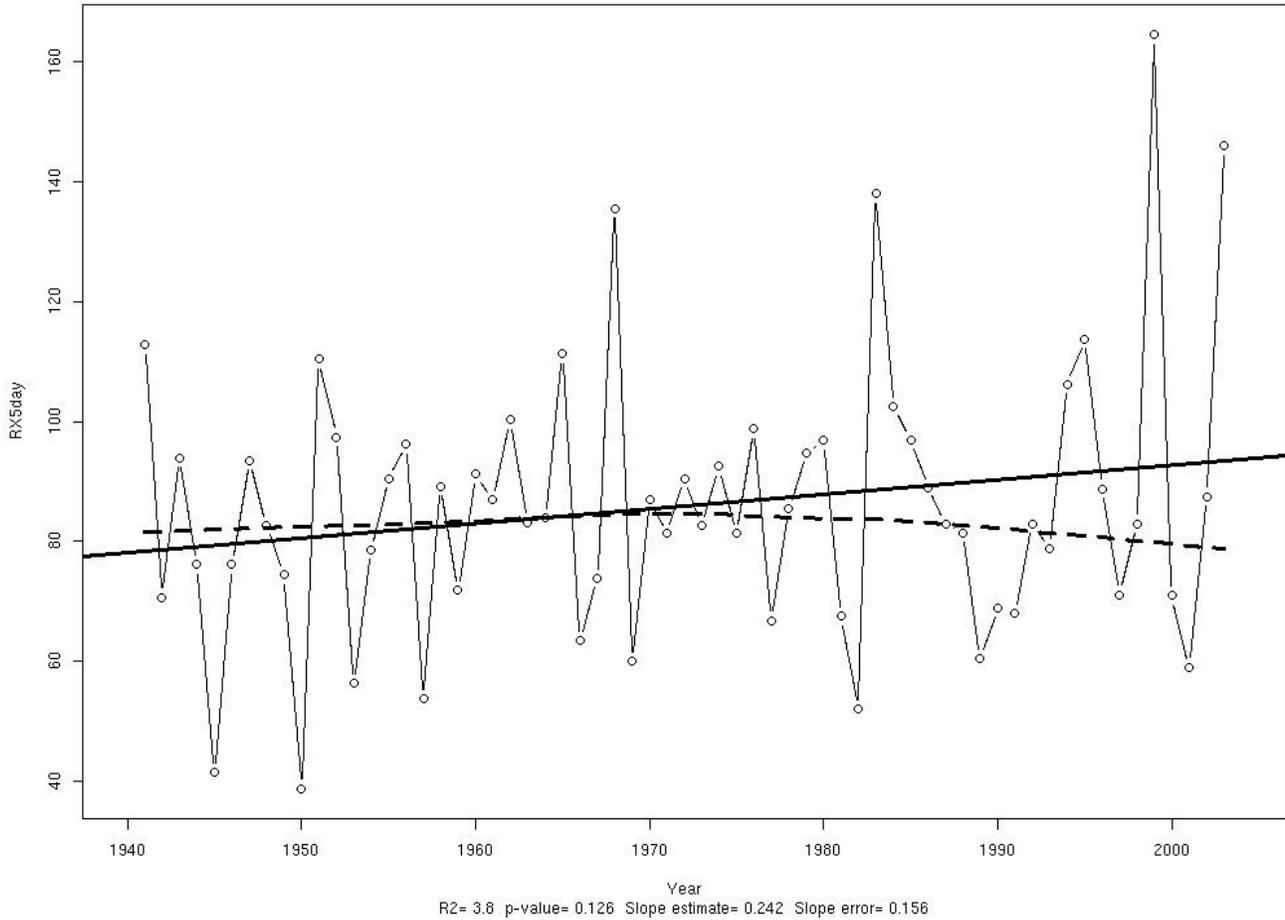


Figura 18-II, Máxima anual en 5 días consecutivos de precipitación.

La tendencia de la máxima anual en cinco días consecutivos de precipitación (figura 18-II) también es positiva y significativa.

Sin embargo, esta misma tendencia no se observa en las máximas mensuales, dónde algunas tendencias no son significativas y otras incluso son negativas. Por tanto no se puede concluir algo concreto.

SDII apizacormac

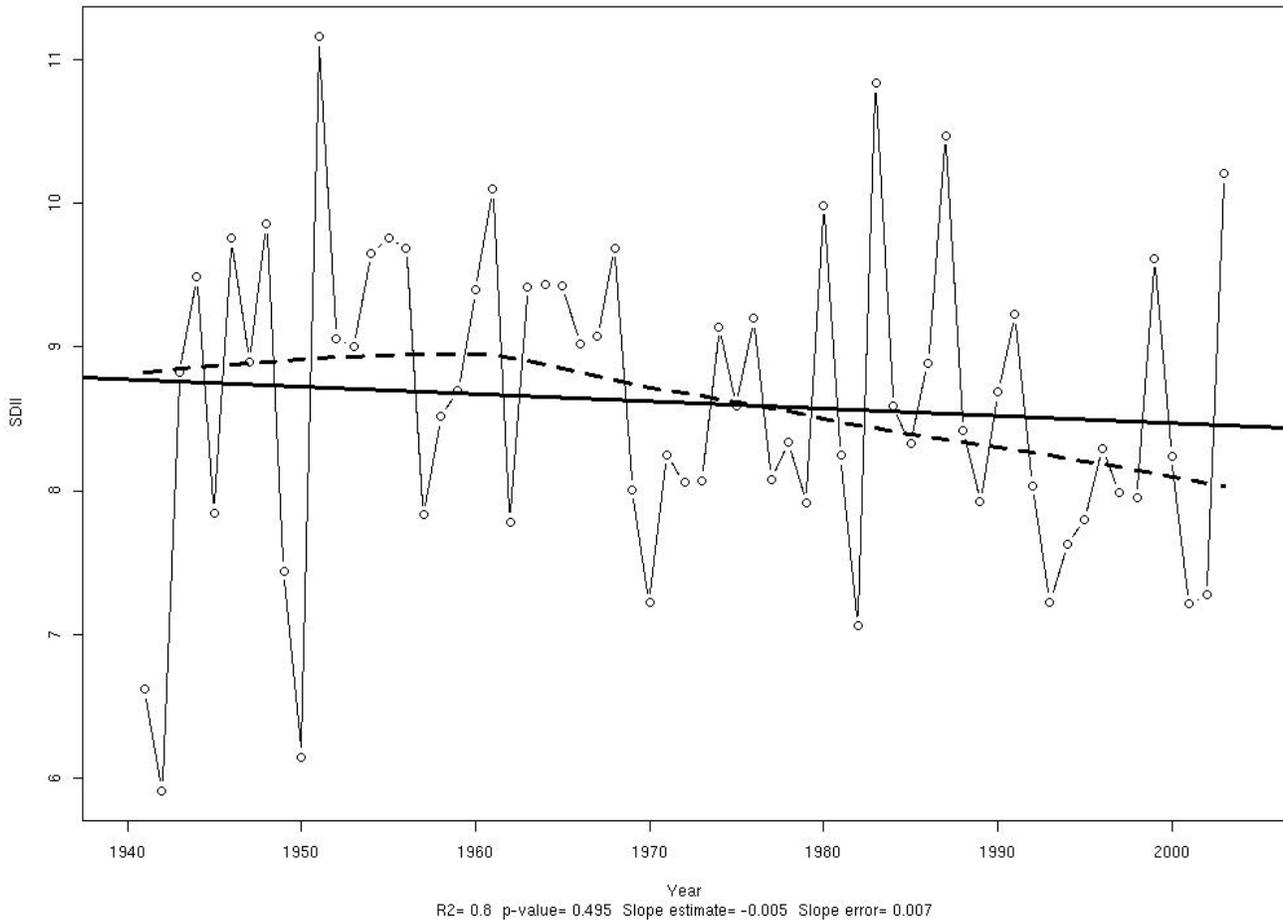


Figura 19-II. Precipitación anual total dividida entre el número de días húmedos, definidos como días con precipitación mayor o igual a 1 mm, en un año.

En la figura 19-II ambas tendencias son significativas y negativas. Este resultado puede ser producto tanto de que la precipitación anual total disminuyó (figura 12-II), o que el número de días con precipitación mayor a 1 mm aumentó o incluso ambas. En el caso de este ejemplo la precipitación anual total no muestra ninguna tendencia significativa, por lo que se puede suponer que ha permanecido sin cambios mientras que el número de días húmedos tendría que haber aumentado. Esto nos dice entonces que los eventos de un día de precipitación están siendo menos intensos, lo cual es contrario a las consecuencias que tendría un ciclo hidrológico más intenso.

## **Análisis**

La precipitación total se ha mantenido constante junto con el número máximo de días húmedos consecutivos y el número de días con precipitación mayor a 10 mm. Mientras que el número de días con precipitación mayor a 20 mm, mayor al 95% y al 99%, la máxima anual en un día de precipitación y la máxima anual en 5 días han aumentado. Lo único que ha disminuido es el índice simple de intensidad. Estos resultados nos dan una caracterización muy específica para la precipitación de la zona, ya que como sabemos la precipitación es una variable compleja y con comportamientos diferentes aún en zonas contiguas. El uso de estos índices es relevante porque de ellos podemos obtener información rápida que nos da las características de la precipitación local.

# Modelos

Para saber si el número de eventos extremos ha cambiado tenemos que conocer la serie a fondo, para ello debemos modelar la forma de los eventos extremos. No es suficiente contar cuántos eventos han habido cada año, además es necesario saber cómo serán en el futuro (sin contemplar cambios en los forzantes actuales), a partir de los datos que poseemos en este momento.

Para hacer el ajuste de los modelos se usó nuevamente una aplicación de R, cuyo nombre es *ExtRemes*<sup>5</sup>. Dicha herramienta usa, igualmente como base teórica las funciones de Coles (2001) en ella se pueden hacer análisis en los dos métodos existentes dentro de la Teoría de Valores Extremos, es decir, en Modelos Asintóticos y Modelos de Umbral.

## DIAGNOSTICO DEL MODELO

La información obtenida a partir del modelo es muy sensible a la exactitud de éste. Así entonces es necesario determinar si el modelo es adecuado. Dada la naturaleza del problema, donde el futuro es desconocido, la única forma de determinar la exactitud del modelo es comparar éste con una amplia población de datos, sin embargo esto no siempre es posible, por tanto lo más práctico es compararlo con los valores introducidos. Entonces, el modelo será adecuado si es capaz de reproducir las variaciones en la distribución de los datos usados para estimarlo.

Suponemos que  $x_1, \dots, x_n$  son realizaciones independientes de una población común con una función de distribución desconocida  $F$ . Un estimador de  $F$ , digamos  $\hat{F}$ , se ha obtenido mediante máxima verosimilitud. Deseamos evaluar la plausibilidad de que  $x_i$  sean una muestra de valores aleatorios de  $\hat{F}$ . Primero una estimación sin modelos de  $F$  puede ser obtenida empíricamente a partir de los datos. Sea  $x_{(1)}, \dots, x_{(n)}$  la muestra ordenada, tal que  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ . Para cualquiera de los  $x_{(i)}$ , exactamente  $i$  de las  $n$  observaciones tienen valores menores o iguales a  $x_{(i)}$ , entonces la estimación empírica de la probabilidad de que una observación sea menor o igual a  $x_{(i)}$  es  $\tilde{F}(x_{(i)}) = i/n$ . Haciendo un ajuste para evitar  $\tilde{F}(x_{(n)}) = 1$ , tenemos  $\tilde{F}(x_{(i)}) = i/(n+1)$ . Esto nos lleva a la siguiente definición.

Dada la muestra ordenada de los observaciones independientes

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

de una población con una función de distribución  $F$ , la distribución empírica de la función se define por:

$$\tilde{F}(x) = \frac{i}{n+1} \quad \text{para } x_{(i)} \leq x < x_{(i+1)}$$

---

<sup>5</sup> Ver apéndice A.4.

Si  $\tilde{F}$  es el estimado de la probabilidad real de la función de distribución  $F$ , entonces debe estar razonablemente de acuerdo con el candidato a modelo  $\hat{F}$ . Así entonces  $\hat{F}$  será un buen estimado de  $F$  si es similar a  $\tilde{F}$ . Por tanto, se hace una comparación entre la probabilidad real  $\tilde{F}$  y la estimada  $\hat{F}$ . Para ello existen dos técnicas, que se muestran a continuación.

#### Plot de Probabilidad

Dada la muestra ordenada de observaciones independientes

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

de una población con una función de distribución aproximada  $\hat{F}$ , un *plot* de probabilidades consiste de los puntos:

$$\left\{ \left( \hat{F} \left( x_{(i)} \right) \frac{i}{n+1} \right) : i = 1, \dots, n \right\}$$

Si  $\hat{F}$  es un modelo razonable de la distribución población, los puntos del *plot* de probabilidad deben caer cerca de la diagonal unidad. Desviaciones sustanciales de esta linealidad proveen evidencia del fallo de como modelo para estos datos.

#### Plot de Cuantiles

Dada la muestra ordenada de observaciones independientes

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

De una población con una función de distribución aproximada, un *plot* de cuantiles consiste de los puntos:

$$\left\{ \left( \hat{F}^{-1} \left( \frac{i}{n+1} \right), x_{(i)} \right) : i = 1, \dots, n \right\}$$

El nombre *plot* de cuantiles se deriva del hecho de que las cantidades  $x_{(i)}$  y  $\hat{F}^{-1}\left(\frac{i}{n+1}\right)$  cada una provee estimaciones del  $\frac{i}{n+1}$  cuantil de la distribución  $F$ . Si  $\hat{F}$  es un estimador razonable de  $F$ , entonces el *plot* de cuantiles debe consistir de puntos alineados cerca de la diagonal unidad.

### Modelos Asintóticos

Estos modelos son la piedra angular de la teoría de valores extremos. Se basan en el comportamiento estadístico de  $M_n = \max\{X_1, \dots, X_n\}$  donde  $X_1, \dots, X_n$  es una secuencia de variables aleatorias independientes que tienen como función de distribución común a  $F$ .  $M_n$  representa el máximo de un proceso con  $n$  observaciones o unidades de tiempo. En teoría la distribución de  $M_n$  podría obtenerse perfectamente para todos los valores de  $n$ :

$$\begin{aligned} \Pr\{M_n \leq z\} &= \Pr\{X_1 \leq z, \dots, X_n \leq z\} \\ &= \Pr\{X_1 \leq z\} \times \dots \times \Pr\{X_n \leq z\} \\ &= \{F(z)\}^n \end{aligned}$$

(1)

Lo anterior es poco útil, ya que no se conoce la función  $F$ . Se pueden usar técnicas estadísticas estándar para determinar  $F$  a partir de los datos observados, sin embargo esto no es muy recomendable ya que pequeñas discrepancias en la estimación de  $F$  se pueden intensificar en  $F^n$ . Otro camino es aceptar que  $F$  es

desconocida y buscar familias de modelos que se aproximen a  $F^n$ , la cual puede ser únicamente estimada basándose en los datos extremos. Se busca el comportamiento de  $F^n$  cuando  $n \rightarrow \infty$ .

Sin embargo sólo esto puede no ser suficiente. Para cualquier  $z < z_+$  donde  $z_+$  es el límite superior de la distribución  $F$ ,<sup>6</sup>  $F^n(z) \rightarrow 0$  si  $n \rightarrow \infty$  entonces la distribución de  $M_n$  degenera en un punto de masa  $z_+$ . Esta dificultad se supera al hacer una renormalización en la variable  $M_n$ :

$$M_n^* = \frac{M_n - b_n}{a_n}$$

(2)

Para las secuencias de constantes  $\{a_n > 0\}$  y  $\{b_n\}$ .

Todas las posibles distribuciones límite para  $M_n^*$  están dadas por el siguiente teorema.

Existen secuencias de constantes  $\{a_n > 0\}$  y  $\{b_n\}$  tales que

$$\Pr \left\{ \frac{(M_n - b_n)}{a_n} \leq z \right\} \rightarrow G(z) \quad n \rightarrow \infty$$

(3)

---

<sup>6</sup>  $z_+$  es el valor más pequeño de  $z$  tal que  $F(z) = 1$ .

Donde  $G$  es una distribución no degenerada, entonces  $G$  pertenece a una de las siguientes familias:

$$\begin{aligned} \text{I: } G(z) &= \exp \left\{ -\exp \left[ -\left( \frac{z-b}{a} \right) \right] \right\} \quad -\infty < z < \infty \\ \text{II: } G(z) &= \begin{cases} 0 & z \leq b \\ \exp \left[ -\left( \frac{z-b}{a} \right)^{-\alpha} \right] & z > b \end{cases} \\ \text{III: } G(z) &= \begin{cases} \exp \left\{ -\left[ -\left( \frac{z-b}{a} \right)^\alpha \right] \right\} & z < b \\ 1 & z \geq b \end{cases} \end{aligned}$$

(4)

Lo interesante de esto es que sólo  $M_n^*$  puede converger en alguna de las familias I,II o III sin importar la forma de la distribución  $F$ . En conjunto estas tres familias son conocidas como las distribuciones de valores extremos,

*Gumbel*, *Fréchet* y *Weibull*, respectivamente. Cada familia tiene parámetros de localización y forma,  $a$  y  $b$ ; adicionalmente *Fréchet* y *Weibull* tienen un parámetro de forma  $\alpha$ .

### Distribución General de Valores Extremos

La distribución de valores extremos o GEV (General Extreme Value, por sus siglas en inglés), es una familia de distribuciones probabilísticas continuas desarrollada dentro de la Teoría de Valores Extremos combinando las distribuciones de *Gumbel*, *Fréchet* y *Weibull*. Su importancia radica en el hecho de que es la distribución límite de los máximos de una secuencia de valores aleatorios independientes e idénticamente distribuidos. Para garantizar la independencia de los valores, que es una característica necesaria, se usa el *método de bloques*. Esta distribución es una herramienta muy adecuada en el estudio de eventos extremos de precipitación.

Las tres formas que tienen las distribuciones de extremos se deben a los diferentes comportamientos en las colas de la distribución  $F(X_i)$ . Esto se obtiene considerando el comportamiento de la distribución límite  $G$  en el límite superior,  $z_+$ . Para la distribución *Weibull*  $z_+$  es finito, mientras que para *Fréchet* y *Gumbel* el límite es infinito. Aún más, la densidad de  $G$  decae exponencialmente para *Gumbel* y polinomialmente para *Fréchet*, lo que corresponde a diferentes tasas de decaimiento en la cola de  $F$ . Así entonces, en la práctica las tres distribuciones dan tres diferentes representaciones del comportamiento de los valores extremos. En un principio era común adoptar alguna de las tres representaciones, sin embargo esto generaba dos dificultades importantes. Primero, era necesaria una técnica para elegir la familia que mejor se adecuara a los datos. Segundo, una vez que la decisión había sido tomada, se hacían ciertas inferencias que determinaban si la elección había sido correcta; sin embargo no permitían que la incertidumbre asociada a tal decisión fuera tomada en cuenta, aún cuando esta incertidumbre no fuera despreciable.

Para superar estos inconvenientes se puede reformular el teorema anterior y combinar las tres familias en una sola, cuya función de distribución es:

$$G(z) = \exp \left\{ - \left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}$$

(5)

Definida en el conjunto  $\left\{ z: 1 + \xi \frac{(z - \mu)}{\sigma} > 0 \right\}$ , donde los parámetros satisfacen  $-\infty < \mu < \infty$ ,  $\sigma > 0$  y  $-\infty < \xi < \infty$ . La ecuación anterior es conocida como la *Distribución General de Valores Extremos* y tiene tres parámetros: localización,  $\mu$ ; escala,  $\sigma$ ; y forma,  $\xi$ . Las familias II y III corresponden a los casos  $\xi > 0$  y  $\xi < 0$  respectivamente. El caso  $\xi = 0$  corresponde a la familia tipo I, es decir Gumbel, con la misma función de distribución que se obtiene del teorema sólo que en vez de  $a$  y  $b$ , se tiene  $\sigma$  y  $\mu$ . Los datos en sí mismos determinarán el parámetro de forma  $\xi$  y con él la familia de valores extremos que mejor se ajusta a los datos, eliminando así tanto el inconveniente de la técnica para elegir la mejor familia, como la exclusión de la incertidumbre que esta elección genera.

### Tamaño de Bloques

La distribución General de Valores Extremos proporciona un modelo para la distribución de los máximos de bloques. El método consiste en agrupar los datos en bloques de igual longitud y posteriormente ajustar la GEV a las máximas de esos bloques. El tamaño de los bloques es importante ya que determinará el número de datos con los que se hará el ajuste. Bloques grandes, implican pocos datos lo que genera varianza grande en la estimación; mientras que bloques chicos implican muchos datos haciendo que el ajuste sea pobre y generando problemas de *sesgo*<sup>7</sup> en la estimación y la extrapolación.

### Estimación de los Parámetros

Suponiendo que  $Z_1, \dots, Z_m$  son variables independientes cuya distribución común es GEV, el logaritmo de la verosimilitud para los parámetros de GEV cuando  $\xi \neq 0$  es.

---

<sup>7</sup> El *sesgo*, o *bias* en inglés, de un estimador se define como:  $Bias(\hat{\theta}_0) = E(\hat{\theta}_0) - \theta_0$  es decir, la diferencia entre el valor esperado del estimador y su valor real. Se espera que esta cantidad sea lo más cercana a cero o cero, puesto que esto significará que el parámetro estimado es el verdadero.

$$\ell(\mu, \sigma, \xi) = -m \log \sigma - (1 + 1/\xi) \sum_{i=1}^m \log \left[ 1 + \xi \left( \frac{z_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^m \left[ 1 + \xi \left( \frac{z_i - \mu}{\sigma} \right) \right]^{-1/\xi} \quad (6)$$

Con  $1 + \xi \left( \frac{z_i - \mu}{\sigma} \right) > 0$  para  $i = 1, \dots, m$ .

En el caso  $\xi = 0$  se usa el límite de Gumbel de la distribución GEV, cuyo logaritmo de la verosimilitud es:

$$\ell(\mu, \sigma) = -m \log \sigma - \sum_{i=1}^m \left( \frac{z_i - \mu}{\sigma} \right) - \sum_{i=1}^m \exp \left[ - \left( \frac{z_i - \mu}{\sigma} \right) \right] \quad (7)$$

Maximizando las ecuaciones (6) y (7) respecto al vector de parámetros  $(\mu, \sigma, \xi)$  se obtiene la estimación de la máxima verosimilitud de toda la familia de GEV.

### Niveles de Retorno

Bajo este modelo los datos se agrupan en bloques de secuencias de observaciones de longitud  $n$ , para un valor dado de  $n$  se genera la siguiente secuencia de máxima de bloques  $M_{n,1}, \dots, M_{n,m}$  para los cuales la distribución GEV será ajustada. Generalmente los bloques escogidos corresponden a un periodo de tiempo de longitud de un año, en cuyo caso  $n$  es el número de observaciones en un año y la máxima del bloque es simplemente la máxima anual. La estimación de los cuantiles de la distribución de máximas anuales se obtiene invirtiendo la función de distribución GEV.

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi} \left[ 1 - \{-\log(1-p)\}^{-\xi} \right] & \xi \neq 0 \\ \mu - \sigma \log \{-\log(1-p)\} & \xi = 0 \end{cases}$$

(8)

Donde  $G(z_p) = 1 - p$ , y  $z_p$  es el nivel de retorno asociado al periodo de retorno  $1/p$ , así entonces se espera que el nivel de retorno sea excedido en promedio una vez cada  $1/p$  años. Es decir  $z_p$  es excedido por el máximo anual en un año en particular con una probabilidad  $P$ .

La gráfica de los niveles de retorno se obtiene haciendo primero una pequeña sustitución:  $y_p = -\log(1 - p)$  y sustituyendo en la ecuación anterior, se obtiene:

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi} [1 - y_p^{-\xi}] & \xi \neq 0 \\ \mu - \sigma \log y_p & \xi = 0 \end{cases}$$

(9)

Así entonces al graficar  $z_p$  contra  $y_p$  en escala logarítmica se obtiene: una relación lineal para el caso de  $\xi = 0$ .

Para el caso  $\xi < 0$ , una curva convexa con un límite asintótico cuando  $p \rightarrow 0$  en  $\frac{\mu - \sigma}{\xi}$ ; y una curva cóncava en el caso de  $\xi > 0$  para el cual no hay límite finito. La gráfica de nivel de retorno permite observar la extrapolación de una mejor manera. En la figura 20-II se muestran niveles de retorno para un rango de valores del parámetro de forma.

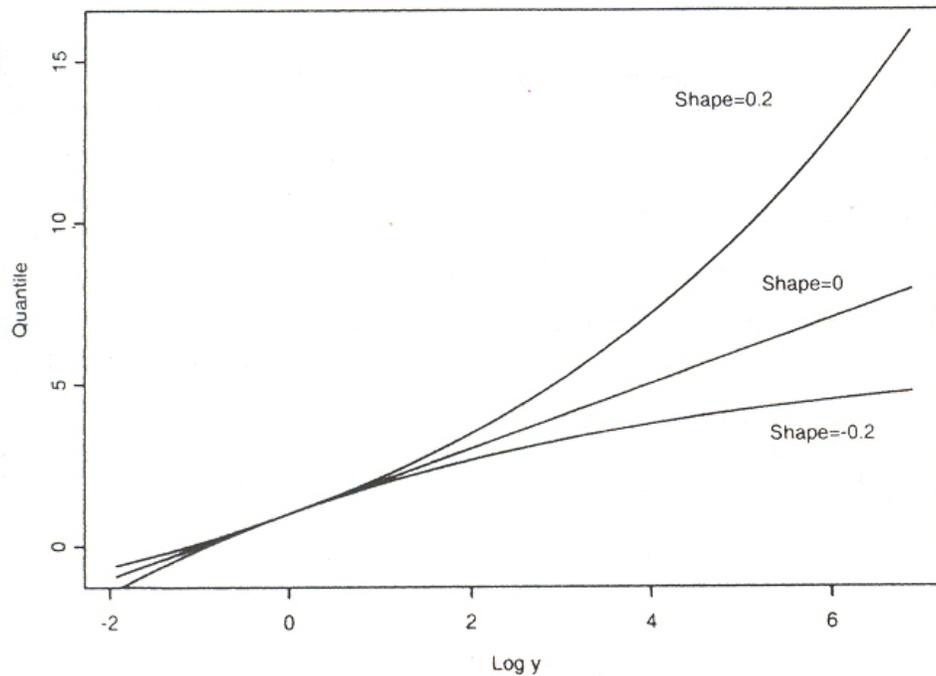


Figura 20-II. Niveles de retorno de la distribución GEV con los parámetros de forma  $\xi = -0.2$ ,  $\xi = 0$ , y  $\xi = 0.2$  respectivamente. (Coles 2001)

### Verificación del Modelo

Debido a que es imposible validar la extrapolación basada en un modelo GEV es necesario hacer una evaluación basada en los datos observados. Aunque lo anterior no es suficiente para justificar la extrapolación, sí es un requisito importante. Para verificar el modelo se pueden usar los *plots* de cuantiles y los *plots* de probabilidades. El *plot* de probabilidad es una comparación de las funciones de distribución empíricas y ajustadas. Ordenando las máximas de los bloques de los datos  $z_1 \leq z_2 \leq \dots \leq z_m$ , la función de distribución empírica evaluada en  $z_i$  está dada por:

$$\tilde{G}(z_i) = \frac{i}{(m+1)}$$

(10)

Sustituyendo los parámetros estimados en (5) la estimación correspondiente basada en el modelo es:

$$\hat{G}(z_i) = \exp \left\{ - \left[ 1 + \xi \left( \frac{z_i - \hat{\mu}}{\hat{\sigma}} \right) \right]^{-1/\xi} \right\}$$

(11)

Así entonces si el modelo GEV funciona adecuadamente, entonces:

$$\hat{G}(z_i) \approx \tilde{G}(z_i)$$

(12)

Para cada  $i$ . Así el *plot* de probabilidad consistirá de los puntos:

$$\left\{ \left( \tilde{G}(z_i), \hat{G}(z_i) \right) \mid i = 1, \dots, m \right\}$$

(13)

Los que deben ser cercanos a la diagonal unidad. Cualquier desviación de esta condición de linealidad significa que existe alguna falla en el modelo.

## Resultados

Todos los parámetros determinados se estimaron mediante el método de máxima verosimilitud.

Los bloques son mensuales, así analizamos el máximo de cada mes en todos los años de la muestra.

Es importante mencionar que el modelo no se ajusta adecuadamente a meses secos, es decir noviembre, diciembre, enero, febrero y marzo. En las figuras 21-II a y b se muestran el ajuste del modelo para el mes de enero y la serie de tiempo de los máximos, respectivamente. Se observa que la serie de tiempo (figura 21-II b) contiene muchos valores cero, lo cual podría ser una razón para que el ajuste del modelo no sea exitoso. En las tablas 3 y 4 se muestran los parámetros obtenidos y la matriz de covarianza, respectivamente. No fue posible determinar el error estándar de los parámetros (tabla 3) dados los valores tan pequeños que se observan en la matriz de covarianza (tabla 4) ya que los errores estándar se calculan usando la raíz cuadrada de los valores en la diagonal de la matriz de covarianza. Claramente se aprecia que el ajuste del modelo para los datos de enero es malo. Resultados similares se presentan para el resto de los meses secos.

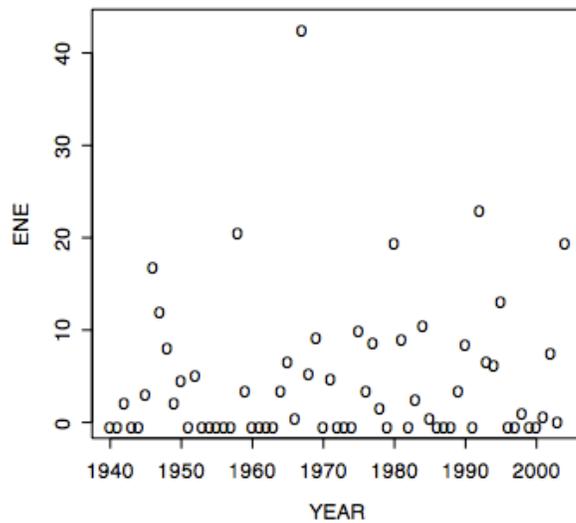


Figura 21-II a. Ajuste de GEV Enero.  
 Figura 21-II b. Serie de tiempo de los máximos de precipitación Enero.

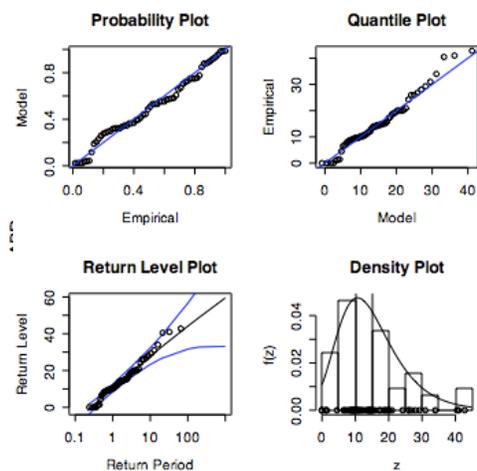
Tabla 3: Parámetros de GEV Enero		
	MLE	Error Estándar
Escala (sigma)	7.6436	NaN
Forma (xi)	6.1794	NaN
Localización (mu)	1.2367	0.0000

El logaritmo negativo de la verosimilitud es 64.1865.

Tabla 4: Covarianza de los Parámetros GEV Enero			
	Localización (mu)	Escala (Sigma)	Forma (Xi)
Localización (mu)	4.0001E-12	9.5467E-12	9.5467E-12
Escala (Sigma)	9.5467E-12	-8.4209E-06	-8.4209E-06
Forma (Xi)	9.5467E-12	-8.4209E-06	-8.4209E-06

#### Abril

El ajuste para el mes de abril fue bueno dado que los puntos en los *plots* de probabilidad y de cuantiles (figura 22-II b) se distribuyen a lo largo de la diagonal.. En la figura 22-II a se observa la serie de tiempo de los máximos de precipitación, se observa que hay menos valores cero, que en el caso del mes de enero (figura 21-II b).



En las tablas 5 y 6 se muestran los parámetros obtenidos y la matriz de covarianza. Figura 22-II a. Ajuste de GEV Abril.  
 Figura 22-II b. Serie de tiempo de los máximos de precipitación Abril.

Tabla 5: Parámetros de GEV Abril		
	MLE	Error Estándar
Escala ( $\sigma$ )	7.7439	0.7669
Forma ( $\xi$ )	-0.0258	0.0869
Localización ( $\mu$ )	10.4768	1.0748

El logaritmo negativo de la verosimilitud es: 234.6481

<b>Tabla 6: Covarianza de los Parámetros GEV Abril</b>			
	<b>Localización (mu)</b>	<b>Escala (Sigma)</b>	<b>Forma (Xi)</b>
<b>Localización (mu)</b>	1.1553	0.2833	-0.0325
<b>Escala (Sigma)</b>	0.2833	0.5881	-0.0195
<b>Forma (Xi)</b>	-0.0325	-0.0195	0.0075

El nivel de retorno para un periodo de 100 años esta entre 40 y 50 mm (figura 22-II b). Para un valor mas exacto del nivel de retorno es mejor usar el *Perfil del Logaritmo de la Verosimilitud*<sup>8</sup> (figura 23-II). De la figura 23-II se puede decir que el intervalo de confianza del Nivel de retorno al 95% para un periodo de retorno de 100 años es (36, 66) mm. Esto significa que en 100 años podemos esperar, con el 95% de confianza, eventos de precipitación entre 36 y 66 mm. Esto es importante si se nota en la figura 22-II a, que no se han observado valores mayores a 45 mm.

---

<sup>8</sup> Ver apéndice A.9.

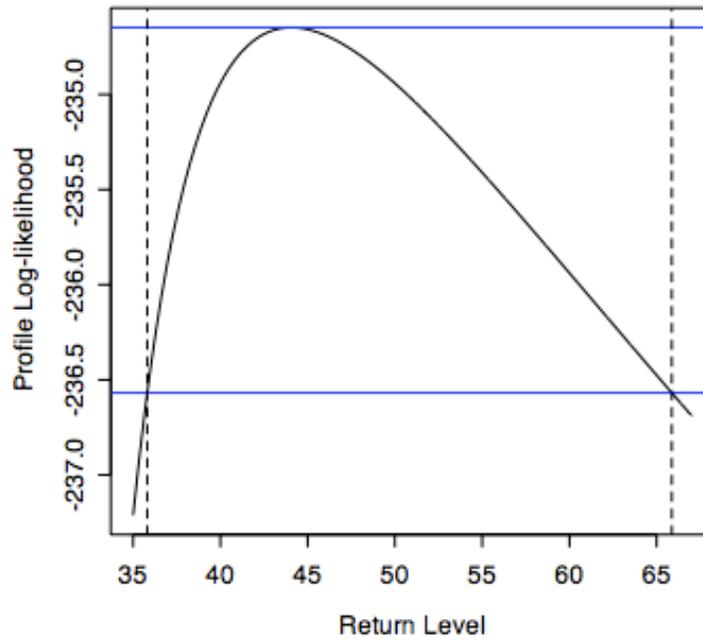
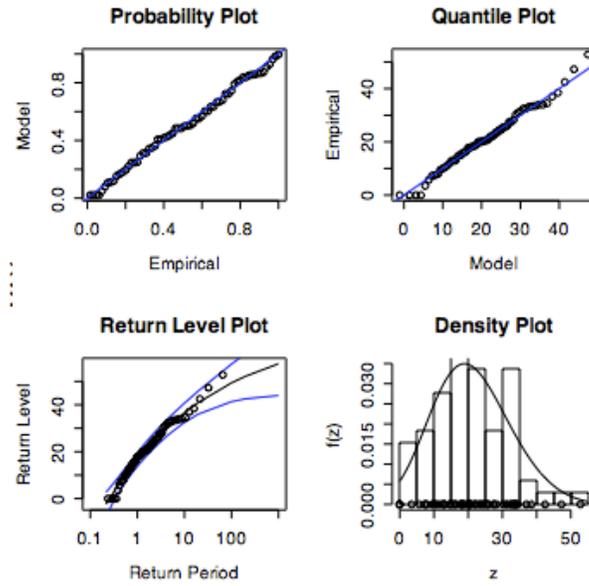


Figura 23-II. Perfil del logaritmo de la verosimilitud para  $\lambda_{0.10}$  y para un periodo de retorno de 100 años.

Mayo

El ajuste para el mes de mayo fue adecuado como se observa en los *plots* de probabilidad y cuantiles de la figura



24-II b. Los parámetros obtenidos y la matriz de covarianza se muestran en las tablas 7 y 8, respectivamente.  
Figura 24-II a. Ajuste de GEV Mayo.  
Figura 24-II b. Serie de tiempo de los máximos de precipitación Mayo.

<b>Tabla 7: Parámetros GEV Mayo</b>		
	<b>MLE</b>	<b>Error Estándar</b>
<b>Escala (sigma)</b>	10.7396	1.0226
<b>Forma (xi)</b>	-0.1940	0.0753
<b>Localización (mu)</b>	16.6462	1.4687

El logaritmo negativo de la verosimilitud es: 249.4971.

<b>Tabla 8: Covarianza de los Parámetros GEV Mayo</b>			
	<b>Localización (mu)</b>	<b>Escala (Sigma)</b>	<b>Forma (Xi)</b>
<b>Localización (mu)</b>	2.1570	0.1905	-0.0400
<b>Escala (Sigma)</b>	0.1905	1.0457	-0.0373
<b>Forma (Xi)</b>	-0.0400	-0.0373	0.0057

De la figura 25-II podemos decir que el intervalo de confianza al 95% del nivel de retorno para un periodo de retorno de 100 años es (43, 63) mm.

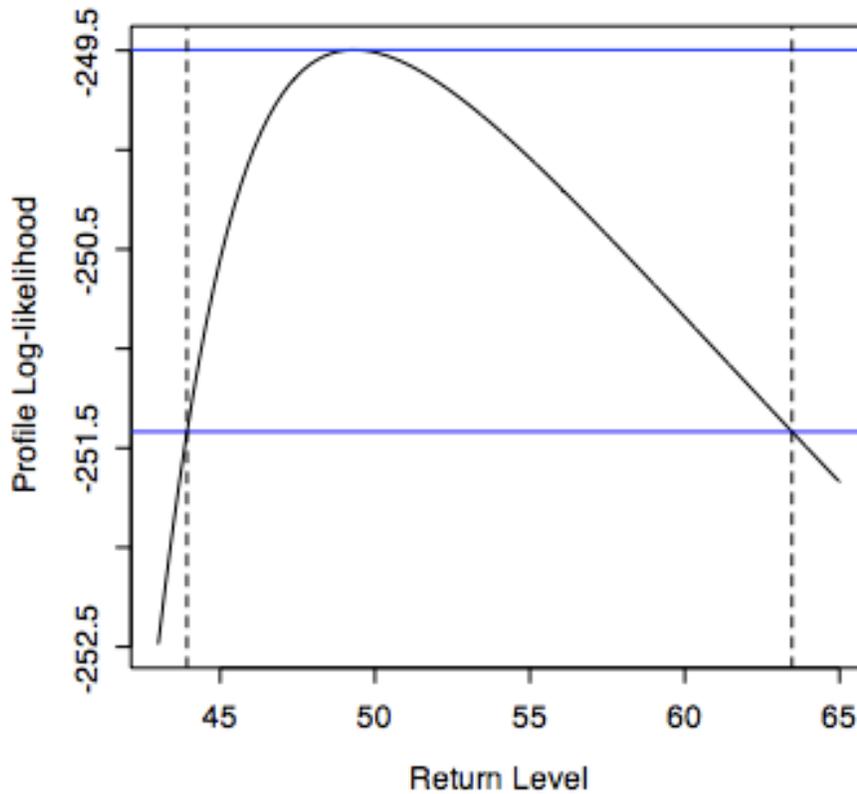


Figura 25-II. Perfil del logaritmo de la Verosimilitud para Mayo y para un periodo de retorno de 100 años

### Junio

De los *plots* de probabilidad y de cuantiles de la figura 26-II b, podemos concluir que el modelo ajustado a los máximos del mes de junio es adecuado ya que los puntos se distribuyen razonablemente a lo largo de las diagonales. Las tablas 9 y 10 muestran los parámetros obtenidos y la matriz de covarianza respectivamente.

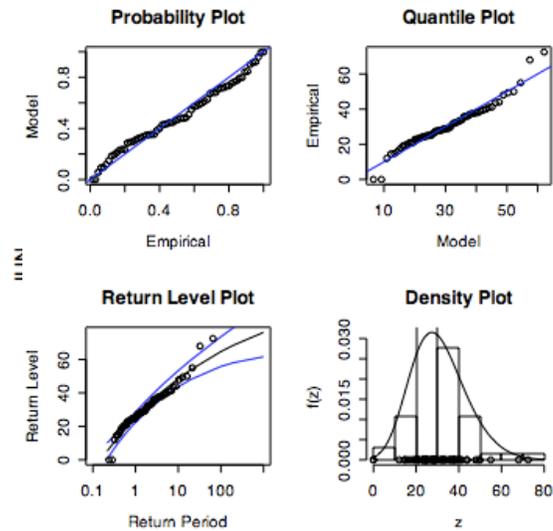


Figura 26-II a. Ajuste de GEV Junio.

Figura 26-II b. Serie de tiempo de los máximos de precipitación Junio.

Tabla 9: Parámetros GEV Junio		
	MLE	Error Estándar
Escala ( $\sigma$ )	11.8039	1.0377
Forma ( $\xi$ )	-0.1505	0.0542
Localización ( $\mu$ )	25.4609	1.5810

El logaritmo negativo de la verosimilitud es: 256.1607.

<b>Tabla 10: Covarianza de los Parámetros GEV Junio</b>			
	<b>Localización (mu)</b>	<b>Escala (Sigma)</b>	<b>Forma (Xi)</b>
<b>Localización (mu)</b>	2.4994	0.1944	-0.0251
<b>Escala (Sigma)</b>	0.1944	1.0769	-0.0186
<b>Forma (Xi)</b>	-0.0251	-0.0186	0.0029

Del perfil del Logaritmo de la Verosimilitud que se muestra en la figura 27-II, se puede decir que el intervalo de confianza al 95% del Nivel de retorno para un periodo de retorno de 100 años es (56,78) mm.

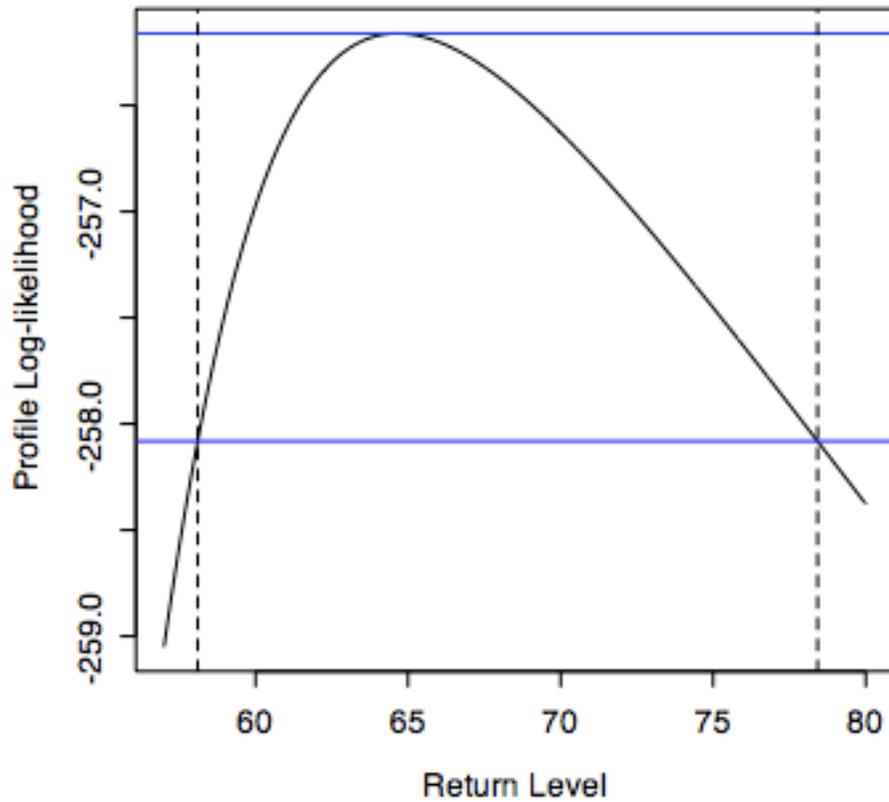


Figura 27-II. Perfil del logaritmo de la Verosimilitud para Junio y para un periodo de retorno de 100 años.

Julio

El modelo se ha ajustado adecuadamente a los datos del mes de julio, tal como se puede ver en la figura 28-II b. La tabla 11 muestra los parámetros obtenidos y sus respectivos errores estándar. La tabla 12 muestra la matriz de covarianza con la que se calcularon dichos errores estándar.

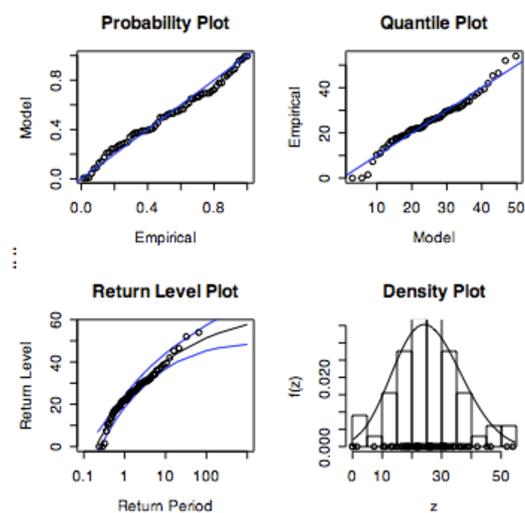


Figura 28-II a. Ajuste de GEV Julio.

Figura 28-II b. Serie de tiempo de los máximos de precipitación Julio.

Tabla 11: Parámetros GEV Julio		
	MLE	Error Estándar
Escala ( $\sigma$ )	10.7502	0.9732
Forma ( $\xi$ )	-0.2391	0.0633
Localización ( $\mu$ )	21.3555	1.4491

El logaritmo negativo de la verosimilitud es: 247.3471.

<b>Tabla 12: Covarianza de los Parámetros GEV Julio</b>			
	<b>Localización (mu)</b>	<b>Escala (Sigma)</b>	<b>Forma (Xi)</b>
<b>Localización (mu)</b>	2.1000	0.0302	-0.0311
<b>Escala (Sigma)</b>	0.0302	0.9471	-0.0299
<b>Forma (Xi)</b>	-0.0311	-0.0299	0.0040

Observando la figura 29-II podemos decir que el intervalo de confianza al 95% del nivel de retorno para un periodo de retorno de 100 años es (45, 62) mm.

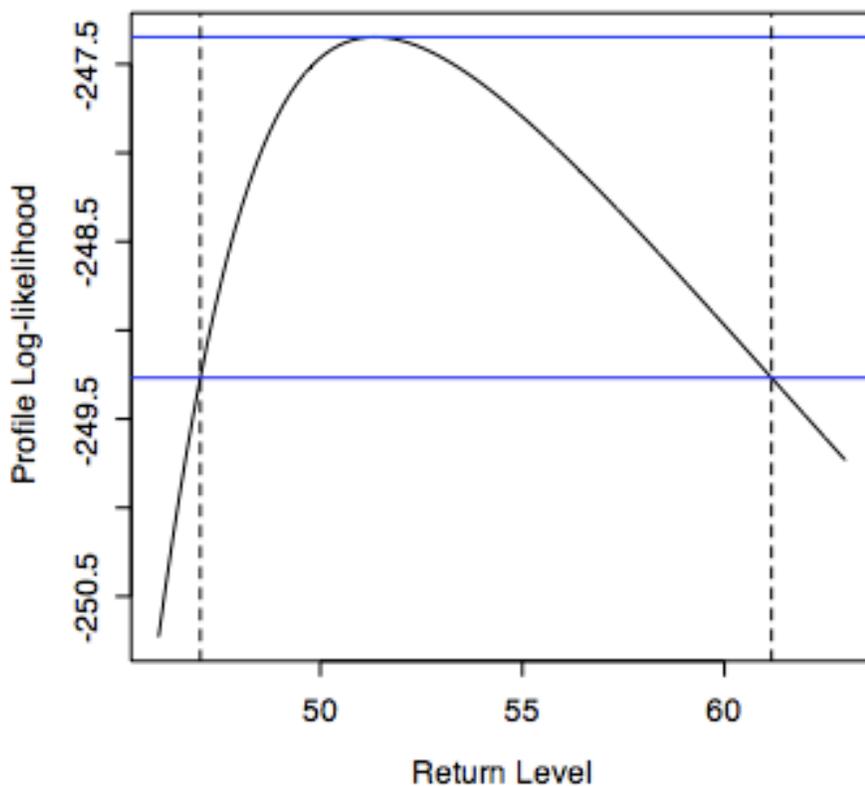


Figura 29-II. Perfil del logaritmo de la Verosimilitud para Julio y para un periodo de retorno de 100 años.

Agosto

El modelo se ajusta adecuadamente aún cuando en los *plots* de probabilidad y de cuantiles (figura 30-II b) se observan pequeñas desviaciones de la diagonal, dichas desviaciones son mínimas y podemos decir que los datos se distribuyen razonablemente a lo largo de ella.

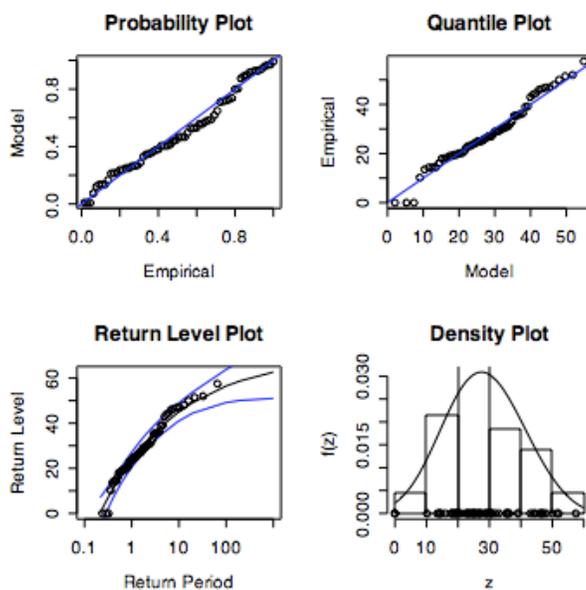


Figura 30-II a. Ajuste de GEV Agosto.

Figura 30-II b. Serie de tiempo de los máximos de precipitación Agosto.

Tabla 13: Parámetros GEV Agosto		
	MLE	Error Estándar
Escala ( $\sigma$ )	12.3988	1.1847
Forma ( $\xi$ )	-0.2693	0.0812
Localización ( $\mu$ )	23.7644	1.6971

El logaritmo negativo de la verosimilitud es: 256.3118.

**Tabla 14: Covarianza de los Parámetros GEV Agosto**

	<b>Localización (mu)</b>	<b>Escala (Sigma)</b>	<b>Forma (Xi)</b>
<b>Localización (mu)</b>	2.8802	0.0583	-0.0526
<b>Escala (Sigma)</b>	0.0583	1.4035	-0.0521
<b>Forma (Xi)</b>	-0.0526	-0.0521	0.0066

Observando la figura 31-II podemos decir que el intervalo de confianza al 95% del nivel de retorno para un periodo de retorno de 100 años es (52, 68) mm.

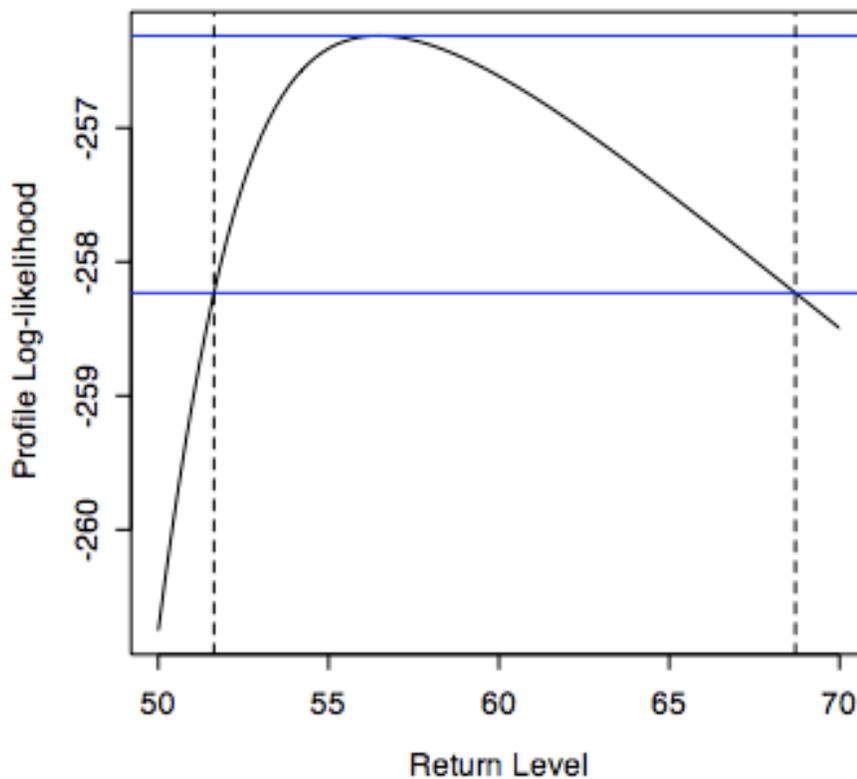


Figura 31-II. Perfil del logaritmo de la Verosimilitud para Agosto y para un periodo de retorno de 100 años.  
Septiembre

El modelo se ha ajustado correctamente como era de esperarse por ser septiembre un mes húmedo y en el cual se ha presentado el valor máximo de 90 mm. En la figura 32-II b se observa en los *plots* de probabilidad y de cuantiles que los puntos se alinean razonablemente en la diagonal.

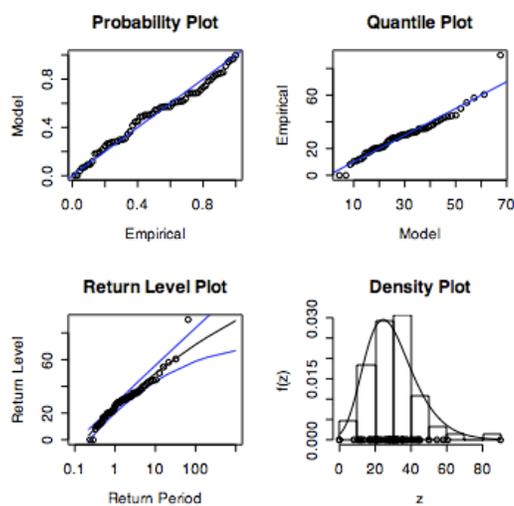


Figura 32-II a. Ajuste de GEV Septiembre.

Figura 32-II b. Serie de tiempo de los máximos de precipitación Septiembre.

**Tabla 15: Parámetros GEV Septiembre**

	MLE	Error Estándar
<b>Escala (sigma)</b>	12.4888	1.1343
<b>Forma (xi)</b>	-0.0823	0.0562
<b>Localización (mu)</b>	23.3873	1.6797

El logaritmo negativo de la verosimilitud es: 262.4223.

**Tabla 16: Covarianza de los Parámetros GEV Septiembre**

	Localización ( $\mu$ )	Escala ( $\sigma$ )	Forma ( $\xi$ )
Localización ( $\mu$ )	2.8212	0.4299	-0.0261
Escala ( $\sigma$ )	0.4299	1.2867	-0.0161
Forma ( $\xi$ )	-0.0261	-0.0161	0.0032

Observando la figura 33-II podemos decir que el intervalo de confianza al 95% del nivel de retorno para un periodo de retorno de 100 años es (62, 91) mm este intervalo incluye el valor de 90 mm que es el máximo valor de todos los datos.

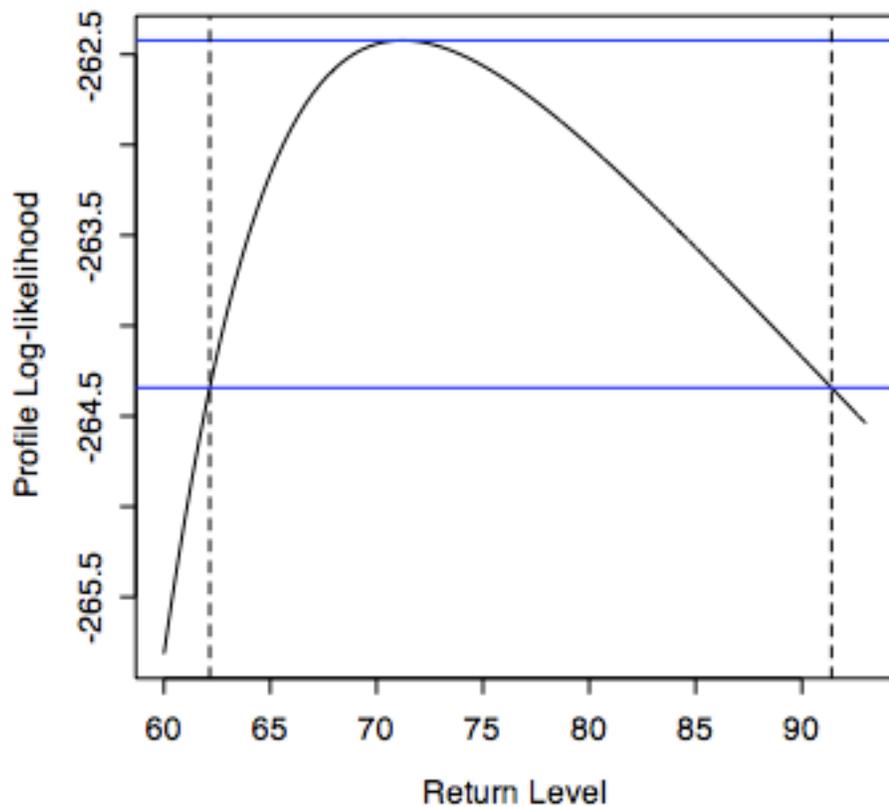


Figura 33-II. Perfil del logaritmo de la Verosimilitud para Septiembre y para un periodo de retorno de 100 años. Octubre

El modelo se ajusta adecuadamente como se aprecia en la figura 34-II b donde los puntos se distribuyen razonablemente a lo largo de la diagonal. En las tablas 17 y 18 se muestran los parámetros determinados y la matriz de covarianza, respectivamente.

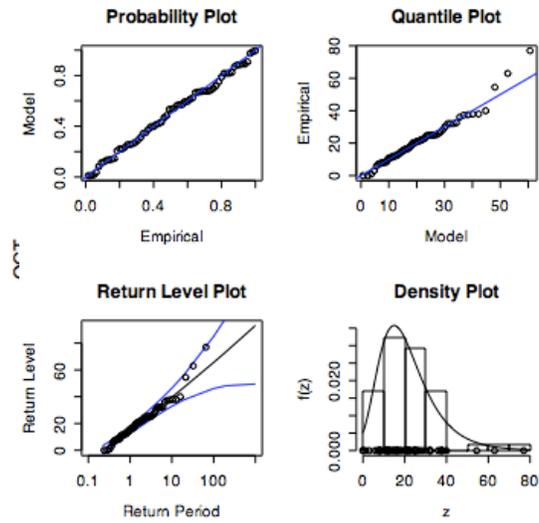


Figura 34-II a. Ajuste de GEV Octubre.

Figura 34-II b. Serie de tiempo de los máximos de precipitación Octubre.

Tabla 17: Parámetros GEV Octubre		
	MLE	Error Estándar
Escala ( $\sigma$ )	10.2882	1.0269
Forma ( $\xi$ )	0.0263	0.0820
Localización ( $\mu$ )	15.1805	1.4203

El logaritmo negativo de la verosimilitud es: 254.8170.

**Tabla 18: Covarianza de los Parámetros GEV Octubre**

	Localización ( $\mu$ )	Escala ( $\Sigma$ )	Forma ( $\Xi$ )
Localización ( $\mu$ )	2.0173	0.5726	-0.0367
Escala ( $\Sigma$ )	0.5726	1.0546	-0.0171
Forma ( $\Xi$ )	-0.0367	-0.0171	0.0067

En la figura 35-II podemos ver que el intervalo de confianza al 95% del nivel de retorno para un periodo de retorno de 100 años es (51, 99) mm. Es interesante observar que este intervalo es más amplio que el correspondiente a septiembre, incluso alcanzando valores de casi 100 mm, es probable que esto se deba a que los valores más grandes en la serie de tiempo (figura 34-II a) se localizan en años más recientes.

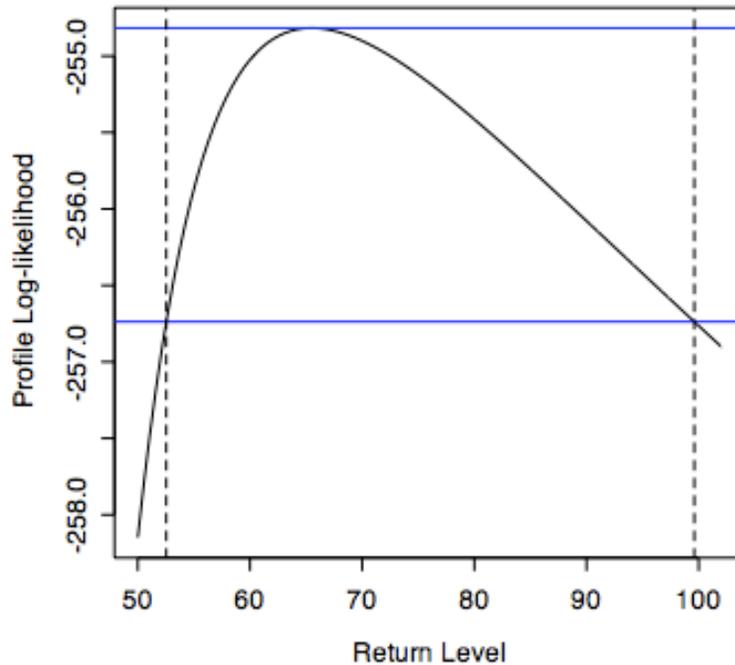


Figura 35-II. Perfil del logaritmo de la Verosimilitud para Octubre y para un periodo de retorno de 100 años.

### Modelos de Umbral

Los Modelos de Umbral surgen de la desventaja que presenta el modelo anterior al no usar todos los datos disponibles. En ocasiones los bloques pueden contener eventos extremos que se descartan al ser menores a la máxima. Los Modelos de Umbral superan esta desventaja al establecer un valor a partir del cual se definirán los eventos extremos. Así entonces se trabaja con los valores por encima de este umbral. Si  $X_1, X_2, \dots$  es una secuencia de variables independientes e idénticamente distribuidas aleatoriamente y teniendo una función de distribución  $F$ , entonces es natural considerar que los eventos extremos son las  $X_i$  que sobrepasan cierto umbral  $u$ . Ahora si  $X$  es un elemento arbitrario de la secuencia de  $X_i$ , la descripción estocástica del comportamiento de los eventos extremos estará dada por la probabilidad condicional:

$$\Pr \{X > u + y | X > u\} = \frac{1 - F(u + y)}{1 - F(u)} \quad y > 0$$

(14)

Si la distribución  $F$  fuera conocida, también lo sería la distribución de los excedentes del umbral. Sin embargo este no es el caso, por lo que es necesario buscar aproximaciones que son generalmente aplicables a los valores extremos. Esto es similar al modelo anterior, donde se usa GEV para aproximar la distribución de los máximos en una secuencia larga, sólo que en este caso la población que la genera no se conoce.

Los modelos de umbral se pueden abordar mediante la distribución *Pareto Generalizada* y *Puntos sobre Umbral*.

### Distribución Pareto Generalizada

La distribución Pareto es una distribución que describe el comportamiento de variables que presentan gran cantidad de valores chicos y algunos grandes, por tanto es una buena aproximación de la distribución que presentan los valores extremos.

Sea  $X_1, X_2, \dots$  una secuencia de variables independientes e idénticamente distribuidas aleatoriamente y teniendo una función de distribución  $F$ , sea  $M_n = \max \{X_1, \dots, X_n\}$ . Ahora  $X_i$  es un elemento arbitrario de la secuencia de  $X_i$  y suponemos que  $F$  satisface las características de los modelos asintóticos, así entonces para una  $n$  grande tenemos:

$$\Pr \{M_n \leq z\} \approx G(z)$$

(15)

Donde

$$G(z) = \exp \left\{ - \left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}$$

(16)

Para  $\mu, \sigma > 0$  y  $\xi$ . Entonces para  $u$  suficientemente grande la función de distribución de  $(X - u)$ , donde  $X > u$ , es aproximadamente:

$$H(y) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi}$$

(17)

Definiendo  $y$  como:  $\{y : y > 0 \text{ y } (1 + \xi y / \sigma) > 0\}$  donde

$$\tilde{\sigma} = \sigma + \xi(u - \mu)$$

(18)

Así entonces  $H(y)$  define una familia de distribuciones conocida como *Pareto generalizada*. Es importante notar que los parámetros de esta distribución de excedencia de umbral son únicamente determinados por aquellos asociados a la distribución GEV de máximos de bloques. En particular el parámetro  $\xi$  es igual al correspondiente en la distribución GEV, ya que es independiente del tamaño del bloque. Mientras que el cálculo de  $\tilde{\sigma}$  en (18) no se modifica por los cambios de  $\mu$  o  $\sigma$ , que se autocompensan.

La dualidad que existe entre GEV y la distribución Pareto generalizada significa que el parámetro de forma,  $\xi$ , es muy importante al determinar el comportamiento cualitativo de la distribución Pareto generalizada, tal como en la distribución GEV. Si  $\xi < 0$  la distribución de los excesos tiene un límite superior,  $u - \sigma/\xi$ ; si  $\xi > 0$  la distribución no tiene límite superior. La distribución tampoco tiene límites si  $\xi = 0$ , al hacer el límite cuando  $\xi \rightarrow 0$  se obtiene:

$$H(y) = 1 - \exp\left(-\frac{y}{\sigma}\right)$$

(19)

con  $y > 0$  y que corresponde a una distribución exponencial con un parámetro  $1/\theta$ .

### Selección del Umbral

En estos modelos se identifica a los eventos extremos mediante un umbral, por tanto es importante establecer un umbral adecuado. El problema de obtener el umbral correcto en estos modelos es análogo al problema de determinar el tamaño de los bloques en la otra familia de modelos, un umbral grande puede dejar fuera muchos datos haciendo que el ajuste no sea muy bueno, por el contrario un umbral chico podría incluir demasiados datos que podrían no ser valores extremos. Anteriormente se mencionó que el IPCC define el umbral en el nonagésimo percentil de la distribución de la precipitación, sin embargo este umbral puede no ser el adecuado para que el modelo funcione, aunque este umbral garantice que se trabajará con el 10% de los datos disponibles éstos pueden no ser suficientes. Para evitar trabajar con un umbral inadecuado es recomendable comparar el umbral definido por el IPCC con el sugerido por el modelo mismo.

El objetivo es elegir el umbral más chico posible tal que incluya el número de datos necesarios para obtener una aproximación razonable. Existen dos métodos disponibles: uno es una *técnica exploratoria* que se debe llevar a cabo antes de la estimación del modelo; el otro es una *evaluación de la estabilidad* de la estimación de los parámetros, se basa en el ajuste de modelos para un rango de diferentes umbrales.

### MÉTODO 1

El primer modelo se basa en la media de la distribución Pareto generalizada. Si  $Y$  tiene una distribución Pareto con los parámetros  $\sigma$  y  $\xi$ , entonces:

$$E(Y) = \frac{\sigma}{1 - \xi}$$

(20)

Con  $\xi < 1$ . Cuando  $\xi \geq 1$  la media es infinita. Si suponemos que la distribución Pareto generalizada es un modelo válido para los excedentes de un umbral  $u_0$  generados por la serie  $X_1, \dots, X_n$  donde se toma a  $X$  como un término arbitrario, entonces de (20):

$$E(X - u_0 | X > u_0) = \frac{\sigma_{u_0}}{1 - \xi}$$

(21)

Con  $\xi < 1$  donde  $\sigma_{u_0}$  es el parámetro de escala correspondiente a los excedentes del umbral  $u_0$ . Si la distribución Pareto generalizada es válida para  $u_0$  también debería serlo para  $u > u_0$  con el respectivo cambio en el parámetro de escala a  $\sigma_u$ . Entonces para  $u > u_0$  y usando (18) obtenemos:

$$E(X - u | X > u) = \frac{\sigma_u}{1 - \xi} = \frac{\sigma_{u_0} + \xi u}{1 - \xi}$$

(22)

Para  $u > u_0$ ,  $E(X - u | X > u)$  es una función lineal de  $u$ . De hecho es simplemente la media de los excedentes del umbral  $u$ , para la cual la media de la muestra de los excedentes de  $u$  provee una estimación empírica. Se espera que dicha estimación varíe linealmente con  $u$ , a niveles de  $u$  donde el modelo Pareto generalizado es apropiado. Esto lleva al siguiente proceso. La localización de los puntos:

$$\left\{ \left( u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_i - u) \right) : u < x \max \right\}$$

(23)

Donde  $x_1, \dots, x_{n_u}$  consiste en las  $n_u$  observaciones que exceden el umbral  $u$ , y  $x_{\max}$  es el máximo de  $X_i$ , este termino es llamado *vida media residual*. En conclusión, por encima del umbral  $u_0$  para el cual la distribución Pareto generalizada es una buena aproximación de la distribución de los excedentes, el diagrama de la vida media residual debe ser aproximadamente lineal en  $u$ .

## MÉTODO 2

El otro método consiste en ajustar la distribución Pareto generalizada a un rango de umbrales buscando la estabilidad de la estimación de los parámetros. Al igual que en el método anterior la idea se basa en que si la distribución Pareto generalizada es una buena aproximación de la distribución de los excedentes del umbral  $u_0$  entonces también lo debe ser para  $u > u_0$ . El parámetro de escala para las distribuciones con umbrales  $u_0$  y  $u$  debe ser el mismo. Sin embargo al denotar  $\sigma_u$  como el parámetro de escala para  $u > u_0$  y por (18) se obtiene:

$$\sigma_u = \sigma_{u_0} + \xi(u - u_0)$$

(24)

De tal forma que para que  $\sigma_u$  no cambie  $\xi = 0$ . Para superar este problema se reparametriza  $\sigma$  como:

$$\sigma^* = \sigma_u + \xi u$$

(25)

El cual es constante respecto a  $u$  usando (24). En consecuencia la estimación de  $\sigma^*$  y de  $\xi$  debe ser constante para  $u_0$ . Sin embargo la variabilidad de la muestra hace que estas cantidades no sean exactamente constantes, pero sí que sean tan estables como los errores de las muestras lo permitan; esto es, mientras el umbral crece los errores o la varianza de la muestra también y se observa en la figura 36-II. Finalmente el umbral correcto  $u_0$  será el mínimo de  $u$  donde las estimaciones permanecen casi constantes.

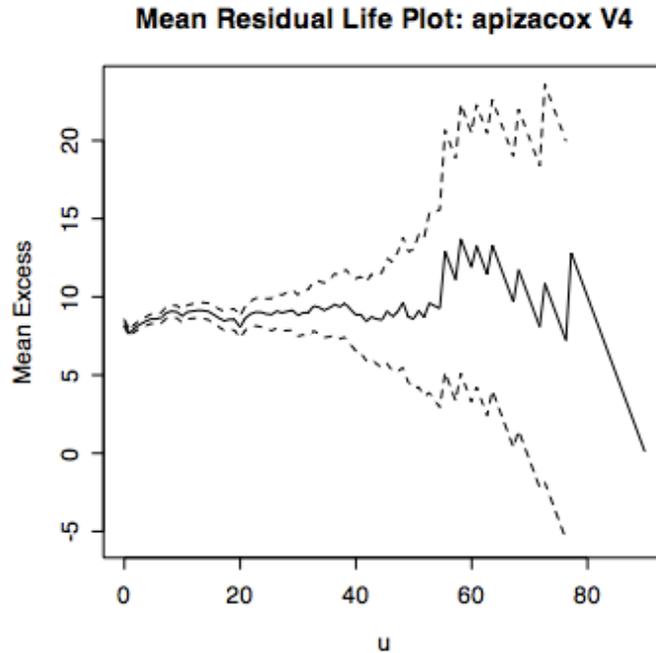


Figura 36-II. Vida Media Residual.

La idea es tomar un valor donde la Vida Media Residual sea aproximadamente lineal, dentro del 95% de confianza, es decir, donde la línea no presente cambios bruscos que no puedan ser descritos mediante funciones lineales. En el caso de nuestros datos esto se da, aproximadamente, como entre 20 y 40 mm. Lo cual es bueno ya que coincide con el percentil 90 de la distribución y que corresponde a 20 mm justamente.

### Estimación de los Parámetros

Una vez que el umbral ha sido determinado, los parámetros para la distribución Pareto generalizada se estiman mediante máxima verosimilitud. Suponemos  $y_1, \dots, y_k$  como los  $k$  excedentes del umbral  $u$ . Para  $\xi \neq 0$  el logaritmo de la verosimilitud se deriva de (18) como:

$$\ell(\sigma, \xi) = -k \log \sigma - (1 + 1/\xi) \sum_{i=1}^k \log(1 + \xi y_i / \sigma)$$

(26)

Donde  $(1 + \xi y_i / \sigma) > 0$  para  $i = 1, \dots, k$ ; de otra forma,  $\ell(\sigma, \xi) = -\infty$ . En el caso  $\xi = 0$  el logaritmo de la verosimilitud se obtiene de (19) como:

$$\ell(\sigma) = -k \log \sigma - (1/\sigma) \sum_{i=1}^k y_i$$

(27)

Debido a que una maximización analítica de esta verosimilitud no es posible, se requieren de técnicas numéricas que eviten la inestabilidad que se presenta en  $y$  que aseguren que el algoritmo no fallará durante la evaluación fuera del espacio permitido por los parámetros. El error estándar y los intervalos de confianza se estiman de manera usual en la teoría estándar de verosimilitud.

### Niveles de Retorno

Es más conveniente interpretar los modelos de valores extremos en términos de cuantiles o niveles de retorno en vez de hacerlo mediante valores de los parámetros. Suponiendo que la distribución Pareto generalizada con parámetros  $\sigma$  y  $\xi$  es un buen modelo de la excedencia de un umbral  $u$  para la variable  $X$ . Esto es, para  $x > u$ .

$$\Pr \{X > x | X > u\} = \left[ 1 + \xi \left( \frac{x - u}{\sigma} \right) \right]^{-1/\xi}$$

(28)

De aquí se sigue que

$$\Pr \{X > x\} = \zeta_u \left[ 1 + \xi \left( \frac{x - u}{\sigma} \right) \right]^{-1/\xi}$$

(29)

Donde  $\zeta_u = \Pr \{X > u\}$ . Entonces el nivel  $x_m$  que es excedido en promedio una vez cada  $m$  observaciones es la solución de:

$$\zeta_u \left[ 1 + \xi \left( \frac{x_m - u}{\sigma} \right) \right]^{-1/\xi} = \frac{1}{m}$$

(30)

Reagrupando

$$x_m = u + \frac{\sigma}{\xi} \left[ (m\zeta_u)^\xi - 1 \right]$$

(31)

Dada  $m$  tiene que ser suficientemente larga para asegurar que  $x_m > u$ . Asumiendo que  $\xi \neq 0$ .

Si  $\xi = 0$ , entonces usando (19) obtenemos

$$x_m = u + \sigma \log(m\zeta_u)$$

(32)

Y  $m$  de nuevo debe ser suficientemente grande.

Por construcción  $x_m$  es el nivel de retorno de la  $m$ -observación. Usando (31) y (32) para graficar  $x_m$  contra  $m$  en escala logarítmica se obtienen las mismas cualidades que las gráficas de los niveles de retorno de los modelos

basados en GEV, esto es: linealidad si  $\xi = 0$ ; concavidad si  $\xi > 0$ ; convexidad si  $\xi < 0$ . Generalmente el nivel de retorno se presenta en una escala anual de manera que el nivel de retorno N es el nivel que se espera se exceda una vez cada N años. Si hay  $n_y$  observaciones por año esto corresponde al nivel de retorno de la m-observación, donde  $m = N \times n_y$ . Entonces, el nivel de retorno del N-año esta definido por:

$$z_N = u + \frac{\sigma}{\xi} \left[ \left( N n_y \zeta_u \right)^\xi - 1 \right]$$

(33)

En el caso en que  $\xi = 0$ ; en caso contrario:

$$z_N = u + \sigma \log \left( N n_y \zeta_u \right)$$

(34)

La estimación de los niveles de retorno requiere la sustitución de los valores de los parámetros  $\sigma$  y  $\xi$  por sus valores estimados mediante *máxima verosimilitud*. Adicionalmente se requiere de  $\hat{\zeta}_u$ , que es la probabilidad de que una observación individual exceda el umbral y cuya estimación natural es:

$$\hat{\zeta}_u = \frac{k}{n}$$

(35)

Que es la muestra de la proporción de puntos que exceden  $u$ . Como el número de excedentes de  $u$  sigue la distribución Binomial,  $\text{Bin} \left( n, \zeta_u \right)$ , entonces  $\hat{\zeta}_u$  es también la estimación de máxima verosimilitud para  $\zeta_u$ .

Los errores estándar o los intervalos de confianza para  $x_m$  se obtienen mediante el *método delta*<sup>9</sup>, la incertidumbre en la estimación de  $\zeta_u$  también se puede incluir en el cálculo. De las propiedades estándar de la distribución binomial se tiene que la varianza es  $\text{var}(\hat{\zeta}_u) \approx \hat{\zeta}_u (1 - \hat{\zeta}_u) / n$ , entonces la matriz de varianza-

covarianza completa para  $(\hat{\zeta}_u, \hat{\sigma}, \hat{\xi})$  es:

$$V = \begin{bmatrix} \hat{\zeta}_u (1 - \hat{\zeta}_u) / n & 0 & 0 \\ 0 & v_{1,1} & v_{1,2} \\ 0 & v_{2,1} & v_{2,2} \end{bmatrix}$$

(36)

Donde  $v_{i,j}$  denotan los  $(i,j)$  términos de la matriz de varianza-covarianza de  $\hat{\sigma}$  y de  $\hat{\xi}$ . Entonces por el método delta obtenemos:

$$\text{var}(\hat{x}_m) \approx \nabla x_m^T V \nabla x_m$$

(37)

Donde

---

<sup>9</sup> Ver apéndice A.12

$$\nabla x_m^T = \left[ \frac{\partial x_m}{\partial \zeta_u}, \frac{\partial x_m}{\partial \sigma}, \frac{\partial x_m}{\partial \xi} \right] = \left[ \sigma m \zeta_u^{\xi-1}, \xi^{-1} \left\{ (m \zeta_u)^\xi - 1 \right\} - \sigma \xi^{-2} \left\{ (m \zeta_u)^\xi - 1 \right\} + \sigma \xi^{-1} (m \zeta_u)^\xi \log(m \zeta_u) \right] \quad (38)$$

Evaluated in  $(\hat{\zeta}_u, \hat{\sigma}, \hat{\xi})$ .

Se obtienen mejores estimaciones en precisión para los parámetros y niveles de retorno al usar los perfiles de verosimilitud apropiados. Para  $\sigma$  y  $\xi$  esto es directo, mientras que para los niveles de retorno se requiere una reparametrización. Se puede ignorar la incertidumbre de  $\zeta_u$  que generalmente es pequeña en comparación a la de los otros parámetros. De (31) y (32) se obtiene:

$$\sigma = \begin{cases} \frac{(x_m - u)^\xi}{(m \zeta_u)^\xi - 1} & \text{si } \xi \neq 0 \\ \frac{x_m - u}{\log(m \zeta_u)} & \text{si } \xi = 0 \end{cases}$$

(39)

Al fijar y sustituir en (26) se obtiene la verosimilitud de un parámetro que puede ser maximizado respecto a  $\xi$ . Como función de  $x_m$ , así entonces este es el perfil de la verosimilitud logarítmica para el nivel de retorno de las  $m$ -observaciones.

### Verificación del Modelo

*Plots* de probabilidad, cuantiles, niveles de retorno y densidad son útiles para determinar la calidad del ajuste del modelo Pareto generalizado. Determinando un umbral  $u$ , sus excedentes  $y_1 \leq \dots \leq y_k$  y estimando el modelo  $\hat{H}$ , el *plot* de probabilidad está formado por los pares:

$$\left\{ \left( \frac{i}{k+1}, \hat{H}(y_i) \right) \mid i = 1, \dots, k \right\}$$

(40)

Donde

$$\hat{H}(y) = 1 - \left( 1 + \frac{\hat{\xi} y}{\hat{\sigma}} \right)^{-1/\hat{\xi}}$$

(41)

Para  $\hat{\xi} \neq 0$ . Si el *plot* se construye usando (19). El *plot* de los cuantiles se obtiene de nuevo suponiendo que  $\hat{\xi} \neq 0$  y es el par:

$$\left\{ \left( \hat{H}^{-1}(i/(k+1)), y_i \right) \mid i = 1, \dots, k \right\}$$

(42)

Donde

$$\hat{H}^{-1}(y) = u + \frac{\hat{\sigma}}{\hat{\xi}} \left[ y^{-\hat{\xi}} - 1 \right]$$

(43)

Si el modelo Pareto generalizado es un modelo razonable para los excedentes de  $u$ , entonces ambos *plots* (el de la probabilidad y el de los cuantiles) serán aproximadamente lineales.

El *plot* de los niveles de retorno consiste en los puntos  $\{(m, \hat{x}_m)\}$  para valores grandes de  $m$ , donde  $\hat{x}_m$  es el nivel de retorno estimado de la  $m$ -observación.

$$\hat{x}_m = u + \frac{\hat{\sigma}}{\hat{\xi}} \left[ (m \hat{\zeta}_u)^{\hat{\xi}} - 1 \right]$$

(44)

De nuevo, si  $\hat{\xi} = 0$  la ecuación anterior debe modificarse. Es usual graficar la curva de los niveles de retorno en escala logarítmica para enfatizar el efecto de la extrapolación, y añadir intervalos de confianza así como estimaciones empíricas de los niveles de retorno.

### Resultados

La ventaja principal de estos modelos es el uso de más información para hacer el análisis. Se ajustaron dos modelos: *Distribución Pareto Generalizada* (GPD, por sus siglas en inglés) y *Procesos Puntuales*.

Bajo el análisis se introduce la serie completa, es decir con días secos y días húmedos.

En la figura 37-II se muestra el diagrama de dispersión de las observaciones diarias para cada año. Se aprecia que la mayoría de los eventos se distribuyen por debajo de los 20 mm.

En la figura 38-II se muestran los eventos de precipitación diaria para cada mes. En esta imagen se aprecia claramente el ciclo estacional.

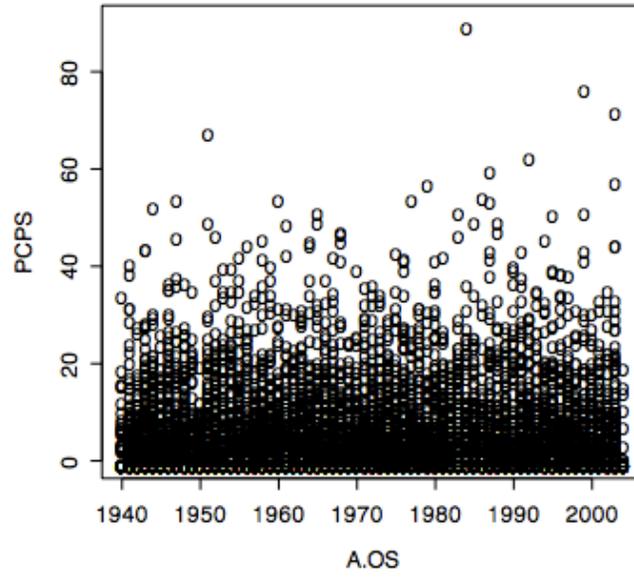


Figura 37-II. Eventos de precipitación diarios.

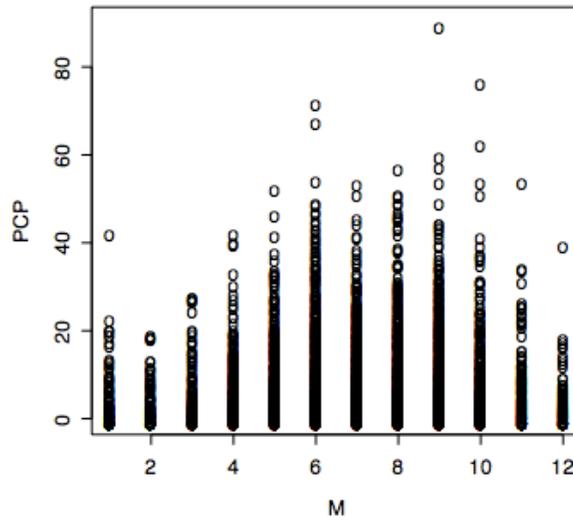


Figura 38-II. Eventos de precipitación diarios para cada mes.

Al introducir los datos se calcula automáticamente la estadística básica de cada columna dentro del archivo, en la tabla 19 se muestra la estadística de la columna de precipitación.

<b>Tabla 19: Estadística Básica</b>	
	<b>PCP</b>
<b>N</b>	23250
<b>Promedio</b>	2.178
<b>Desv. Est.</b>	5.703
<b>Mínimo</b>	0.000
<b>Q1</b>	0.000
<b>Mediana</b>	0.000
<b>Q3</b>	1.000
<b>Máximo</b>	90.0
<b>Valores Faltantes</b>	0

Donde N es la cantidad de valores introducidos. Q1 y Q3, son los cuartiles 1 y 3, que corresponden al primer y tercer cuarto de la distribución. El valor de Q3, que marca  $\frac{3}{4}$  de la distribución, claramente muestra que la mayor parte de los eventos de precipitación diarios se encuentran por debajo de 1.0 mm de precipitación. Esto significa que el último cuarto de distribución es la parte que nos interesa, y es consistente con la *Teoría de Valores Extremos* donde se analizan las colas de la distribución, al tratarse de precipitación nos concentramos únicamente en la cola que representa a los valores máximos.

En este modelo se ajusta primeramente la función Pareto Generalizada a las observaciones que sobrepasan cierto umbral, posteriormente el modelo se valida y finalmente se hacen extrapolaciones.

Lo primero que se hizo fue ajustar la distribución Pareto y el método de optimización fue el de *BFGS*<sup>10</sup>. El umbral que se eligió fue 20.0 mm, que es el que corresponde a el percentil 90 de la distribución.

Se encontró que 567 puntos exceden el umbral y que la tasa promedio de excesos por año es: 8.91. Los parámetros determinados mediante *MLE* se muestran en la tabla 20.

<b>Tabla 20: Parámetros GPD</b>		
	<b>MLE</b>	<b>Error Estándar</b>
<b>Escala (Sigma)</b>	8.6682	0.5203
<b>Forma (Xi)</b>	0.0202	0.0429

El valor positivo del parámetro de forma nos sugiere una distribución sin límites. El intervalo de confianza al 95% es (-0.05,0.11) en su mayor parte tiene valores por encima de cero. Sin embargo contiene a 0, por lo tanto

<sup>10</sup> Ver apéndice.A.10

formalmente no se puede rechazar la posibilidad de que  $\xi = 0$  o incluso que  $\xi < 0$ , al contener también algunos valores negativos.

En la tabla 21 se muestra la matriz covarianza de los parámetros. Las raíces cuadradas de la diagonal de la matriz nos dan el valor de los respectivos errores estándar.

Tabla 21: Covarianza de los Parámetros GPD		
	Escala (Sigma)	Forma (Xi)
Escala (Sigma)	0.2707	-0.0156
Forma (Xi)	-0.0156	0.0018

En la figura 39-II se muestra la salida del ajuste gráficamente.

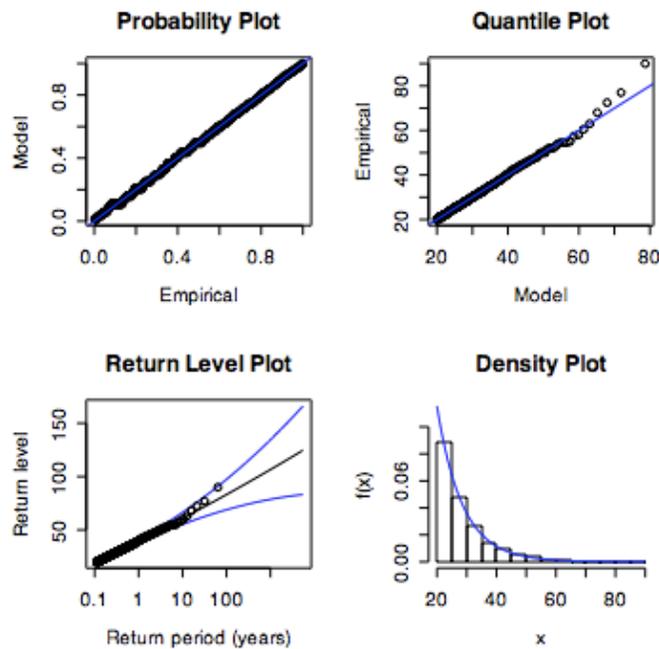


Figura 39-II. Ajuste del modelo GPD.

Al observar el *plot* de probabilidad podemos ver que los puntos se ajustan a una línea, lo que significa que la probabilidad dada por el modelo es muy similar a la probabilidad empírica. En el caso del *plot* de cuantiles el modelo y la distribución empírica no coinciden del todo, especialmente en los cuantiles más grandes, aún así podemos decir que los puntos se distribuyen a lo largo de la diagonal. El resultado de ambos *plots* nos da la confianza de que el ajuste del modelo ha sido el adecuado.

La forma de los periodos de retorno se ajusta a lo esperado dado el valor del parámetro de forma  $\xi$ , positivo y cercano a cero. Esto es, al ser cercano a cero los valores se distribuyen linealmente y al ser ligeramente positivo los valores forman una curva ligeramente convexa.

El nivel de retorno estimado para 100 años es: 83.1156 mm. El periodo de retorno de 100 años tiene un nivel de confianza del 95% en el intervalo (72.3,94.3).

Usando el *Perfil del Logaritmo de la Verosimilitud* el intervalo de confianza al 95% para el parámetro de forma y el periodo de retorno se muestran en las figuras 40-II y 41-II, respectivamente. Se obtienen al dibujar una curva con máximo alrededor del valor estimado para el parámetro de forma. Con máximo en el valor estimado para el periodo de retorno elegido (100 años). Así entonces se observa que el intervalo de confianza estimado para el parámetro de forma es (-0.08,0.13); esto se determina al elegir un límite superior y uno inferior que corresponden a las líneas punteadas, cuando este límite es el adecuado, las líneas punteadas intersectan la curva en el mismo punto que lo hacen las líneas horizontales. Este nuevo intervalo de confianza es más amplio que el anterior de (-0.05,0.11); sin embargo, en la gráfica se aprecia la asimetría del intervalo, que incluye más valores positivos que negativos, incluyendo a 0.

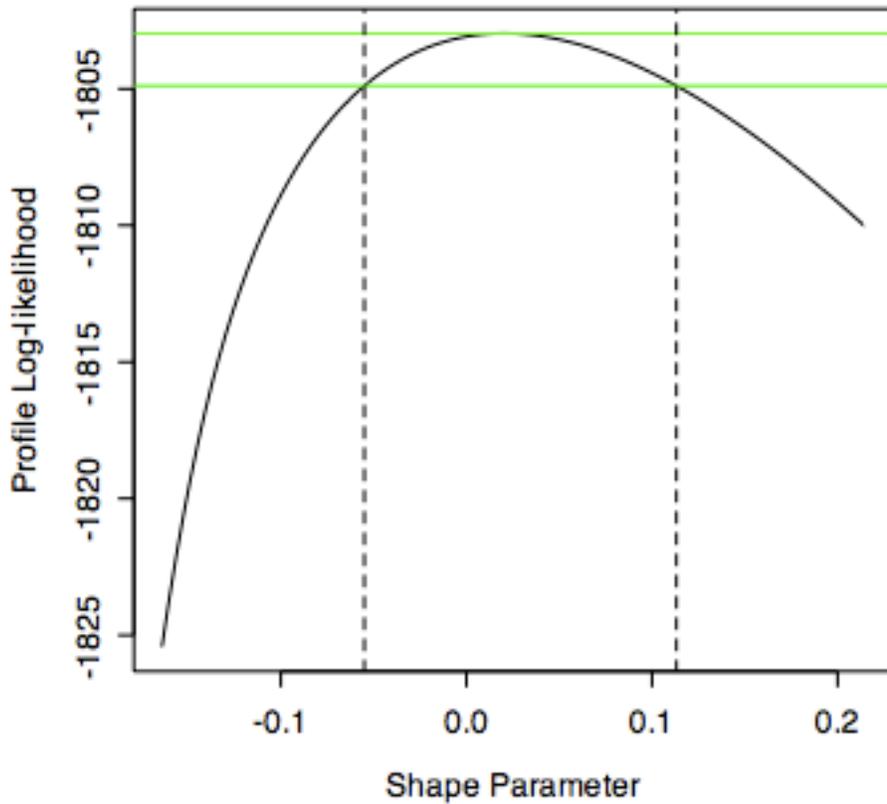


Figura 40-II. Intervalo de confianza del parámetro de forma para GPD.

En el caso del periodo de retorno, el anterior intervalo de confianza era (72.3,94.3), mientras que el nuevo se ha ampliado también (73, 103) lo que significa que se pueden alcanzar valores más grandes que los determinados anteriormente. Ambos intervalos de confianza incluyen el valor máximo de 90 mm. Lo cual es importante si se tienen registros históricos de daños al haberse presentado este u otro valor contenido dentro de los intervalos de confianza.

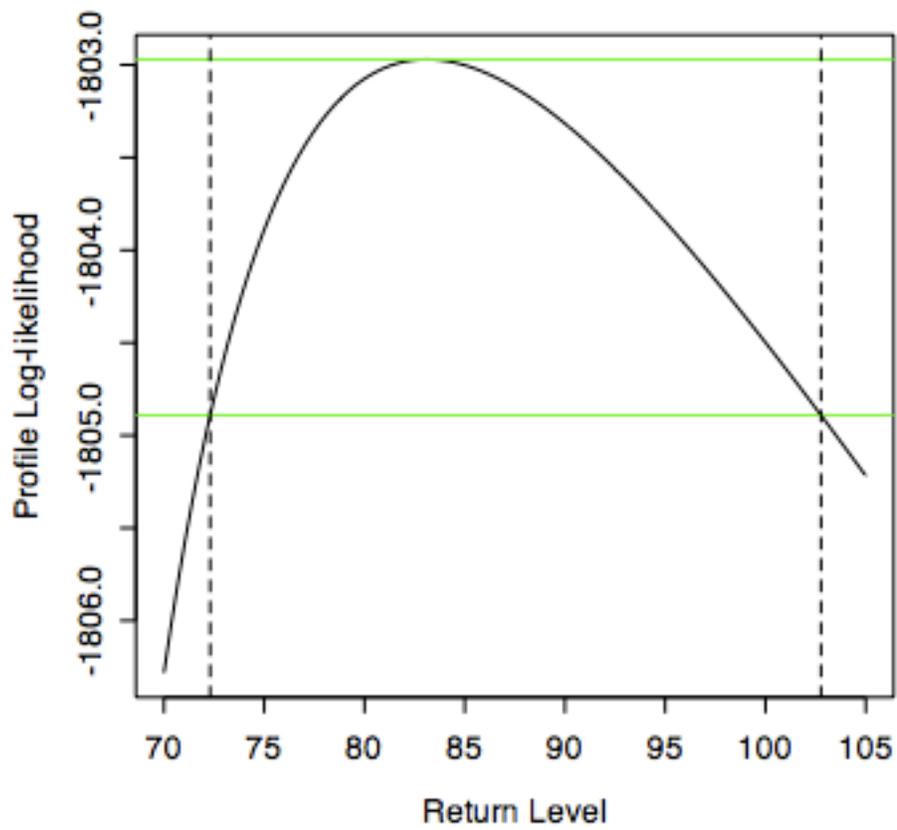


Figura 41-II. Intervalo de confianza del nivel de retorno para un periodo de retorno de 100 años.

## Procesos Puntuales

Existen diferentes formas de caracterizar el comportamiento de los valores extremos de un proceso. En particular esta es una formulación más formal y elegante que se deriva de la *Teoría de Procesos Puntuales (PP)*. Presenta dos ventajas: Primero, da una interpretación del comportamiento de los valores extremos que unifica todos los otros modelos. Segundo, el modelo lleva directamente a una verosimilitud que permite una formulación más natural que la obtenida mediante la Distribución Pareto Generalizada para la excedencia del umbral en el caso de series no estacionarias. Esto es, en esta formulación, las veces que se da la excedencia del umbral y los valores de éstas se combinan en un proceso plasmado en un *plot* bidimensional de las veces que se excede el umbral y valores de los excedentes. La teoría asintótica de los excedentes del umbral muestra que, bajo una normalización razonable, este proceso se comporta como un *Proceso de Poisson No-Homogéneo*<sup>11</sup>.

### Teoría

Un proceso puntual en un conjunto  $A$  es una regla estocástica de ocurrencia y posición de eventos puntuales. En el contexto del modelo  $A$  representa un periodo de tiempo y el proceso puntual es la ocurrencia de un evento. Ya sea que la ocurrencia se estime en cierto periodo de tiempo  $A$ ; o dada la ocurrencia de un evento determinar el periodo de tiempo esperado hasta que se de el siguiente. El conjunto  $A$  puede ser multidimensional.

Una forma de caracterizar las propiedades estadísticas de un proceso puntual es definir un conjunto de variables aleatorias con valores enteros no negativos,  $N(A)$ , para cada  $A \subset A$ , tal que  $N(A)$  es el número de puntos en el conjunto  $A$ . Así, al especificar de manera consistente la distribución de probabilidad para cada  $N(A)$  se determinan las características del proceso puntual, al que llamamos  $\Lambda$ . En particular:

$$\Lambda(A) = E \{N(A)\}$$

(45)

Es el número esperado de puntos en cualquier subconjunto  $A \subset A$ , y se define como medida de intensidad del proceso. Suponiendo  $A = [a_1, x_1] \times \dots \times [a_k, x_k] \subset \mathbb{R}^k$ , la derivada de la función:

$$\lambda(x) = \frac{\partial \Lambda(A)}{\partial x_1 \dots \partial x_k}$$

---

<sup>11</sup> Ver apéndice A.11

(46)

Es la función de intensidad (densidad) del proceso.

El proceso puntual canónico es el *Proceso de Poisson Homogéneo Unidimensional*. Con el parámetro  $\lambda$ , este es un proceso en que satisface:

1. Para toda  $A = [t_1, t_2] \subset \mathbb{R} \Rightarrow N(A): Poi(\lambda(t_1 - t_2))$  (47)

2. Para todos los subconjuntos sin intersección  $A$  y  $B$  de  $\mathbb{R}$ ,  $N(A)$  y  $N(B)$  son variables aleatoria independientes.

En otras palabras, el número de puntos en un intervalos dado sigue la distribución de Poisson, donde la media proporcional a la longitud del intervalo y el número de puntos que ocurren en intervalos separados son mutuamente independientes. El *Proceso de Poisson* con un parámetro  $\lambda$  puede llegar a ser un modelo estocástico apropiado para puntos que ocurren aleatoriamente en el tiempo a la tasa uniforme de  $\lambda$  por intervalo de tiempo. La medida de intensidad correspondiente es  $\Lambda[t_1, t_2] = \lambda(t_1 - t_2)$  y la función de densidad de intensidad es  $\lambda(t) = \lambda$ .

El *Proceso de Poisson Homogéneo* puede generalizarse a un modelo de puntos que ocurren aleatoriamente en el tiempo, pero a una tasa variable  $\lambda(t)$ . Esto es el *Proceso de Poisson Unidimensional No-Homogéneo*, que tiene la misma propiedad de cuentas independientes en subconjuntos que no se intersectan, sólo se modifica la propiedad de que para toda  $A = [t_1, t_2] \subset \mathbb{R} \Rightarrow N(A): Poi(\Lambda(A))$ , donde:

$$\Lambda(A) = \int_{t_1}^{t_2} \lambda(t) dt$$

(48)

Implícitamente la medida de intensidad y la función de densidad son  $\Lambda(\cdot)$  y  $\lambda(\cdot)$ , respectivamente.

El *Proceso de Poisson No-Homogéneo* se puede generalizar a puntos que ocurren en un subconjunto  $k$ -dimensional del espacio. Además existe una propiedad intrínseca en los *Procesos de Poisson* donde los puntos ocurren independientemente unos de otros, por lo que lo hace un modelo ideal para dispersiones aleatorias. Si embargo es un modelo pobre para procesos donde la dispersión de los puntos sigue una regla natural, tal es el caso de árboles en un bosque o tormentas, ya que aunque variaciones en el número de puntos en diferentes sub-regiones del espacio son admitidas a través de una función de intensidad de medida no constante, ésta puede ser consecuencia del número promedio de puntos en algunas regiones comparadas con otras.

### Ley de convergencia

Tal como se ha hecho en los modelos anteriores, sea  $X_1, X_2, \dots$  una serie de variables aleatorias independientes e idénticamente distribuidas, con una función de distribución común  $F$ . Suponemos que las  $X_i$  se comportan adecuadamente en el sentido de valores extremos. Esto es, con  $M_n = \max \{X_1, X_2, \dots\}$ , una secuencia de constantes  $\{a_n > 0\}$  y  $\{b_n\}$  tal que:

$$\Pr \left\{ \frac{(M_n - b_n)}{a_n} \leq z \right\} \rightarrow G(z)$$

(49)

Con:

$$G(z) = \exp \left\{ - \left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}$$

(50)

Para los parámetros  $\mu, \sigma > 0$  y  $\xi$ . Ahora definimos una secuencia de Procesos Puntuales  $N_n$  en  $\mathbb{R}^2$  como:

$$N_n = \left\{ \left( \frac{i}{n+1}, \frac{(X_i - b_n)}{a_n} \right) : i = 1, \dots, n \right\}$$

(51)

Se re-escala la primera ordenada para asegurar que el eje temporal siempre tenga valores en el intervalo (0,1). Al re-escalar la segunda ordenada se establece el comportamiento de los extremos cuando  $n \rightarrow \infty$ . El resultado final es que en las regiones de la forma  $[0,1] \times [u, \infty)$ ,  $N_n \xrightarrow{d} N$  mientras  $n \rightarrow \infty$ , donde  $N$  es un *Proceso de Poisson No-Homogéneo*.

Así entonces si consideramos una región de la forma  $A = [0,1] \times (u, \infty)$  para algún valor grande de  $u$ . Entonces, cada uno de los  $n$  puntos de  $N_n$  tiene la probabilidad  $p$  de caer dentro de  $A$ , donde:

$$p = \Pr \left\{ \frac{(X_i - b_n)}{a_n} > u \right\} \approx \frac{1}{n} \left[ 1 + \xi \left( \frac{u - \mu}{\sigma} \right) \right]^{-1/\xi}$$

(52)

Como los  $X_i$  son mutuamente independientes,  $N_n(A)$  tiene una distribución binomial:

$$N_n(A) : \text{Bin}(n, p)$$

(53)

Por la convergencia estándar de la distribución binomial en el límite de la distribución Poisson. La distribución límite de  $N_n(A)$  cuando  $n \rightarrow \infty$  es  $\text{Poi}(\Lambda(A))$  con:

$$\Lambda(A) = \left[ 1 + \xi \left( \frac{u - \mu}{\sigma} \right) \right]^{-1/\xi}$$

(54)

De la homogeneidad del proceso en la dirección del tiempo, se sigue que, para cualquier región de la forma  $A = [t_1, t_2] \times (u, \infty)$  con  $[t_1, t_2] \subset [0, 1]$  la distribución límite de es también , donde.

$$\Lambda(A) = (t_2 - t_1) \left[ 1 + \xi \left( \frac{u - \mu}{\sigma} \right) \right]^{-1/\xi}$$

(55)

Este límite de Poisson para todos esos conjuntos  $A$  , junto con el hecho de que las distribuciones de  $N(A)$  en en conjuntos sin intersección están determinados a ser independientes por construcción, es suficiente para establecer el Proceso de Poisson límite, con una intensidad de medida de los procesos límite dada por la ecuación anterior.

## Resultados

Ahora usando el método de Puntos Sobre Umbral, para el umbral de 20.0 mm y con el método de optimización de *Nelder-Mead*<sup>12</sup>, obtenemos los parámetros que se muestran en la tabla 22. En la tabla 23 se muestra la matriz de covarianza de los parámetros. El logaritmo negativo de la verosimilitud fue: 1130.01.

---

<sup>12</sup> Ver apéndice A.5.

<b>Tabla 22: Parámetros Procesos Puntuales</b>		
	<b>MLE</b>	<b>Error Estándar</b>
<b>Escala (sigma)</b>	9.0640	0.9193
<b>Forma (xi)</b>	0.0207	0.0429
<b>Localización (mu)</b>	39.3815	0.9119

<b>Tabla 23: Covarianza de los Parámetros Procesos Puntuales</b>			
	<b>Localización (mu)</b>	<b>Escala (Sigma)</b>	<b>Forma (Xi)</b>
<b>Localización (mu)</b>	0.8451	0.3718	0.0044
<b>Escala (Sigma)</b>	0.3733	0.3718	0.0202
<b>Forma (Xi)</b>	0.0044	0.0202	0.0018

Al igual que en la matriz de covarianza anterior, las raíces cuadradas de los números en la diagonal dan los respectivos valores del error estándar para los parámetros de la tabla 22.

En la figura 42-II se muestra el ajuste del modelo, el cual, a juzgar por el *plot* de probabilidad ha sido bueno, ya que los valores se distribuyen a lo largo de la línea con pendiente unitaria. En la figura 43-II se aprecia al histograma.

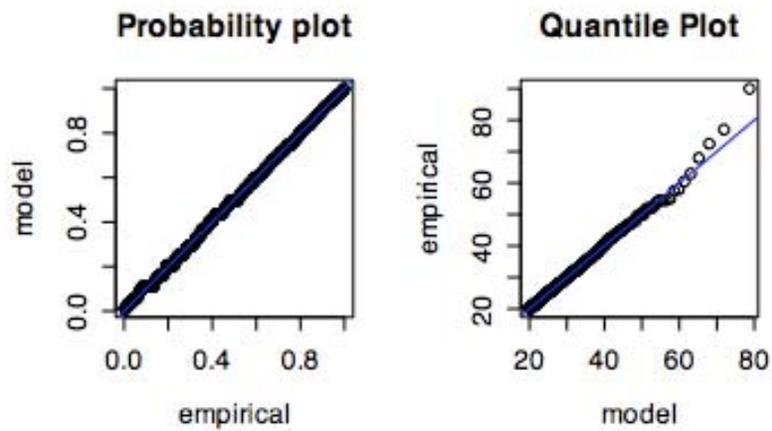


Figura 42-II. Ajuste del modelo de Procesos Puntuales.

Debido a que los puntos se distribuyen sobre la diagonal podemos decir que el modelo es un buen ajuste de los datos introducidos.

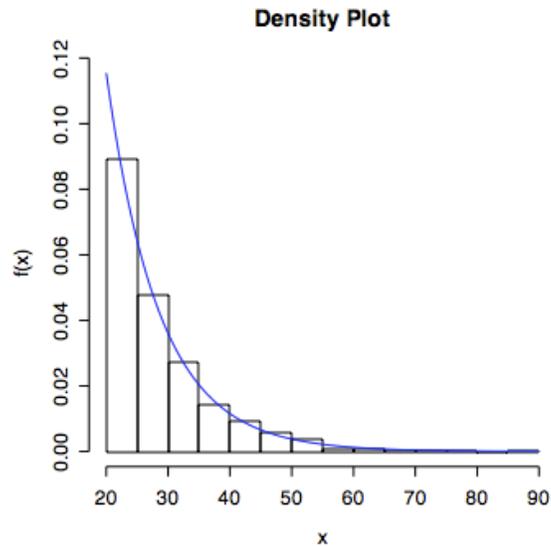


Figura 43-II. Histograma Procesos Puntuales.

Como se mencionó anteriormente la ventaja del Modelo de Procesos Puntuales radica en que se puede obtener un ajuste a la distribución GEV haciendo algunas transformaciones en los parámetros.

Usando el método de optimización de BGFS-quasi Newton se hacen las transformaciones necesarias a los parámetros para obtener un ajuste a GEV. El umbral sigue siendo 20.0 mm. El logaritmo negativo de la verosimilitud es 1130.1. En las tablas 24 y 25 se muestran los parámetros y la matriz de covarianza de estos, respectivamente.

**Tabla 24: Parámetros GEV**

	<b>MLE</b>	<b>Error Estándar</b>
<b>Escala (sigma)</b>	9.0596	0.6086
<b>Forma (xi)</b>	0.0202	0.0429
<b>Localización (mu)</b>	39.3809	0.9188

**Tabla 25: Covarianza de los Parámetros GEV**

	<b>Localización (mu)</b>	<b>Escala (Sigma)</b>	<b>Forma (Xi)</b>
<b>Localización (mu)</b>	0.8441	0.3723	0.0043
<b>Escala (Sigma)</b>	0.3723	0.3704	0.0201
<b>Forma (Xi)</b>	0.0043	0.0201	0.0018

En la figura 44-II se muestra el ajuste del modelo y en la 45-II el histograma.

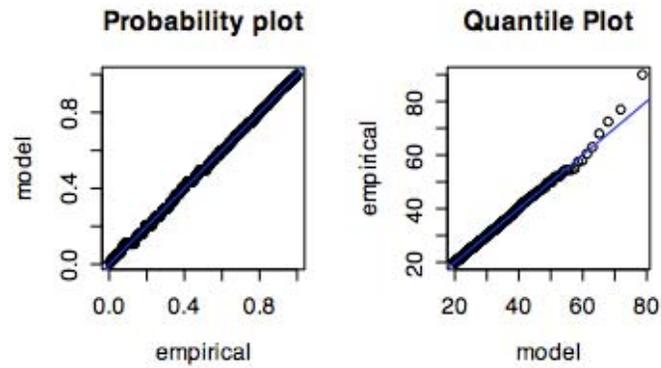


Figura 44-II. Ajuste de l Modelo de Puntos sobre umbral con el método de optimización de BFGS.

Una vez más el ajuste del modelo fue el adecuado, ya que la mayoría de los valores se distribuyen en la diagonal.

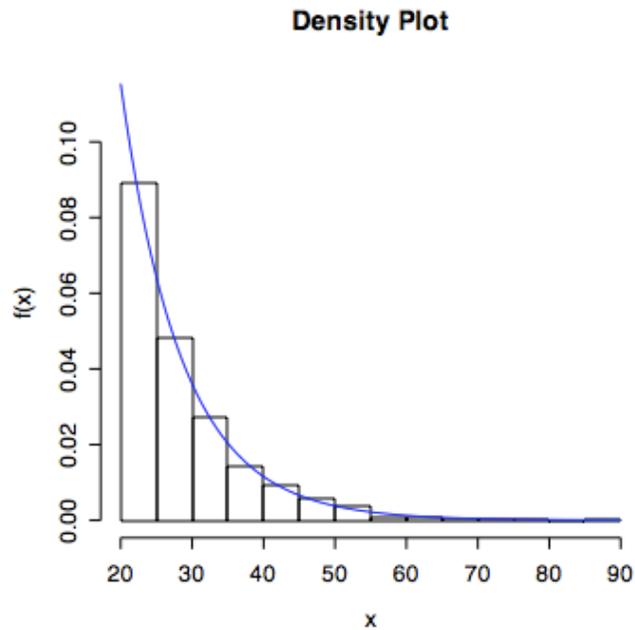


Figura 45-II. Histograma modelo GEV.

### Comparación

Los parámetros obtenidos para GPD con el método de optimización BFGS son: Escala  $\sigma = 8.6682$  (0.5203) y Forma  $\xi = 0.0202$  (0.0429)

Mientras que los obtenidos mediante Procesos Puntuales con el mismo método de optimización son:

Localización  $\mu = 39.3809$  (0.9188), Escala  $\sigma = 9.0596$  (0.6086) y Forma  $\xi = 0.0202$  (0.0429).

Para ambos casos el umbral fue de 20.0 mm. Los parámetros de forma y escala,  $\xi$  y  $\sigma$ , son muy similares en ambos métodos, y las diferencias están dentro del rango del error estándar. Como se ha mencionado una de las ventajas del de Procesos Puntuales es que se puede obtener una GEV a partir de los parámetros obtenidos.

### Serie No Estacionaria

Cuando la serie no es estacionaria o incluye algún tipo de tendencia, ya sea un ciclo anual o estacional, como la precipitación, es necesario hacer algunos cambios en el análisis.

Uno de los métodos es incluir la tendencia tanto en los parámetros de escala o de localización, permitiendo que estos cambien en el tiempo. Variar el parámetro de forma no es muy recomendable, aunque es posible, ya que no se puede modelar una función que varíe suavemente y por lo mismo estimar los parámetros no es algo sencillo. Una posibilidad es permitir que varíe en una escala mayor, por ejemplo ajustando una distribución para cada estación. Cambios en el parámetro de localización, para la GEV, significarían que la media de la distribución ha cambiado, o en otras palabras, que existe cierta tendencia en la media de la distribución con base en la cual se están dando los eventos extremos.

Para el caso de la *Distribución Pareto Generalizada*,  $GP(\sigma, \xi)$ , cambios en el parámetro de escala,  $\sigma$ , se traducirían en cambios en el valor mínimo de la distribución, que en el caso de precipitación y análisis de eventos extremos serían cambios en el umbral a partir del cual se van a tomar los datos para ajustar el modelo; o como es el caso del ciclo estacional, donde el umbral cambia según la estación. Mientras que cambios en el parámetro de forma,  $\xi$ , serían justamente cambios en la forma en que estos se distribuyen, lo que podría significar, por ejemplo, más eventos de precipitación intensa. Esto entonces se podría aplicar si ya sabemos que existe una tendencia importante en la frecuencia de dichos eventos. Aunque este no es el caso de nuestros datos.

Por ejemplo cambios en el parámetro de escala se expresan en modelos con exponenciales, esto para asegurar que  $\sigma > 0$  para todos los valores, o lo que es lo mismo:  $\sigma(t) > 0 \forall t$ . Para este caso, *ExtRemes* permite obtener automáticamente los logaritmos y con ellos expresar el modelo en forma lineal, donde el objetivo es encontrar los coeficientes adecuados y con ellos incluirlos ya sea directamente en el umbral o en el parámetro.

En el siguiente caso introducimos la tendencia del ciclo anual en el parámetro de escala,  $\sigma$ , pero para ello tenemos que hacer algunas transformaciones a los datos y añadiendo éstas como nuevas columnas en el archivo.

La primera transformación consiste en asignar 1 a los datos que estén por encima del umbral y 0 a todos los demás.

Sean  $X_1, X_2, \dots$  observaciones diarias y denotamos  $s(t)$  como la estación en la que el día  $t$  cae. Para cada estación habrá umbrales diferentes  $u_s(t)$ . Si  $GP(\sigma, \xi)$  entonces el modelo se expresa como:

$$\left( X_t - u_{s(t)} \mid X_t > u_{s(t)} \right): GP(\sigma_{s(t)}, \xi_{s(t)})$$

donde  $(\sigma_{s(t)}, \xi_{s(t)})$  son los parámetros de la distribución Pareto Generalizada en la estación  $s(t)$ .

Así entonces es necesario distinguir los valores por encima del umbral del resto de la serie, marcándolos con la etiqueta 1 y 0 para los demás.

La segunda es una transformación trigonométrica con el objeto de establecer el ciclo anual. Esta transformación

añade dos columnas al archivo, una con la función  $\sin\left(\frac{2\pi t}{365.25}\right)$  y la otra con la función  $\cos\left(\frac{2\pi t}{365.25}\right)$  donde  $t = 1, 2, \dots, 365.24$

así se crean las columnas obs.sin365 y obs.cos365 respectivamente.

Dicha transformación se hace sobre una columna donde se muestren el número de observaciones.

Ahora incorporaremos el ciclo en algunos parámetros del modelo *Poisson-GP*, para ello lo primero que hay que hacer es ajustar el parámetro  $\lambda$  de Poisson en función del tiempo, de hecho lo que se pretende es establecer la forma lineal del modelo:

$$\log \lambda(t) = \beta_0 + \beta_1 \sin\left(\frac{2\pi t}{365.25}\right) + \beta_2 \cos\left(\frac{2\pi t}{365.25}\right)$$

donde las funciones trigonométricas son reemplazadas por las respectivas columnas.

El parámetro  $\lambda$  es el número esperado de ocurrencias en un cierto tiempo.

Para los datos de precipitación no fue posible obtener ningún resultado para este método.

En este otro caso ajustamos GPD con el mismo ciclo anual igualmente incorporándolo en el parámetro de escala, pero en este caso como covariante, esto es, como si se tratara de explicar el comportamiento de la serie mediante un agente externo, en este caso el mismo ciclo anual.

El parámetro de escala se modeló de la siguiente manera:

$$\log \sigma(t) = \sigma_0 + \sigma_1 \sin\left(\frac{2\pi t}{365.25}\right) + \sigma_2 \cos\left(\frac{2\pi t}{365.25}\right)$$

En la tabla 26 se muestran los parámetros obtenidos y que corresponden a los coeficientes del polinomio. En la 27 se muestra la matriz de covarianza. La verosimilitud logaritmo-negativa es: 1802.3.

**Tabla 26: Parámetros Serie No-Estacionaria**

	MLE	Error Estándar
log Scale	2.1370	0.0928
OB.sin365 (beta1)	0.0287	0.0833
OB.cos365 (beta2)	0.0796	0.0971
Forma (xi)	0.0129	0.0434

**Tabla 27: Covarianza de los Parámetros Serie No-Estacionaria**

	<b>log Scale</b>	<b>OB.sin365 (beta1)</b>	<b>OB.cos365 (beta2)</b>	<b>Forma (xi)</b>
<b>log Scale</b>	0.0086	0.0049	-0.0061	-0.0019
<b>OB.sin365 (beta1)</b>	0.0049	0.0069	-0.0040	-0.0002
<b>OB.cos365 (beta2)</b>	-0.0061	-0.0040	0.0094	-0.0001
<b>Forma (xi)</b>	-0.0019	-0.0002	-0.0001	0.0019

En la figura 46-II se muestra el ajuste del modelo.

### Residual Probability Plots and Quantile Plot (Exptl.)

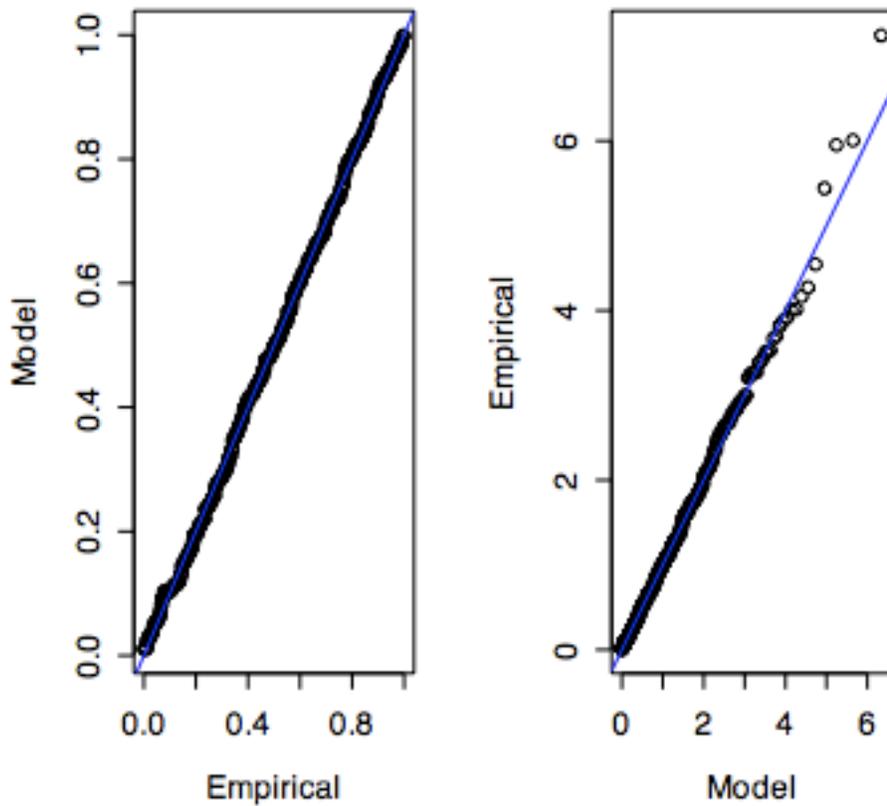


Figura 46-II. Ajuste del Modelo GPD introduciendo un ciclo anual en el parámetro de escala. Se observa que la mayoría de los puntos se distribuyen a o largo de la diagonal, lo que significa que el ajuste del modelo es adecuado.

Determinar parámetros que varíen con el ciclo anual es un avance importante ya que nos permite trabajar con la distribución completa.

# III-Conclusiones

La meta de este proyecto fue desarrollar un método para estimar cómo han cambiado en intensidad, frecuencia y duración los eventos extremos de precipitación diarios. Además de ajustar modelos para saber cómo se comportarán éstos en el futuro a partir de los eventos observados. En conjunto este análisis será una herramienta para la adaptación y la toma de decisiones presentes y futuras.

El análisis completo se llevó a cabo en tres partes. La primera consistió en caracterizar la climatología local; es decir la determinación de la precipitación en el periodo normal 1961-1990, la obtención de las anomalías de precipitación respecto a este periodo y la determinación de las distribuciones mensuales y quincenales. Este primer acercamiento nos permite tomar decisiones futuras dentro del análisis, tal como ocurrió al observar el ciclo anual de la precipitación y así diferenciar los meses húmedos de los meses secos. En esta primera parte también se hizo un breve análisis de la relación entre las máximas de precipitación y El Niño Oscilación Sur, ENOS; el resultado fue que estas máximas en las series de tiempo de todos los meses ocurrieron en años neutros, y sólo en algunos meses se tuvieron eventos extremos en años de Niño o Niña, por lo que no podemos concluir que el ENOS sea un fenómeno que influya en la precipitación local.

La segunda parte consistió en la determinación de once índices de precipitación, de donde podemos concluir que: El número máximo de días consecutivos con muy poca precipitación ha disminuido, mientras que el número máximo de días con precipitación mayor o igual a 1mm aumentó. La precipitación anual total disminuyó muy poco. El número de días con precipitación mayor a 20 mm parece haber aumentado, por tanto parece ser que los eventos extremos están aumentando su frecuencia, aunque esto depende de la tendencia que se tome, la tendencia polinomial en realidad muestra una ligera disminución a partir de los 80. La precipitación anual total mayor al 95% parece haber aumentado en la tendencia lineal y disminuido ligeramente en la polinomial, lo mismo ocurre con el percentil 99. La tendencia lineal muestra un aumento de la máxima mensual en un día de precipitación. Y por ultimo, la precipitación anual total dividida entre el número de días húmedos (precipitación mayor o igual a 1 mm) en un año muestra una disminución en ambas tendencias. Estos resultados muestran que la media ha cambiado muy poco, mientras que el cambio en la variabilidad es mayor. Lo anterior es especialmente importante

dado que esta región presenta graves problemas de fertilidad en el suelo y eventos extremos de precipitación contribuyen en gran medida a la pérdida de los pocos nutrientes del suelo. Al mismo tiempo la mayor parte de la producción agrícola son cultivos de temporal, por lo que cambios en la variabilidad de la precipitación representan pérdidas importantes en este rubro. Por tanto, para esta región los cambios en la variabilidad de la precipitación son más importantes que los cambios en la media.

En la tercera parte se introdujo la Teoría de Valores Extremos para llevar a cabo un análisis más riguroso de los cambios en extremos. Se usaron dos tipos de modelos: Asintóticos y de Umbral. En los modelos asintóticos se usó el método de bloques ajustando la distribución General de Valores Extremos (GEV) a los eventos de precipitación máxima de cada mes durante todo el periodo. El resultado más importante es que el modelo sólo se pudo ajustar exitosamente a meses húmedos (Abril a Octubre), mientras que para meses secos el ajuste del modelo fue pobre; la información más importante se observa en los niveles de retorno para el periodo de cien años que muestran que es muy probable (95% de confianza) que se presenten eventos extremos de más de 90 mm, máximo observado hasta ahora, en los meses de septiembre e incluso casi 100 mm en octubre. En los modelos de Umbral se usaron dos métodos: distribución Pareto Generalizada (GPD) y Procesos Puntuales (PP). El umbral usado fue de 20 mm, correspondiente al percentil 90 de la distribución. El modelo se ajustó exitosamente en el caso de la distribución GPD devolviendo un nivel de retorno para 100 años que incluye el valor máximo de 90 mm, llegando incluso a más de 100 mm, lo cual es importante si eventos como éste han causado algún problema en la región. En el caso de los de los Procesos Puntuales el modelo se ajustó exitosamente también.

Es importante notar que en el análisis se consideró que los forzantes permanecían constantes, esto significa que el cambio climático no se ha tomado en cuenta. Sin embargo, cambios en los eventos extremos pueden ser un indicador de cambio climático, por lo que sería una gran mejoría incluir cambios en los forzantes climáticos en futuros trabajos. Los cambios observados hasta el momento son una buena aproximación de los efectos que posibles eventos futuros puedan tener. Esto significa que se deben desarrollar planes que al menos disminuyan la vulnerabilidad actual que presenta la zona de interés.

# Referencias

- Beniston, M. S., D. B., 2004: Extreme climatic events and their evolution under changing climatic conditions. *Global and Planetary Change*, **44**, 1-9.
- Cavazos, T. R., D., 2004: Variability of extreme precipitation events in Tijuana, México. *Climate Research*, **25**, 229-243.
- Coles, S., 2001: An Introduction to Statistical Modeling of Extreme Values. *Springer Series in Statistics*, Springer-Verlag London.
- Conde, C., Ferrer, R., Orozco, S., 2006: Climate Change and Climate Variability impacts on rainfed agricultural activities and possible adaptation measures. A Mexican case of study. *Atmósfera*, **19(3)** 181-194.
- Goodess, C., Hulme M. y Osborn, T., 2001: The identification and evaluation of suitable scenario development methods for the estimation of future probabilities of extreme weather events.
- Houghton, J., 1994. *Global Warming: The Complete Briefing*. Cambridge, Cambridge University Press.
- INEGI (Instituto Nacional de Estadística, Geografía e Informática), 1996. Atlas Agropecuario del Estado de Tlaxcala. VII Censo Agropecuario. 1991. 80 pp.
- — — 2005: Anuario Estadístico Tlaxcala. 3 - 14 pp.
- IPCC, 2001: *Climate Change. The Scientific Basis Vol. I*, Cambridge University Press, 881 pp.
- — —, 2007: *Climate Change. The Physical Science Basis Vol. I*, Cambridge University Press.
- Karl T. R., K. R. W., Easterling D. R. y Quayle R. G., 1996: Indices of Climate Change for the United States. *Bulletin of the American Meteorological Society*, **77 No.2**, 279-292.
- Karl T. R., E. D. R., 1999: Climate extremes: Selected review and future research directions. *Climatic Change*, **42**, 309-325.
- Katz, R. W. B., B. G., 1992: Extreme events in a changing climate: variability is more important than average. *Climatic Change*, **21**, 289-302.
- Magaña, V., J.A. Amador y S. Medina, 1999: The mid-summer drought over Mexico and Central America. *J. Climate*, **12**, 1577-1588.
- Magaña, V. (editor). 1999. *Los Impactos de El Niño en México*. México. UNAM, IAI; SG. 228 pp. <http://ccaunam.atmosfcu.unam.mx/cambio/nino.htm>.
- Reyes Coca, Sergio 2002: *Introducción a la Meteorología*. UABC. 428 pp.
- Salby Murry L., 1996: *Fundamentals of Atmospheric Physics. Vol 61*, *International Geophysics Series*, Academic Press, 204 pp.
- Saldaña Zorrilla, Sergio O., 2007. Socio-Economic Vulnerability to Natural Disasters in Mexico: rural poor, trade and public response. *Serie Estudios y Perspectivas de la CEPAL*, No. 92. LC/MEX/L.819. ISSN: 1680-8800. ISBN: 978-92-1-121661-5. Comisión Económica para América Latina y el Caribe (CEPAL). Organización de las Naciones Unidas (ONU). México, DF.

- SEGOB, 1988: Los Municipios de Tlaxcala. Vol. 29, *Colección: Enciclopedia de los Municipios de México*, 248 pp.
- SEMARNAT (Secretaría de Medio Ambiente y Recursos Naturales) 1995. Estadísticas Selectas. Agua. Balance de agua superficial y subterránea. [www.semarnap.gob.mx/naturaleza/estadistica-am](http://www.semarnap.gob.mx/naturaleza/estadistica-am).
- Storch, H. v., 1999: Statistical analysis in climate research. New York, *Cambridge University Press*. 197-215 pp.
- Wilks, D. S., 1995: Statistical Methods in the Atmospheric Sciences. Vol. 59, *International Geophysics Series*, Academic Press, 444 pp.

# Apéndice

## A.1 Términos

- Virtualmente cierto: > 99% de probabilidad de ocurrencia.
- Extremamente probable: >95%.
- Muy Probable: > 90%.
- Probable: > 66%.
- Más probable que no: > 50%
- No probable: < 33%.
- Muy improbable: < 10%
- Extremamente improbable: < 5%.

## A.2 Temperatura de Brillo

Es una medida de la intensidad de radiación térmica emitida por un objeto, está dada en unidades de temperatura porque es una correlación entre la intensidad de radiación emitida y la temperatura física de un objeto radiante, que está dada por Stefan-Boltzmann.

## A.3 Absorción y Emisión en la Atmósfera

Las moléculas pueden almacenar energía en estas diferentes formas:

Trasnacional: energía cinética como resultado del movimiento.

Rotacional: Movimiento rotacional con respecto al eje en su centro de gravedad.

Vibracional: Los átomos están unidos entre sí por fuerzas que actúan como resortes, de manera que éstos se pueden mover alrededor de una posición de equilibrio.

Electrónica: Los electrones cambian su nivel de energía.

Los tres últimos están cuantizados.

## A.4 Cambio de Temperatura por Forzamiento Radiativo

Suponemos el sistema en equilibrio, bajo condiciones naturales. Al perturbarlo comenzará a oscilar hasta alcanzar un nuevo equilibrio. La respuesta a la perturbación o forzamiento radiativo, tendrá un efecto en la temperatura, es decir un cambio  $\Delta T_s$ .

La relación entre el cambio de temperatura superficial que sufrirá el sistema y el forzante está dada mediante:  $\Delta T_s = \lambda RF$

Donde  $\lambda$  es el parámetro de sensibilidad climática con  $\lambda = T_s/4F_{TOA}$

Esto significa que el sistema responderá de manera diferente a un mismo forzamiento si la sensibilidad climática es diferente, esta puede incluir, cambios en el albedo o retroalimentaciones. En general

$$\lambda = -\left(\frac{dFR}{dT_s}\right)^{-1}$$

Por ejemplo podemos obtener la respuesta al forzamiento radiativo para el CO<sub>2</sub>:

Con  $T_s = 288$  K y  $F_{TOA} = 235$  Wm<sup>-2</sup>

$$\lambda = 288 \text{ K}/4(235 \text{ Wm}^{-2}) = 0.3064 \text{ K/Wm}^{-2}$$

Con  $RF = 1.6$  Wm<sup>-2</sup>

$$\Delta T_s = (0.3064 \text{ K/Wm}^{-2})RF$$

$$\Delta T_s = (0.3064 \text{ K/Wm}^{-2})(1.6 \text{ Wm}^{-2}) = 0.49 \text{ K}$$

Finalmente  $\Delta T_s = 0.49$  K

sería la respuesta a el forzamiento radiativo de 1.6 Wm<sup>-2</sup>.

## A.5 R

R es la versión de S-Plus en software libre. Se obtiene de la siguiente dirección:

<http://www.r-project.org>

## A.6 Rclimdex

RClimDex es una interfase gráfica que funciona sobre la plataforma de R y que calcula 27 índices de eventos extremos basados en las funciones de Coles (2001). Fue desarrollado por Xuebin Zhang y Yang Feng del Climate Research Branch.

El software se obtiene de la siguiente dirección:

<http://cccma.seos.uvic.ca/ETCCDMI/software.shtml>

## A.7 ExtRemes

ExtRemes es una herramienta diseñada para facilitar la aplicación de la Teoría de Valores Extremos en estudios climáticos que involucren eventos extremos, usa como base teórica las funciones de Stuart Coles (Coles 2001). Funciona sobre la plataforma de R, generalmente se obtiene al instalar R. Es necesario cargar el programa cada vez que se acceda a R. Instrucciones para instalarlo y archivos de ejemplos se encuentran en la siguiente dirección:

<http://www.isse.ucar.edu/extremevalues/evtk.html>

## A.8 Diagrama de Caja

También son conocidos como *Boxplots* por su nombre en inglés. Este diagrama es una representación sintética de una distribución. Se construye calculando 3 valores básicos: El primer cuartil ( $Q_1$ ), la mediana y el tercer cuartil ( $Q_3$ ). La caja representa el espacio entre el primer cuartil y el tercer cuartil donde la línea intermedia representa la mediana, el ancho de la caja es también llamado Rango Inter-Cuartilico (IQR). Las líneas que salen de la caja se obtienen calculando lo siguiente:  $Q_1 - 1.5 \cdot \text{IQR}$  y  $Q_3 + 1.5 \cdot \text{IQR}$ . Así entonces los valores atípicos u *outliers* son aquellos que se encuentran por encima o por debajo de los valores antes calculados.

## A.9 Perfil del Logaritmo de la Verosimilitud

Se puede obtener mayor exactitud en el cálculo de los intervalos de confianza, al trabajar con el *Perfil de Verosimilitud*.

El logaritmo de la verosimilitud para  $\theta$  se escribe formalmente como  $\ell(\theta_i, \theta_{-i})$ , donde  $\theta_{-i}$  denota a los componentes de  $\theta$  excluyendo a  $\theta_i$ . El *perfil del logaritmo de la verosimilitud* para  $\theta_i$  se define como:

$$\ell_p(\theta_i) = \max_{\theta_{-i}} \ell(\theta_i, \theta_{-i})$$

Esto significa que para cada valor de  $\theta_i$ , el perfil del logaritmo de la verosimilitud es el logaritmo de la verosimilitud maximizado con respecto a los otros componentes de  $\theta$ . Es decir,  $\ell_p(\theta_i)$  es el perfil de la superficie del logaritmo de la verosimilitud visto desde el eje  $\theta_i$ .

Esta definición se puede generalizar a la situación donde  $\theta$  se puede descomponer en  $(\theta^{(1)}, \theta^{(2)})$ , donde  $\theta^{(1)}$  es el vector  $k$ -dimensional que nos interesa y  $\theta^{(2)}$  corresponde a los  $(d - k)$  componentes restantes. Así entonces el perfil del logaritmo de la verosimilitud para  $\theta^{(1)}$  se define como:

$$\ell_p(\theta^{(1)}) = \max_{\theta^{(2)}} \ell(\theta^{(1)}, \theta^{(2)})$$

Si  $k = 1$  entonces (pag 35 Coles) se reduce a (ecuación anterior).

Este resultado nos da un procedimiento para aproximar las inferencias en el MLE de  $\theta^{(1)}$ .

Así entonces,  $x_1, \dots, x_n$  sea realizaciones independientes de una distribución con una familia paramétrica  $F$ , y sea  $\hat{\theta}_0$  el estimador de máxima verosimilitud del parámetro  $\theta_0 = (\theta^{(1)}, \theta^{(2)})$  del modelo  $d$ -dimensional, donde es el subconjunto  $k$ -dimensional de  $\theta_0$ . Entonces, bajo condiciones de regularidad aceptables y para  $n$  grande:

$$D_p(\theta^{(1)}) = 2 \left\{ \ell(\hat{\theta}_0) - \ell_p(\theta^{(1)}) \right\} \sim \chi_k^2$$

Finalmente, el intervalo de confianza, para un sólo componente  $\theta_i$ , es:

$$C_\alpha = \left\{ \theta_i : D_p(\theta_i) \leq c_\alpha \right\}$$

Donde  $c_\alpha$  es el cuantil  $(1 - \alpha)$  de la distribución  $\chi_1^2$ .

## A.10 Métodos de Optimización

Broyden–Fletcher–Goldfarb–Shanno (BFGS): Es un método para resolver problemas de optimización no lineales sin restricciones. Es un método derivado del método de optimización de Newton, que es una técnica de optimización que busca puntos estacionarios de una función, o sea, cuando el gradiente es 0. El método de Newton asume que la función puede ser localmente aproximada mediante expansiones de Taylor cuadráticas alrededor del punto óptimo y usando la primera y segunda derivadas para encontrar el punto estacionario.

Nelder-Mead: Es un método de optimización también conocido como método simple cuesta-abajo o método de amiba. Es un algoritmo comúnmente utilizado en problemas de optimización no-lineales. Es un método numérico para minimizar una función en un espacio de varias dimensiones. El método trata de encontrar soluciones optimas aproximadas en un problema con N variables cuando la función varía suavemente. El algoritmo genera nuevas posiciones prueba extrapolando el comportamiento de la función en cada punto prueba. Así entonces el algoritmo decide sustituir uno de los puntos prueba con este nuevo punto.

### A.11 Procesos de Poisson

En los procesos de Poisson los eventos individuales son independientes en el sentido de que su probabilidad de ocurrencia en un intervalo dado depende sólo del tamaño del intervalo. El número de ocurrencias no depende de dónde en el tiempo se localiza el intervalo o que tan frecuentes han sido en otros intervalos sin intersección. Así entonces los eventos de Poisson ocurren aleatoriamente, pero a una tasa promedio. La secuencia de estos eventos es lo que se conoce como un proceso de Poisson.

### A.12 Método Delta

Es un método para derivar una distribución de probabilidad aproximada para una función con estimadores asintóticamente normales, esto es que al aumentar la muestra los estimadores tienden a ser los de la distribución normal o Gaussiana. La derivación se hace a partir de la varianza limitante de ese estimador. El método delta se puede considerar una generalidad del teorema de límite central. Puede ser usado para obtener intervalos de confianza de los componentes individuales de dichos estimadores, permite obtener intervalos de confianza para , usando la aproximada normalidad de .

Sea  $\hat{\theta}_0$  estimador de máxima verosimilitud para una muestra grande del parámetro d-dimensional, , con matriz de varianza-covarianza  $V_\theta$  . Entonces si  $\phi = g(\theta)$  es una función escalar, el estimador de máxima verosimilitud de  $\phi_0 = g(\theta_0)$  , satisface:

$$\hat{\phi}_0 \sim N(\phi_0, V_\phi)$$

Donde

$$V_\phi = \nabla \phi^T V_\theta \nabla \phi$$

Con

$$\nabla\phi = \left[ \frac{\partial\phi}{\partial\theta_1}, \dots, \frac{\partial\phi}{\partial\theta_d} \right]^T$$

Evaluado en  $\hat{\theta}_0$ .