



COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN EN
CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

POSTGRADO DE SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA
ESTADÍSTICA

**MODELACIÓN DEL RIESGO DE INUNDACIONES EN
EL ESTADO DE TABASCO**

ANA LAURA REYES OLVERA

T E S I S

PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO DE:

MAESTRO EN CIENCIAS

MONTECILLO, TEXCOCO, EDO. DE MÉXICO

2012


La presente tesis titulada: **MODELACIÓN DEL RIESGO DE INUNDACIONES EN EL ESTADO DE TABASCO**, realizada por la alumna: **ANA LAURA REYES OLVERA**, bajo la dirección del consejo particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

MAESTRA EN CIENCIAS

SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA
ESTADÍSTICA

CONSEJO PARTICULAR

CONSEJERO



Dr. Javier Suarez Espinoza

DIRECTOR



Dr. Eduardo Gutiérrez González

ASESOR



Dr. José A. Villaseñor Alva

ASESOR



Dr. M. Michel Rosengaus Moshinsky

Montecillo, Texcoco, Estado de México, Mayo de 2012

Modelación del riesgo de inundaciones en el estado de Tabasco

Ana Laura Reyes Olvera

Colegio de Postgraduados, 2012

Los problemas ocasionados por las inundaciones en México se han incrementado de manera considerable, específicamente en el estado de Tabasco las pérdidas por inundaciones son preocupantes. Por esta razón resulta de gran interés la construcción de un índice que mida el riesgo de inundación en una área determinada.

En el presente trabajo se construye un indicador del riesgo de inundación en el estado de Tabasco, para esto se consideran los siguientes tres componentes: la **vulnerabilidad**, el **costo** y el **peligro**. Con el componente de vulnerabilidad se mide que tan susceptible es el área de interés frente a una inundación y para esto se toma en cuenta tanto elementos sociales como naturales, en el primero se considera la disposición de los servicios con los que cuenta el municipio en estudio y para el segundo se incluyen los recursos hidrológicos que atraviesan o rodean al municipio. En cuanto al costo se responde a cuánto y cómo podría afectar el daño que provoca una inundación, y en este caso se tienen dos factores a considerar: el número de habitantes por comunidad y los bienes con los que cuenta la población. En el caso del componente peligro éste da información sobre la probabilidad de ocurrencia de una inundación, para medir la probabilidad se consideran las precipitaciones máximas anuales de cada municipio y se calculan sus periodos de retorno. Para verificar el modelo probabilístico se propone una prueba de bondad de ajuste “bootstrap” basada en el coeficiente de correlación muestral. El tamaño y potencia de la prueba fueron estimados usando simulación. La potencia de la prueba propuesta fue comparada contra varias distribuciones alternas. Bajo la prueba “bootstrap” se pudo verificar que los datos cumplen con la distribución Gumbel y la distribución de valores extremos generalizada (DVEG), para elegir el mejor modelo se consideraron tres criterios: log-verosimilitud, AIC y la correlación entre la distribución empírica y el ajuste de la distribución. Finalmente se concluye que ambas distribuciones son buenas para modelar este tipo de eventos extremos y que el índice de riesgo de inundación propuesto se puede considerar un indicador apropiado ya que éste describe adecuadamente la realidad.

Palabras clave: Índice de riesgo, Vulnerabilidad, Costo, Peligro, Valores Extremos.

Modelling the risk of flooding in the state of Tabasco

Ana Laura Reyes Olvera

Colegio de Postgraduados, 2012

The problems caused by flooding in Mexico have increased significantly; particularly in the state of Tabasco, flood losses are troubling. For this reason it is of great interest to construct an index that measures the risk of flooding in a given area.

In this paper we construct a flood risk indicator in the state of Tabasco, for this we consider the following three components: the **vulnerability**, the **cost** and **danger**. The vulnerability component measures how susceptible is the area of interest against a flood and this takes into account both social and natural elements, the first considers the provision of services of the municipality under study and the second includes water resources through or around the municipality. As for responding to cost and how it might affect on how much damage it causes a flood. In this case there are two factors to consider: the number of people per community and property there are in the population. In the case of danger, this component provides information on the likelihood of a flood, to measure the probability is it considered maximum annual rainfall of each municipality and their return periods are calculated. For assessing the model probability we propose a goodness of fit test based on bootstrap and the sample correlation coefficient. The size and power of the test were estimated using simulation. The power of the proposed test was compared against several alternative distributions. Under the test bootstrap it is verified that the data comply with the Gumbel distribution and generalized extreme value distribution (GEVD). To choose the best model three criteria are considered: log-likelihood, AIC and the correlation between the empirical distribution and adjustment of the distribution. Finally it is concluded that both distributions are good for modeling this type of extreme events and that the rate proposed flood risk can be considered as a indicator suitable because it describes reality fairly well.

Key words: Risk Index, Vulnerability, Cost, Risk, Extreme Values.

Agradecimientos

Al consejo nacional de ciencia y Tecnología por el apoyo económico brindado para la realización de mis estudios de Maestro en Ciencias.

Al Colegio de Postgraduados por la oportunidad que me brindó.

Al Consejo particular integrado por:

- Dr. Eduardo Gutierrez Gonzalez,
- Dr. Javier Suarez Espinosa,
- Dr. Jose A. Villaseñor Alba,
- Dr. M. Michel Rosengaus Moshinsky,

por el apoyo recibido y el tiempo dedicado a esta tesis.

A mis padres, mi esposo y mis hermanos por la confianza y el apoyo brindado.

A cada uno de los profesores que contribuyeron en mi formación académica.

A mis compañeros y al personal administrativo que de alguna manera ayudaron a la culminación de este proyecto.

A la Línea Prioritaria de Investigación No. 15, Estadística, Modelado y Tecnologías de la información Aplicados a la Agricultura y al Medio Rural del CP, por el apoyo recibido para la realización de mi trabajo de tesis.

Dedicatoria

A mis padres, Cecilia Olvera González y Martín P. Reyes Hernández.

A mi esposo, Magin Zuñiga Estrada.

Índice general

Índice de figuras	IX
Índice de cuadros	IX
Introducción	1
I.1. Revisión bibliográfica	1
I.2. Descripción del problema	3
I.3. Objetivos	5
I.3.1. Objetivo general	5
I.3.2. Objetivos específicos	5
I.4. Justificación	5
I.5. Alcances y limitaciones del estudio	5
I.6. Resultados esperados	6
I.7. Metodología General	6
1. Teoría de valores extremos e índices de riesgo	7
1.1. Introducción	7
1.2. Distribución de Valores Extremos	7
1.3. Construcción de índices	11
1.3.1. Componentes Principales	12
1.3.2. Cálculo de las componentes principales	12
1.4. Riesgo	13
1.5. Índice de Riesgo de Inundación	14

1.5.1. Vulnerabilidad	15
1.5.2. Costo de los Bienes Expuestos	16
1.5.3. Peligro	17
1.6. Mapas de Riesgo de Inundación	18
2. EMV y Pruebas de bondad de ajuste	20
2.1. Introducción	20
2.2. Estimación de los parámetros	20
2.2.1. Método de Máxima Verosimilitud para estimar parámetros	21
2.2.2. Propiedades de los Estimadores de Máxima Verosimilitud	22
2.2.3. Criterio de Información de Akaike (AIC)	23
2.3. Intervalos de Confianza	23
2.4. Pruebas de Bondad de Ajuste	24
2.4.1. Prueba Bootstrap basada en el coeficiente de Correlación muestral	25
2.5. Niveles y Periodos de Retorno	33
2.5.1. Relación con la distribución de probabilidad acumulada	34
2.5.2. Distribución geométrica	34
3. Modelación del riesgo de inundaciones en Tabasco	35
3.1. Introducción	35
3.2. Resultados del Factor de Vulnerabilidad	35
3.3. Factor de Costo	37
3.4. Factor de Peligro	39
3.4.1. Obtención de Datos de Precipitación	39
3.4.2. Análisis de los Datos	45
3.4.3. Resultados del Factor Peligro	48
3.5. Resultados del Índice de Riesgo	65
Conclusiones	67
Bibliografía	67
Anexos	71
A.1. Anexo A: Programas para la estimación de CP en SAS	71

A.2. Anexo B: Información de disposición de servicios en los municipios de Tabasco	71
A.3. Anexo C: Información de bienes con los que cuenta la población por municipio	73
A.4. Anexo D: Breve explicación de la matriz de covarianzas, valores y vectores propios	74
A.5. Anexo E: Precipitación Máxima Anual en mm, por municipio	74
A.6. Anexo F: Subcuencas para cada municipio	78
A.7. Anexo G: Programas para la estimación del tamaño y potencia de la prueba .	81
A.8. Anexo H: Programas para la estimación de parámetros y gráficas	82

Índice de figuras

I.1. Inundaciones en Tabasco	4
1.1. Máximos de la Función de Densidad	8
3.1. Índice de Vulnerabilidad	37
3.2. Nodos en la República Mexicana	40
3.3. Región Hidrográfica 29 y 30	43
3.4. Región de Interés, Tabasco	45
3.5. Gráficas del municipio de Balancán	49
3.6. Gráficas del municipio de Cárdenas	50
3.7. Gráficas del municipio de Centla	51
3.8. Gráficas del municipio de Centro	51
3.9. Gráficas del municipio de Comalcalco	52
3.10. Gráficas del municipio de Cunduacan	53
3.11. Gráficas del municipio de Emiliano Zapata	53
3.12. Gráficas del municipio de Huimanguillo	54
3.13. Gráficas del municipio de Jalapa	55
3.14. Gráficas del municipio de Jalpa de Méndez	55
3.15. Gráficas del municipio de Jonuta	56
3.16. Gráficas del municipio de Macuspana	57
3.17. Gráficas del municipio de Nacajuca	57
3.18. Gráficas del municipio de Paraíso	58
3.19. Gráficas del municipio de Tacotalpa	59
3.20. Gráficas del municipio de Teapa	59

3.21. Gráficas del municipio de Tenosique	60
3.22. Mapa de peligro a un nivel de retorno de 100mm.	64
3.23. Mapa de peligro a un nivel de retorno de 150mm.	65
3.24. Mapa de peligro a un nivel de retorno de 200mm.	65
3.25. Mapa de peligro a un nivel de retorno de 250mm.	65
3.26. Índice de Riesgo de Inundaciones	66
A.27. Nodos de interés en el municipio de Balancán.	78

Introducción

I.1. Revisión bibliográfica

El agua es uno de los recursos naturales más valioso de cualquier país debido a los beneficios sociales y económicos que se derivan de su consciente explotación; sin embargo, junto con las ventajas existen también situaciones extremas tales como las inundaciones y las sequías.

A nivel mundial las inundaciones están aumentando más rápidamente que ningún otro desastre. Las inundaciones han causado severos problemas, tanto sociales como económicos, por esta razón se han desarrollado diversas investigaciones para analizar eventos extremos y la forma de evaluar el riesgo de inundación.

El estudio de las inundaciones es complejo ya que por ejemplo, el daño económico directo a un edificio inundado no sólo depende de la profundidad del agua y uso del edificio. Otros factores importantes son las características del edificio, las variables socioeconómicas, la calidad de la respuesta de emergencia, entre otras (Smith, 1994) [33].

Existen diferentes formas de estudiar las inundaciones, Hunt (2002) [25] analiza las inundaciones como un desastre natural en el que se debe identificar los peligros y la vulnerabilidad para poder superarlos. El CENAPRED (2006)[12] (Centro Nacional de Prevención al Desastre) por el contrario no maneja a los desastres como naturales sino como producto de las condiciones de vulnerabilidad y exposición, entonces si lo que se desea es prevenir los desastres, es necesario hablar de riesgo. Existen diferentes definiciones de riesgo, éstas se basan principalmente en los factores de los que depende su estimación. Además es importante mencionar que los efectos de las inundaciones en la sociedad no pueden ser representados en una cotización monetaria (Green *et al.* 1994) [19].

En algunos otros trabajos como Young (2002) [39] se combinan los métodos estadísticos y determinísticos para analizar datos pasados y predecir los niveles de inundación y de ríos; otro más es Bayliss & Reed (2001) [2] en el que se discuten diferentes formas de estudiar las inundaciones históricas y revisa métodos para incorporar datos históricos al análisis de frecuencias de inundaciones. Se puede incorporar información meteorológica para fortalecer la

estimación del riesgo, la lluvia puede ser un notorio indicador fiable de la inundación (Duncan 2002) [13]. Sin embargo el riesgo abarca dos aspectos: Peligro y Vulnerabilidad (Mileti, 1999) [28] y aunque si bien no es la única definición que se maneja para el riesgo es una de las más utilizadas. Ordaz (1996) [29], incluye un tercer factor para la expresión del cálculo de riesgo por algún fenómeno natural, éste es el costo, es decir, el valor de los bienes expuestos, quedando finalmente como el producto de tres factores $R = C \times V \times P$. Nótese que en realidad se trata de la misma definición ya que la vulnerabilidad está compuesta de dos elementos la exposición y la susceptibilidad (daño potencial y pérdida) (Begum *et al.*, 2007) [3]. Por lo tanto para obtener una medida del riesgo se debe trabajar sobre los tres factores.

Debido al impacto de las inundaciones sobre la sociedad y las pérdidas monetarias es importante la construcción de un índice de vulnerabilidad Tapsell *et al.* (2002) [35]. La vulnerabilidad tiene diferentes conceptos, por ejemplo Blaikie *et al.* (1994) [4] analizan las condiciones socio-económicas que aumentan el grado de vulnerabilidad.

En cuanto al peligro se deben obtener las probabilidades de que ocurra un fenómeno que pueda dañar lo expuesto (Ordaz, 1996) [29], en el caso de las inundaciones puede ser generado por:

- Lluvias (de invierno, de verano o convectivas)
- Ciclones tropicales
- Falla de obras hidráulicas (diseño inadecuado, mala operación o falta de mantenimiento)

En caso de ser por lluvias, las inundaciones se presentan cuando se registran valores extremos de precipitación, es aquí donde se introduce la teoría sobre valores extremos.

En 1928 Fisher y Tippett [14] derivaron la forma límite de la distribución del valor máximo de una muestra aleatoria, encontrando las tres posibles distribuciones: Gumbel, Weibull y Fréchet. A partir de esto se desarrollan aportaciones de interés tales como las de Gumbel (1934) [20], que además es quien presenta el primer libro de importancia (Gumbel, 1958) [22], en el que trabajo con eventos extremos contemplando aplicaciones. En 1943 Gnedenko [17] presenta en forma general y con una prueba rigurosa el Teorema de los tipos de distribución de extremos propuestos por Fisher y Tippett. Otra aportación importante la dio Von Mises (1936) [38] & Jenkinson (1955) [26] que trabajaron sobre la distribución de valores extremos y propusieron una distribución que combina las tres familias de distribución de valores extremos, la cual es conocida como DVEG. De 1990-2011 el desarrollo estadístico ha sido sustancial en el área, por mencionar algunos, Tawn (1992) [37], Rosbjerg & Madsen (1996) [31], Coles & Dixon (1999) [8] y Coles (2001) [7].

Una vez que se ha obtenido una medida del riesgo se recomienda el uso de mapas de peligro y riesgo por inundaciones para ubicar las zonas en diferentes categorías de riesgo y desarrollar un plan de manejo (Tawatchi & Mohammed, 2005) [36], Lehner & Döll (2001)

[27] dan un ejemplo para aproximar a escala global el análisis con mapas concernientes a la situación de inundación en Europa, en cuanto a México se tiene el desarrollo de los mapas de Riesgo a escala municipal por inundaciones (CENAPRED, 2006) [12]

I.2. Descripción del problema

El agua es uno de los recursos naturales más valioso de cualquier país debido a los beneficios sociales y económicos que se derivan de su consciente explotación; sin embargo, junto con las ventajas existen desventajas en situaciones extremas tales como las **inundaciones** y las sequías.

Inundación: es aquel evento que debido a la precipitación, oleaje, marea de tormenta, o falla de alguna estructura hidráulica provoca un incremento en el nivel de la superficie libre del agua de los ríos o el mar mismo, generando invasión o penetración de agua en sitios donde usualmente no la hay y, generalmente, daños en la población, agricultura, ganadería e infraestructura.

El tipo de inundación puede definirse desde diversos puntos de vista de acuerdo con:

- Duración de la inundación, éstas pueden ser lentas o súbitas
- Mecanismo que las genere, pueden ser pluviales, fluviales, costeras o por falla de obras hidráulicas.

En el pasado, las poblaciones entendían la naturaleza de las inundaciones, no en términos estadísticos, sino como un elemento del medio ambiente con el que se mantenían en estrecho contacto. Aunque la gente vivía cerca de los ríos para aprovecharlos como medio de transporte y fuente de abastecimiento de agua para su consumo y el cultivo de sus fértiles llanuras de inundación, sus hogares eran construidos en terrenos altos, para evitar afectaciones debidas a las inundaciones. Pero en la actualidad, por el aumento de la población es difícil mantener ese respeto hacia la naturaleza y los problemas de inundación han aumentado.

A nivel mundial las inundaciones están aumentando más rápidamente que ningún otro desastre. De acuerdo con la Cruz Roja Internacional, durante el periodo de 1919-2004, han colaborado con ayuda en más eventos de inundaciones que de cualquier otro tipo.

Una estimación de las víctimas fatales en México de 1980 a 1999 a consecuencia de fenómenos hidrometeorológicos arroja 2767 personas, lo que representa un promedio cercano a los 140 individuos fallecidos anualmente. La cantidad de daños totales por este tipo de fenómenos en el mismo periodo fue de 4,537 millones de dólares, lo que en promedio arroja 227 millones de dólares en pérdidas anuales.

En las planicies de los grandes ríos de México, prácticamente todos los años se producen inundaciones derivadas de sus desbordamientos. La causa principal es la pérdida de la capacidad hidráulica de esas corrientes, una vez que dejan la zona de sierras y se adentran en las planicies.

En el caso de México el impacto de daños y pérdidas asociadas a estas lluvias es especialmente severo en Tabasco. Se conoce que Tabasco es la entidad de la República que presenta un mayor escurrimiento de agua a lo largo del año; la red hidrológica de esta región es la más compleja del país, caracterizándose por entramados sinuosos de corrientes superficiales y una gran densidad de cuerpos de agua. Grandes extensiones del terreno tabasqueño son propensas a sufrir inundaciones durante los meses de lluvia crecida, presentándose numerosos cuerpos de agua intermitentes.



Figura I.1: Inundaciones en Tabasco

La ocurrencia de lluvias máximas impacta en la generación de desastres geológicos y riesgos hidrometeorológicos en zonas vulnerables del sureste del país. En el estado Tabasco, por ejemplo, las inundaciones presentan frecuentemente el carácter de catástrofe. Es por esta razón que surge la necesidad de hacer una estimación del **riesgo de inundación**, para evaluar las zonas más propensas a este tipo de problemas en el Estado de Tabasco.

I.3. Objetivos

El trabajo tiene diferentes metas a desarrollar.

I.3.1. Objetivo general

Modelar el riesgo de inundaciones en el estado de Tabasco por causas climáticas extremas.

I.3.2. Objetivos específicos

- Proponer un modelo estadístico basado en la teoría de valores extremos que permita determinar el comportamiento y tendencias de las precipitaciones extremas.
- Calibrar el modelo estadístico a desarrollar usando una base de datos histórica de precipitación y niveles hidrométricos del estado de Tabasco.

- Proponer una metodología para determinar un indicador del riesgo de inundaciones que permita alertar a la ciudadanía del lugar en cuestión.
- Construir mapas de riesgo bajo diferentes escenarios de inundaciones por municipio en el estado de Tabasco.

I.4. Justificación

Las inundaciones en Tabasco, México, son un problema que aqueja a gran parte de la población de este estado y no ha sido estudiado con el suficiente rigor estadístico, por esta razón este trabajo está encaminado en proponer un índice de riesgo de inundaciones por municipio del estado de Tabasco. El indicador comprende tres factores: Vulnerabilidad, Costo y Peligro, para los dos primeros se desarrolla un indicador con la información municipal, pertinente para cada factor, tomada de la base de información de INEGI y en cuanto al peligro se analiza con la teoría de valores extremos las precipitaciones históricas de cada municipio, información dada por el SMN.

I.5. Alcances y limitaciones del estudio

La información que se tiene disponible para el análisis del componente vulnerabilidad y del componente costo no es suficiente para medirlos ya que se dejan fuera algunos otros factores que pueden aumentar o disminuir el riesgo de inundación. Por ejemplo, no se cuenta con información completa para analizar las pérdidas o la vulnerabilidad en el ámbito de los sectores económicos.

Se propone una modelación de las precipitaciones extremas para cada municipio que al ser calibrada con series históricas resulta ser una buena estimación.

En algunos municipios sus indicadores de vulnerabilidad y de costo obedecen a cómo se encuentra el municipio en general pero existen algunas regiones que podrían sufrir un daño mayor al estimado.

Se construye un índice de riesgo de inundaciones a escala municipal que se puede considerar un buen indicador ya que se encuentra muy cercano de la realidad.

I.6. Resultados esperados

Con el trabajo se espera dar respuesta a los objetivos planteados. Es decir, proponer un índice del riesgo de inundaciones por municipio en el estado de Tabasco, México y un modelo estadístico basado en la teoría de valores extremos que permita determinar el comportamiento y tendencias de las precipitaciones extremas.

Además de construir un mapa de peligro de inundación con diferentes escenarios y un mapa del riesgo de inundación por municipio. Se desea que la metodología propuesta pueda ser aplicada a cualquier otro estado de la República Mexicana.

I.7. Metodología General

La solución al problema será desarrollado de la siguiente forma.

- Revisar los diferentes conceptos de Riesgo y elegir el más adecuado para el problema.
- Estudiar a cada componente que influye en el riesgo (vulnerabilidad, costo y peligro) y revisar con qué información se cuenta para su análisis.
- Construir un indicador de vulnerabilidad y un indicador de costo con la información que se tenga disponible para cada caso.
- Revisar la teoría de valores extremos y de periodos de retorno que será utilizada para el componente peligro.
- Construir una prueba de bondad de ajuste para aplicarla a los datos de precipitación.
- Elegir la región de estudio, es decir seleccionar la información de precipitación que será analizada.
- Obtener las precipitaciones máximas anuales y ajustarlas a un modelo dado por las distribuciones de valores extremos (Gumbel y VEG).
- Elegir el mejor modelo y con este obtener los periodos de retorno para los 17 municipios.
- Construir el índice de riesgo de inundación con el producto de los tres componentes obtenidos.
- Construir el mapa del índice de riesgo de inundación a escala municipal.

Teoría de valores extremos e índices de riesgo

1.1. Introducción

En el presente capítulo se estudiarán las distribuciones para los eventos extremos y su relación con algunos resultados clásicos de la Teoría de la Distribución de Valores Extremos (DVE). Se revisará el Teorema de Fisher y Tippett con el cual se sustenta el uso de las DVE en los datos en que se tienen valores extremos y su generalización con el Teorema demostrado por Gnedenko.

El estudio inicia con un bosquejo sobre la historia de la teoría de valores extremos. Posteriormente se dará una pequeña revisión a la teoría de componentes principales que será utilizada para la construcción de algunos indicadores necesarios para la solución del problema.

Finalmente serán revisados los conceptos del riesgo y de cada uno de los componentes involucrados, así como la información que será procesada en el desarrollo de la investigación.

1.2. Distribución de Valores Extremos

Sea X_1, X_2, \dots, X_n una sucesión de variables aleatorias independientes e idénticamente distribuidas con función de distribución F y observadas a través del tiempo. Si M es el máximo de las variables aleatorias, esto es $M_n = \max \{X_1, \dots, X_n\}$, entonces

$$\begin{aligned}
 P(M_n \leq z) &= P(X_1 \leq z, \dots, X_n \leq z) \\
 &= P(X_1 \leq z) \cdot \dots \cdot P(X_n \leq z) \\
 &= \prod_{i=1}^n P(X_i \leq z) \\
 &= \{F(z)\}^n
 \end{aligned}
 \tag{1.2.1}$$

La función de densidad de probabilidad de M_n está dada por

$$nF'(x)F^{n-1}(x) \tag{1.2.2}$$

NOTA 1.1. La distribución F es desconocida, la cual si bien es posible estimar por medio de los datos observados no es recomendable ya que se pueden tener variaciones pequeñas de la estimación que implican diferencias para $\{F(z)\}^n$.

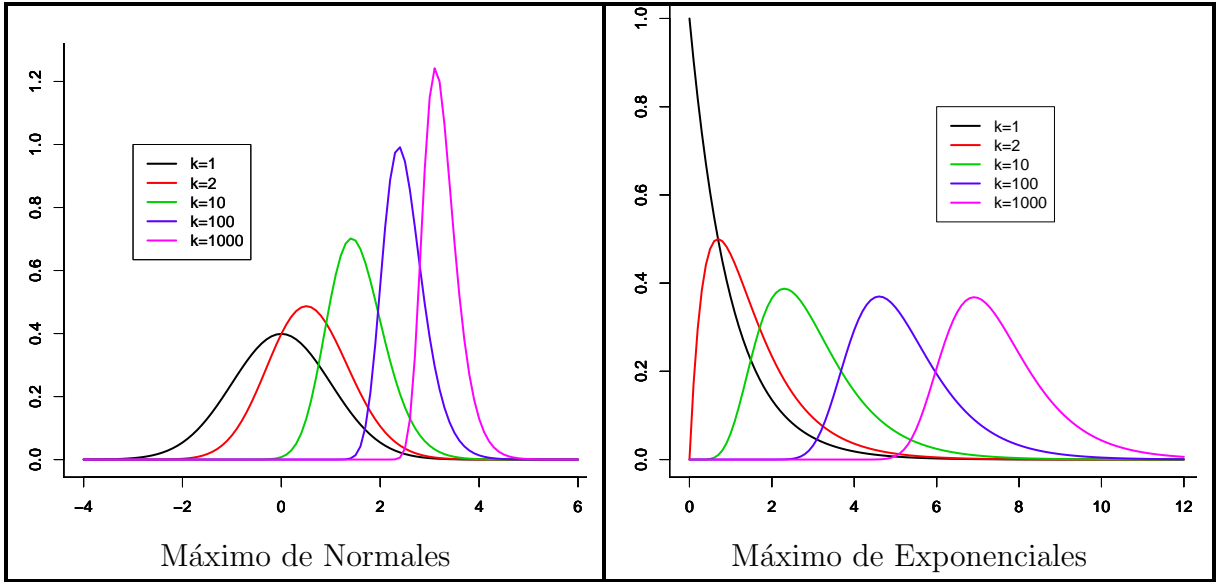


Figura 1.1: Máximos de la Función de Densidad

Al observar la figura 1.1 se puede decir que la distribución límite de $\{F(z)\}^n$ depende de la varianza $\sigma_n(x)$ y la media μ_n de los datos, ya que muestra el comportamiento de dos funciones de densidad (Normal y Exponencial) al aumentar el tamaño de la potencia n . Por lo que es posible encontrar una distribución límite para $\{F(z)\}^n$. Una manera de encontrar dicha distribución es por medio de aproximaciones de la distribución de M_n . Por ejemplo para aproximar la distribución de la suma de variables aleatorias se aplica el Teorema del Límite Central (TLC). Sean Y_1, Y_2, \dots, Y_n una sucesión de variables aleatorias *iid* con función de distribución H . En la práctica no siempre se conoce a H , sin embargo, suponiendo que Y_i tiene media y varianza finita, entonces por el TLC se cumple que

$$\frac{\sum_{i=1}^n Y_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0, 1) \tag{1.2.3}$$

Si se tienen estimaciones para μ y σ , entonces la convergencia dada en (1.2.3) se puede

usar para aproximar la distribución de la suma con n grande, esto es,

$$\frac{\sum_{i=1}^n Y_i - n\mu}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{} N(0, 1) \quad (1.2.4)$$

De manera similar a la distribución de la suma, se encuentran las distribuciones límites para aproximar $\{F(z)\}^n$. Denotando al extremo derecho de F por

$$x_F = \sup \{x \in \mathbb{R} : F(x) < 1\} \quad (1.2.5)$$

entonces, se cumple que $M_n \rightarrow x_F$, es decir, la distribución de máximos converge a una distribución degenerada. La afirmación anterior resulta puesto que para $x < x_F$, se tiene $F(x) < 1$ y

$$P(M_n \leq x) = F^n(x) \rightarrow 0, \text{ cuando } n \rightarrow \infty, \quad (1.2.6)$$

por lo que M_n converge en probabilidad a x_F , y dado que la sucesión $\{M_n\}$ es no decreciente, entonces converge casi seguramente, y en consecuencia, también en distribución.

El argumento anterior muestra que es necesario estandarizar a M_n , es decir, se busca garantizar una convergencia débil de la forma

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n) \rightarrow H(x), \quad (1.2.7)$$

con $\{a_n\}$ y $\{b_n\}$ sucesiones de números reales y H una distribución límite no degenerada.

La distribución de valores extremos generalizada, es la distribución que se ajusta a los máximos de bloques independientes de observaciones adecuadamente normalizadas y surge a partir del teorema propuesto por Fisher y Tippett en 1928 [14], demostrado rigurosamente por Gnedenko en 1943, que va en la dirección mencionada.

Teorema 1.1 (Fisher-Tippett, Gnedenko). *Sea $\{X_n\}$ una sucesión de variables aleatorias iid con una función de distribución común $F(x)$. Si existen sucesiones de constantes de normalización $\{a_n > 0\}$ y $\{b_n\} \in \mathbb{R}$ y una función de distribución H no degenerada, tal que $\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq z\right) = H(z)$, entonces $H(z)$ pertenece a alguna de las siguientes familias:*

1).- *Gumbel (colas medias):*

$$\Delta(x) = \exp \left\{ - \exp \left[- \left(\frac{x - b}{a} \right) \right] \right\}, \quad x \in \mathbb{R}$$

2).- *Fréchet (colas pesadas):*

$$\Phi(x) = \begin{cases} \exp \left[- \left(\frac{x-b}{b} \right)^{-\alpha} \right] & \text{si } x > b \\ 0 & \text{si } x \leq b \end{cases}$$

3).- Weibull (colas cortas o ligeras):

$$\Psi(x) = \begin{cases} 1 - \exp \left\{ - \left(\frac{x-b}{a} \right)^\alpha \right\} & \text{si } x > b \\ 0 & \text{si } x \leq b \end{cases}$$

donde $b \in \mathbb{R}$, $a > 0$ y $\alpha > 0$. El parámetro a es de escala, el parámetro b de localización y en los casos Fréchet y Gumbel determina el soporte de la distribución. Finalmente, α es un parámetro de forma.

Las tres familias pueden ser combinadas en un solo modelo, la Distribución de Valores Extremos Generalizada (DVEG) dado por

$$F(x) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]_+^{-\frac{1}{\xi}} \right\}, \quad (1.2.8)$$

tal que $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \xi < \infty$, donde μ es el parámetro de localización, σ el parámetro de escala y ξ el parámetro de forma. Además, para $\xi \neq 0$ el soporte está determinado por una combinación de dichos parámetros, en particular, para $\xi > 0$ la distribución tiene un extremo izquierdo finito, mientras que para $\xi < 0$ el extremo derecho es finito.

Cuando $\xi > 0$ resulta la función de distribución Fréchet, mientras que si $\xi < 0$ se tiene la función de distribución Weibull, finalmente para $\xi = 0$, se obtiene un subconjunto de la DVEG, que es interpretado como el límite de $H(z)$ cuando $\xi \rightarrow 0$, dicho subconjunto representa la familia Gumbel.

Además es posible determinar los parámetros de (a, b, α) que le corresponden a G cuando se escribe como una DVE.

$$\text{Cuando } \xi > 0 : \begin{cases} b = \mu - \frac{\sigma}{\xi} \\ a = \frac{\sigma}{\xi} \\ \alpha = \frac{1}{\xi} \end{cases} \quad \text{caso contrario } \xi < 0 : \begin{cases} b = \mu - \frac{\sigma}{\xi} \\ a = -\frac{\sigma}{\xi} \\ \alpha = -\frac{1}{\xi} \end{cases}$$

Finalmente, para $\xi = 0$ se tiene $b = \mu$ y $a = \sigma$.

Además de poder identificar con el signo de ξ el tipo de DVE a la que corresponde una DVEG, también es posible determinar el elemento estándar con el que está relacionada.

$$\begin{cases} \xi = \alpha^{-1} > 0, & \text{se tiene una distribución tipo Fréchet} \\ \xi = 0, & \text{se tiene una distribución tipo Gumbel} \\ \xi = -\alpha^{-1} < 0, & \text{se tiene una distribución tipo Weibull} \end{cases} \quad (1.2.9)$$

El resultado anterior se enuncia en el siguiente teorema.

Teorema 1.2 (Distribución de Valores Extremos Generalizada). *Si existen sucesiones de con-*

stantes $\{a_n > 0\}$ y $\{b_n\}$ tales que

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = H(x)$$

donde $H(x)$ es una función no degenerada, entonces H es miembro de la familia de valores extremos generalizada dada por:

$$F(x) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]_+^{-\frac{1}{\xi}} \right\}$$

definido sobre $\{x : 1 + \xi \left(\frac{x - \mu}{\sigma} \right) > 0\}$, donde $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \xi < \infty$.

La distribución tipo Gumbel tiene una cola superior no acotada que disminuye rápidamente, algunas distribuciones que tienen este comportamiento en su soporte son: normal, exponencial, gamma, logística y log-normal. Aunque la distribución tipo Fréchet también tiene una cola superior no acotada, ésta disminuye lentamente y sus momentos son infinitos para todos los órdenes mayores $1/\xi$ (por ejemplo, la varianza es infinita si $\xi > 0.5$; la media es infinita si $\xi > 1$), algunas distribuciones que tienen este comportamiento dentro de su soporte son: Pareto, Burr, log-gamma, Cauchy y t-student. En cuanto a la distribución tipo Weibull tiene un límite superior finito en $x = \mu - (\sigma/\xi)$, algunas distribuciones que pertenecen a este tipo son: Uniforme y Beta.

Si una variable aleatoria X tiene *DVEG*, entonces la variable estandarizada $(X - \mu)/\sigma$ tiene una distribución que no depende de μ ni de σ , sino únicamente de ξ . Se tiene que el parámetro de localización especifica en donde se encuentra centrada la distribución, el parámetro de escala especifica su propagación, y el parámetro ξ está asociado al tipo de la cola de la distribución, en cuanto más grande sea el valor de ξ más pesada es la cola de la distribución.

1.3. Construcción de índices

Uno de los principales problemas que se tienen en la búsqueda de buenos índices consiste en determinar ponderaciones apropiadas para los componentes de los factores que intervienen en el índice. Por otro lado, del análisis multivariado se sabe que si se tienen p variables, x_1, x_2, \dots, x_p y de cada una n observaciones, al acomodar la información en una matriz $n \times p$, se tiene

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2p} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \cdots & x_{np} \end{bmatrix}.$$

Para la comparación se conocen dos formas: por filas o por columnas. La comparación que se requiere para los pesos es entre variables, o lo que es lo mismo entre columnas. Más aún se requieren no sólo su comparación sino los pesos que se deben asignar a cada una de ellas. Es decir, se requieren combinaciones lineales entre las variables para ver cuál es la más adecuada.

Como es de imaginarse existen infinidad de combinaciones lineales que se pueden proponer entre las variables. Por otro lado, del análisis multivariado se conoce una técnica llamada componentes principales que en su parte inicial de tipo descriptivo da una respuesta adecuada.

1.3.1. Componentes Principales

Sean las p variables x_1, x_2, \dots, x_p de interés, se llaman componentes principales a las nuevas variables, resultantes de las combinaciones lineales de ellas

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p \\ y_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p \\ \vdots \\ y_p = a_{p1}x_1 + a_{p2}x_2 + \dots + a_{pp}x_p \end{cases}$$

Que cumplen la característica de no ser correlacionadas.

De esta forma las componentes principales permiten ver con mayor claridad las diferencias entre cada uno de los n valores de las variables, jerarquizándolos por el criterio que usa en la construcción de y_i sobre las variables x_i .

Por otro lado, se desea que las componentes principales cumplan

- $cov(y_i, y_j) = 0$ para $i \neq j$, no correlacionadas.
- $var(y_1) \geq var(y_2) \geq \dots \geq var(y_p)$, la varianza de la primera componente sea mayor.

1.3.2. Cálculo de las componentes principales

El método que será revisado fue introducido originalmente por Karl Pearson en 1901 y desarrollado por Hotelling en 1933.

1).- Si las variables x_1, x_2, \dots, x_p son conmensurables.

- a).- Con la matriz \mathbf{X} calcule su matriz de covarianzas, Σ .
- b).- Calcular los valores y vectores propios de la matriz de covarianzas. Para que cumplan la segunda propiedad deseada sobre la varianza, se deben buscar los vectores propios, iniciando con el valor propio más grande. Es decir, los valores propios, se indexan para que queden $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, con λ_i valores propios de la matriz de covarianzas. De esta forma se cumplirá $var(y_1) \geq var(y_2) \geq \dots \geq var(y_p)$.
- c).- De esta forma las p componentes principales son los p vectores propios.

- d).- Las ponderaciones de las variables estarán dadas por los elementos de las componentes principales, por medio de la siguiente regla. Sea la variable k , ($k = 1, 2, \dots, p$) la que el decisor considere que tenga mayor peso, si el k -ésimo elemento de la primer componente principal es el mayor, entonces las ponderaciones de las p variables son los p elementos del componente principal, respectivamente. En caso contrario, tal vez sea necesario realizar sumas entre los elementos de las componentes principales.
- 2).- Si las variables x_1, x_2, \dots, x_p no son conmensurables.
- a).- Con la matriz \mathbf{X} calcule su matriz de correlaciones, Σ_ρ .
- b).- Los demás incisos se repiten, pero con la matriz de correlaciones.

1.4. Riesgo

En muchas ocasiones se cree que los peligros son los únicos responsables de los desastres o emergencias. Por ejemplo, se suele señalar al huracán o a las lluvias extremas como responsables de las pérdidas que existen durante un desastre. En realidad es en buena medida la sociedad la que se expone con su infraestructura, organización y cultura que tiene que enfrentarse a este tipo de fenómenos que conllevan un alto grado de peligro. Por lo tanto se debe entender que los desastres no son naturales, es decir, son producto de condiciones de vulnerabilidad y exposición (CENAPRED, 2006) [12].

Si lo que se desea es prevenir los desastres, es necesario hablar de riesgo. Existen diferentes definiciones de riesgo, sin embargo es necesario definir y dar los factores de los que depende el riesgo. El riesgo se puede definir como la probabilidad de sufrir algún daño o pérdida. Entonces se puede decir que el análisis del riesgo es un conjunto de conocimientos que se evalúan, y con ellos se puede calcular la probabilidad de efectos adversos de un proceso natural, tecnológico, industrial o algún otro (Begum *et al.*, 2007) [3].

La definición de riesgo, (R), que será utilizada es aquella mencionada en Ordaz (1996) [29] y una de las más aceptadas. Es común, encontrar la siguiente definición:

$$R = C \times V \times P, \quad (1.4.1)$$

es decir, se define al riesgo como el producto de tres factores, los cuales son: el valor de los bienes expuestos, C , la vulnerabilidad, V , y el peligro, P , entendiendo a éste como la probabilidad de ocurrencia de un fenómeno que pueda dañar lo expuesto.

NOTA 1.2. La ecuación (1.4.1) que define al riesgo se refiere a la ocurrencia de un solo evento, sin embargo el principal interés es medir el riesgo asociado a los diferentes eventos que pueden ocurrir a lo largo de un cierto lapso. Por esta razón se requiere desarrollar y precisar el significado de cada uno de los factores que influyen en la obtención del riesgo.

Además, se debe tomar en cuenta que encontrar una medida que represente el riesgo y sus cambios a través del tiempo, puede ser estadísticamente difícil de modelar. Luego, encontrar la manera de modelar el riesgo resulta un problema complejo.

Una forma de tratar este problema estadísticamente puede ser introduciendo los números índice, que no son más que una especie de barómetros de cambios tanto sociales, económicos, entre otros. Los índices desde su aparición han llegado a ser cada vez más importantes como indicadores de los diferentes cambios que rodean a la sociedad. Es entonces por lo que se utiliza un índice para medir el riesgo.

1.5. Índice de Riesgo de Inundación

Para la construcción del índice de riesgo de inundación se debe tener en cuenta los distintos aspectos que pueden caracterizar al riesgo de inundación y que incidirán en el valor final del indicador. Tomando en cuenta que si bien el conocimiento de la probabilidad de peligro de inundación es un elemento importante en la prevención, no se debe olvidar que el análisis del riesgo de inundaciones envuelve las características de los elementos del riesgo, es decir todos los elementos de la sociedad y medio ambiente. Tales elementos son la población, construcciones, actividades económicas, ecosistemas, etcétera. Puesto que el daño de una inundación no depende únicamente de ésta, sino también de la vulnerabilidad del área inundada y de las pérdidas que se pudiesen presentar.

Por estas razones en el desarrollo de esta investigación el indicador del riesgo de inundación se obtiene a partir de tres factores relevantes: Costo de los bienes expuestos, vulnerabilidad frente a un fenómeno que pudiera ocasionar inundaciones y finalmente el peligro ante este fenómeno.

Se debe obtener un indicador para cada componente, éstos se analizan por separado y de forma diferente según la información que influya en el valor de cada uno y sobre todo con la que se tenga disponible. Para el cálculo del componente peligro se utilizan los registros de lluvia diaria y se analizan por medio de la teoría desarrollada para valores extremos, mientras que para la vulnerabilidad se hace uso de la información concerniente a la infraestructura con la que cuenta la entidad geográfica así como sus principales recursos hidrológicos (ríos y lagunas) y finalmente para el indicador del costo se toma en cuenta el tamaño de la población, basándose en el supuesto “Que a mayor población, mayor riqueza”, y finalmente se pondera con un índice de los bienes con los que cuenta la familia (enferos domésticos, automóvil, computadora, etc.).

En el manejo de los datos de infraestructura y bienes se construyen indicadores, los cuales tendrán que cumplir con dos criterios; por un lado obtener en la combinación lineal el máximo de varianza, ya que esto garantiza que discrimine del mejor modo entre las entidades geográficas a las que se aplique, y por otro, obtener coeficientes propios, ya que esto garantiza explicar adecuadamente la situación de riesgo en la que se encuentra. La manera fue encontrar las componentes principales (CP) de cada conjunto de datos, se sumaban las componentes

principales y se examinaban si los coeficientes y la proporción de la varianza de la suma resultante eran pertinentes, esto se hizo hasta que se habían sumado todas las CP del conjunto. Teniendo todas estas combinaciones lineales se eligió la que tenía coeficientes acordes con lo esperado por la teoría, el máximo de varianza.

En las siguientes secciones se explica en qué consiste cada uno de los factores que influyen en el índice de riesgo, se detallan las variables que intervienen en el valor final del factor así como la forma de obtenerlo.

1.5.1. Vulnerabilidad

La vulnerabilidad puede tener diferentes conceptos, por ejemplo Blaikie (1994) [4] analiza las complejas condiciones socio-económicas que crean un alto grado de vulnerabilidad. Begum *et al.* (2007) [3] mencionan que el acceso a los recursos es usualmente el factor más crítico con el que se logra asegurar el sustento o la recuperación efectiva de un desastre.

En esta investigación aplicamos la siguiente definición: La vulnerabilidad, V , es una medida de qué tan susceptible es un bien expuesto a la ocurrencia de un fenómeno perturbador (Ordaz, 1996) [29]. En el análisis de la vulnerabilidad se está dando respuesta a ¿Quién o qué puede ser afectado por alguna inundación?.

En el análisis de la vulnerabilidad entran varios factores, como el uso que tiene la construcción (vivienda, infraestructura pública, manufacturera, etc.), las características de la construcción (material, tipo de construcción, el lugar en el que se construyó, etc.), variables naturales (ríos, lagunas, obras hidráulicas, etc.) y variables socio-económicas (servicios que hay en el área de estudio, calidad de respuesta ante una inundación, estatus financiero, etc.).

El análisis del índice de vulnerabilidad de esta investigación no contempla todas las variables descritas arriba ya que no se cuenta con toda la información, por lo que se trabaja con la información que se tenga de manera concreta. Por lo tanto, se utilizan dos variables importantes: los servicios con los que cuenta la comunidad (energía eléctrica, agua potable y drenaje) y sus principales recursos hidrológicos (ríos).

En primer lugar se construye un índice de servicios con los tres factores que se describen en la tabla 1.1:

Variables	Descripción
PO_NDEE	Porcentaje de ocupantes que no disponen de energía eléctrica
PO_NDAE	Porcentaje de ocupantes que no disponen de agua entubada
PO_NDD	Porcentaje de ocupantes que no disponen de drenaje

Tabla 1.1: descripción de las variables incluidas en el índice de servicios

Mediante el uso de componentes principales se construye el índice de servicios ya que con esto se garantiza obtener una combinación lineal con máxima varianza para discriminar al mejor y obtener coeficientes propios para explicar adecuadamente la situación de la comunidad.

Ahora, pasando a la segunda variable, los recursos hidrológicos con los que cuenta cada municipio, ésta es considerada ya que la región de Tabasco se encuentra a muy poca altura sobre el nivel de mar, es una zona de ríos, lagunas y pantanos. Además a lo largo de su historia ha sufrido inundaciones por el desbordamiento de los ríos que atraviesan la entidad. Por ejemplo en la inundación del 2007 en el municipio de Paraíso se debió al exceso de lluvia que coincidió con oleaje en el Golfo de México lo cual evitó las descargas de los ríos Grijalva y Carrizal hacia el mar. Además es importante mencionar que la población se ha establecido cada vez más en zonas de alto riesgo a los márgenes de los ríos.

En la tabla 1.2 se muestran los principales ríos que riegan a cada municipio y como información adicional se tiene el régimen de precipitaciones que caracteriza a cada municipio del estado de Tabasco con un total de caída de agua en mm.

Con la información del Nivel de Agua Máximo Ordinario (NAMO) de los ríos mencionados en la tabla 1.2 se realiza una ponderación para cada municipio y se obtiene un indicador de los ríos.

Municipio	Ríos	Precip. Anual
Balancan	Usumacinta y San Pedro Mártir	1,500 mm
Cárdenas	San Felipe, Naranjeño, Santana, y Chicozapote.	2,643 mm
Centla	Grijalva, Usumacinta, San Pedro y San Pablo	1,696 mm
Centro	Grijalva, Samaria, Carrizal y río Viejo	2,237 mm
Comalcalco	Cuxcuchapa, Seco, Tular y Cocohital	2,052 mm
Cunduacán	Samaria, Guayabal y Cuxcuchapa	1,947 mm
Emiliano Zapata	Usumacinta y Chacamax	1,947 mm
Huimanguillo	Mezcalapa, Blasillo, Tonalá	2,290 mm
Jalapa	De la sierra, Teapa, Puyacatengo, Puente Grande y San Cristóbal	3,783 mm
Jalpa de Méndez	Nacajuca, El Naranjo, Cucuxchapa y Chiquistero	1,692 mm
Jonuta	Usumacinta, Chico, San Antonio, Palizada y San Pedro y San Pablo	2,019 mm
Macuspana	Tepetitán, Puxcatán, Tulija, Maluco y Chilapa	3,186 mm
Nacajuca	Carrizal, Samaria, González, Calzada, San Cipriano y Jahuactal	1,707 mm
Paraíso	González, Seco, El Corcho, Tupilco, Cocohital, El Corinto, Arroyo Hondo y Arroyo Hondo	1,751 mm
Tacotalpa	Puxcatan, Oxolotán, Amatán, Chinal y de la Sierra	4, 014 mm
Teapa	Teapa, Puyacatengo y Pichucalco	3,863 mm
Tenosique	Usumacinta y San Pedro Mártir	3,286 mm

Tabla 1.2: Recursos hidrológicos de los municipios de Tabasco

1.5.2. Costo de los Bienes Expuestos

La componente costo, C , de la ecuación 1.4.1 mide la cuantía de lo que es susceptible de afectarse durante la ocurrencia de un fenómeno perturbador (Ordaz 2006) [29]. Al evaluar el costo se responde ¿Cómo podría afectar el daño que provoca una inundación a los elementos expuestos?

Los datos que pueden ayudar a obtener un valor que represente al costo son: el valor de las construcciones expuestas (viviendas, fabricas, escuelas, etc.), los bienes con las que cuenta la población (enseres domésticos, automóvil, maquinaria, etc.), valores económicos de la agricultura, ganadería o algún otro sector que se desarrollan en la comunidad y el valor de infraestructura con la que cuenta la comunidad (carreteras, espacios públicos, etc.).

En esta investigación no se cuenta con toda la información para desarrollar el análisis del Costo con todas las variables mencionadas por lo que se adapta a la información concreta que se tiene disponible. Entonces el indicador de Costo se construye con dos factores: algunos de los bienes con los que cuenta la población y el número de habitantes por comunidad.

En el ¿Porqué del número de habitantes? la respuesta está basada en el supuesto que dice “A mayor población, mayor riqueza”, aunque, por conocimiento empírico se podría decir que esto no siempre es cierto. Pero dado que no se cuenta con información suficiente para medir la riqueza que existe en las comunidades, será suficiente en esta investigación analizar la información de habitantes por comunidad, además algo que se puede tomar como cierto es el hecho de que “A mayor población, mayor pérdida”.

El factor de la población será complementado con otro más, los bienes con lo que cuenta la población, para analizar este factor se construye un índice de bienes con las variables que se muestran en la tabla 1.3:

Variables	Descripción
PODT	Porcentaje de ocupantes que disponen de televisión
PODR	Porcentaje de ocupantes que disponen de refrigerador
PODL	Porcentaje de ocupantes que disponen de lavadora
PODC	Porcentaje de ocupantes que disponen de computadora
PODA	Porcentaje de ocupantes que disponen de automóvil ó camioneta

Tabla 1.3: Descripción de las variables incluidas en el índice de bienes

1.5.3. Peligro

El peligro, P , de inundación en el contexto de esta investigación se define como la probabilidad de ocurrencia de una situación de inundación potencialmente dañina en un área dada y con cierto periodo especificado. Aclarando que la medida del peligro de inundación no trasmite información sobre las consecuencias de tales inundaciones en la sociedad o en el medio ambiente, dado que estas consecuencias dependen de otros factores, como la intensidad de la

inundación, la susceptibilidad y la exposición. Por ejemplo si un área especificada se encuentra inundada y en ésta hay una casa construida con elementos estructurales y una construcción sin elementos estructurales, la inundación causará daños mayores a la segunda construcción pero seguramente el valor de la primera construcción es mayor, con esto se muestra que el riesgo envuelve diferentes elementos además del peligro de inundación, aunque, éste puede ser usado para cuantificar la intensidad del proceso de inundación (Begum *et al.*, 2007) [3].

Diferentes estudios identifican a la profundidad del agua como la característica que tiene mayor influencia sobre el daño de la inundación (Penning-Rowsell & Fordham, 1994) [30]. Un criterio más, que también se considera importante para la intensidad de la inundación, es la velocidad con la que ocurre ya que esto ocasiona una mayor vulnerabilidad. Otro indicador para la intensidad de la inundación es la duración en situación de inundación; por ejemplo en muchos casos las fallas en los niveles de los ríos o de las obras hidráulicas son ocasionadas por la duración del nivel del agua.

Para el análisis del factor de peligro son varias las fuentes de información que pueden ayudar, esto depende del tipo de inundación que se analice, pueden definirse de acuerdo con:

- La duración de la inundación, éstas pueden ser lentas o súbitas.
- El mecanismo que las genere, pueden ser pluviales, fluviales, costeras o por falla de obras hidráulicas.

En este caso se calcula un valor para el peligro tomando a la lluvia como un indicador fiable de la inundación, no se estudian las inundaciones dadas directamente por ciclones tropicales o por fallas de obras hidráulicas, aunque es importante hacer notar que cuando se presentan valores extremos de precipitación es muy probable que ocurran tanto fallas en obras hidráulicas, como ciclones tropicales. Por lo tanto, para el cálculo del factor Peligro se analizan los datos de precipitación en Tabasco y zonas cercanas, estas son consideradas ya que las inundaciones no dependen solamente de la precipitación local, sino de la precipitación en la cuenca aguas arriba del punto de desborde.

1.6. Mapas de Riesgo de Inundación

Los mapas de riesgo de inundación son un paso necesario en la construcción de estrategias para disminuir los daños provocados por las inundaciones. Además pueden ayudar en diferentes sectores como:

- Proporcionar conocimiento a la gente para tomar decisiones de mercado.
- Proveer información para el desarrollo urbano o planear inversiones.
- Ayudar a tomar medidas para el control de inundaciones como obras hidráulicas.
- Permitir preparar la forma más eficiente de manejar situaciones de emergencia.

Los mapas de inundación que incluyen información sobre las pérdidas son limitados a daños económicos, sin tomar en cuenta que existen otros tipos de daños, como los ocasionados a las personas (traumas, estrés, muertes, etcétera). Un ejemplo de mapas sobre daño a personas proviene de Jonkman *et al.* (2003), quienes estiman y mapean el riesgo individual por año para pérdidas de vida debido a ahogamiento en los países bajos.

La mayoría de los mapas se limitan a identificar las áreas de inundación para ciertos escenarios, en algunos casos dan información adicional como la profundidad del agua. Mapas que ilustren las posibles consecuencias de la inundación o información que ayude a mitigar los daños son raros.

En esta investigación se muestran los mapas de vulnerabilidad, de peligro en diferentes escenarios y finalmente uno para el riesgo, todos basados en los índices que son construidos como se explicó en este capítulo.

EMV y Pruebas de bondad de ajuste

2.1. Introducción

En este capítulo se dará un breve enfoque de alguna de las formas en que es posible modelar los eventos extremos. Continuando con un resumen del método de estimación por máxima verosimilitud y algunas de sus propiedades, así como un resumen sobre estimación por intervalos de confianza.

Se plantea una prueba bootstrap con el coeficiente de correlación muestral para las distribuciones Gumbel y Valores Extremos Generalizada. Posteriormente se calcula el tamaño de la prueba y utilizando diferentes distribuciones alternativas se calcula su potencia.

El capítulo termina con una pequeña explicación sobre la teoría de los periodos de retorno, su relación con la distribución geométrica y la forma de obtenerlos a partir de la función de distribución acumulada.

2.2. Estimación de los parámetros

Sea X_1, X_2, \dots, X_n una sucesión de variables aleatorias independientes y con la misma función de distribución F desconocida. El problema de interés es determinar el comportamiento estadístico de M_n , donde

$$M_n = \text{máx} \{X_1, X_2, \dots, X_n\} \quad (2.2.1)$$

En la sección anterior se mostraron dos teoremas que ayudan a encontrar las posibles distribuciones límite para M_n las cuales son la *DVE* y su generalización en *DVEG*. Suponga que existe $z = a_n x + b_n$ con $\{a_n > 0\}$ y $\{b_n\}$ sucesiones de constantes tales que

$$P(M_n \leq z) \approx G\left(\frac{z - b_n}{a_n}\right) \quad (2.2.2)$$

donde G es una *DVE*. Por lo tanto, se ha encontrado que la distribución de M_n se puede aproximar por una *DVE*.

Con lo anterior se tiene un enfoque para modelar eventos extremos, que es conocido como análisis de máximos por bloques. Sea x_1, x_2, \dots, x_m una muestra de alguna variable aleatoria X . Se forman bloques, que en este trabajo corresponden a un periodo igual a un año, de n observaciones formando los vectores

$$\begin{aligned}\mathbf{x}^{(1)} &= (x_1^1, \dots, x_n^1) \\ \mathbf{x}^{(2)} &= (x_1^2, \dots, x_n^2) \\ &\vdots \\ \mathbf{x}^{(k)} &= (x_1^k, \dots, x_n^k)\end{aligned}$$

Dados los máximos por bloques, específicamente los máximos anuales, esto es, $M_{n,i} = \max\{x_1^i, \dots, x_n^i\}$, $i = 1, \dots, k$. Asumiendo que tienen una *DVE* ó *DVEG*, se estiman los parámetros μ , σ y ξ .

Algunos métodos de estimación son:

- Técnicas Gráficas.
- Estimadores de momentos.
- Estimadores de momentos ponderados.
- Máxima verosimilitud.

El método de estimación que se usará es Máxima Verosimilitud, la razón es que dichos estimadores cuentan con buenas propiedades, tales como aproximaciones simples para obtener los errores estándar y los intervalos de confianza. Así como propiedades óptimas cuando el tamaño de muestra es grande y que son mencionadas más adelante.

2.2.1. Método de Máxima Verosimilitud para estimar parámetros

En primer lugar se describe en forma general el método de máxima verosimilitud y posteriormente el caso particular de la estimación de los parámetros de la *DVE*.

Este método se basa en la función de densidad conjunta de n variables aleatorias X_1, \dots, X_n , dependientes de los parámetros $\{\theta_1, \dots, \theta_m\}$, sobre los que maximizamos la función de densidad conjunta para el caso de una realización x_1, \dots, x_n .

Definición 2.1. (Función de verosimilitud) La función de verosimilitud de n variables aleatorias X_1, \dots, X_n está definida como la densidad conjunta de las n variables, es decir, $f(\mathbf{x}; \boldsymbol{\theta})$, la cual es la considerada como una función de $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$. En particular, si X_1, \dots, X_n es una muestra aleatoria de densidades $f_X(x; \boldsymbol{\theta})$, entonces la función de verosimilitud es

$$f(\mathbf{x}; \boldsymbol{\theta}) = \prod_{i=1}^n f_{X_i}(x_i; \theta_1, \dots, \theta_m).$$

Note que la función de verosimilitud es una función de $\theta_1, \dots, \theta_m$ y suele utilizarse la notación $L(\theta_1, \dots, \theta_m; x_1, \dots, x_n) = f(\mathbf{x}; \boldsymbol{\theta})$.

Definición 2.2. (Estimador de máxima verosimilitud) Sea $L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}; x_1, \dots, x_n)$ la función de verosimilitud para las variables aleatorias X_1, \dots, X_n . Si $\hat{\boldsymbol{\theta}}$ es el valor de $\boldsymbol{\theta} \in \Omega$ con el que se maximiza $L(\boldsymbol{\theta})$, entonces la variable aleatoria $\hat{\boldsymbol{\Theta}} = g(X_1, \dots, X_n)$ es el estimador de máxima verosimilitud de $\boldsymbol{\theta}$. Mientras que $\hat{\boldsymbol{\theta}} = g(x_1, \dots, x_n)$ es el valor del estimador de máxima verosimilitud de $\boldsymbol{\theta}$ para la realización x_1, \dots, x_n .

Cabe mencionar que es usual trabajar con la función log-verosimilitud, esto es,

$$l(\boldsymbol{\theta}; \mathbf{x}) = \sum_{i=1}^n \log f(x_i; \boldsymbol{\theta}). \quad (2.2.3)$$

Dado que la función logaritmo es monótona creciente, entonces maximizar $L(\boldsymbol{\theta}; \mathbf{x})$ es equivalente maximizar $l(\boldsymbol{\theta})$ para obtener el EMV $\hat{\boldsymbol{\theta}}$.

2.2.2. Propiedades de los Estimadores de Máxima Verosimilitud

Los estimadores de máxima verosimilitud tienen buenas propiedades asintóticas, cuando éstos cumplen ciertas condiciones de regularidad.

Teorema 2.1. Sea $F = \{F_\theta : \theta \in \Theta\}$ una familia de funciones de distribución, donde Θ es un subconjunto abierto de \mathbb{R} . Supóngase que las distribuciones F_θ poseen funciones de densidad $f(x; \theta)$ y que la familia F satisface las siguientes condiciones de regularidad.

1).- Para cada $\theta \in \Theta$, las derivadas

$$\frac{\partial \log [f(x, \theta)]}{\partial \theta}, \frac{\partial^2 \log [f(x, \theta)]}{\partial \theta^2}, \frac{\partial^3 \log [f(x, \theta)]}{\partial \theta^3}$$

existen para toda x .

2).- Para cada $\theta \in \Theta$,

$$0 < E_\theta \left\{ \left(\frac{\partial \log [f(X, \theta)]}{\partial \theta} \right)^2 \right\} < \infty$$

Bajo las condiciones anteriores y considerando observaciones de variables aleatorias iid de F_θ , para θ un elemento de Θ . Entonces, con probabilidad 1, las ecuaciones de verosimilitud admiten una sucesión de soluciones $\hat{\theta}_n$ tales que

$$(\hat{\theta}_n - \theta) \sqrt{nl_\theta} \xrightarrow{d} N(0, 1),$$

donde $l_\theta = -E_\theta \left(\frac{\partial^2 f(\theta; X_i)}{\partial \theta^2} \right)$ es la información esperada de una sola observación, y $I_\theta = nl_\theta$ es la función de información esperada de Fisher. Es decir, el estimador de máxima verosimil-

itud es asintóticamente normal. Además, con apropiadas generalizaciones de las condiciones anteriores, el resultado sigue siendo válido para el caso multidimensional.

Los EMV también cumplen la propiedad de invarianza como se muestra en los siguientes teoremas.

Teorema 2.2. Si $\hat{\theta}$ es el estimador de máxima verosimilitud para θ , y $\psi = h(\theta)$ es una función escalar, entonces el estimador de máxima verosimilitud de ψ está dado por $\hat{\psi} = h(\hat{\theta})$.

Teorema 2.3. Bajo las condiciones de regularidad del Teorema 2.1, el estimador de máxima verosimilitud $\hat{\theta}$ es asintóticamente normal y denotamos a su matriz de covarianza asintótica como V_{θ} . Si $\psi = h(\theta)$ es una función escalar, entonces el estimador de máxima verosimilitud $\hat{\psi} = h(\hat{\theta})$ tiene una distribución aproximada $N(\psi, V_{\psi})$, donde

$$V_{\psi} = \Delta\psi^t V_{\theta} \Delta\psi,$$

con

$$\Delta\psi = \left[\frac{\partial\psi}{\partial\theta_1}, \dots, \frac{\partial\psi}{\partial\theta_k} \right]$$

evaluado en $\hat{\theta}$.

2.2.3. Criterio de Información de Akaike (AIC)

El criterio de información de Akaike (AIC) propuesta por Akaike (1974) es una medida que ayuda en la selección de un modelo. El AIC no es una prueba en el sentido de una prueba de hipótesis, más bien es una herramienta que se usa para medir la calidad del ajuste del modelo estadístico estimado. Cuando los modelos estadísticos son estimados mediante máxima verosimilitud el AIC se obtiene con la siguiente expresión:

$$AIC(k) = -2 \log(L(\theta_k)) + 2k \quad (2.2.4)$$

donde, k es el número de parámetros del modelo y $\log(L(\theta_k))$ es el logaritmo de la verosimilitud del modelo. El AIC es un criterio útil para comparar modelos estimados, que utiliza como criterio de decisión aquel modelo que cuente con el menor valor de AIC.

2.3. Intervalos de Confianza

Supóngase que se tiene una población de la cual se desconoce su parámetro θ , y que bajo ciertas condiciones, se encuentra que $\theta \in (\hat{\theta}_i, \hat{\theta}_s)$, donde los puntos extremos $\hat{\theta}_i$ y $\hat{\theta}_s$ dependen del valor de la estadística $\hat{\Theta}$ para una realización particular de una muestra aleatoria y, se conocen como extremo inferior y extremo superior, respectivamente. Debido a que los extremos del intervalo, $\hat{\theta}_i$ y $\hat{\theta}_s$, dependen de la realización de la muestra, resulta que sólo son valores particulares de dos variables aleatorias $\hat{\Theta}_i$ y $\hat{\Theta}_s$, respectivamente. Las cuales son función de la estadística $\hat{\Theta}$

correspondiente a la muestra aleatoria y a su distribución muestral. Con base en las variables aleatorias anteriores podemos calcular la probabilidad de que el parámetro θ se encuentre en el intervalo establecido. Es decir, si se simboliza por $1 - \alpha$ con $\alpha \in (0, 1)$ a la probabilidad mencionada se tendrá:

$$P\left(\hat{\Theta}_i < \theta < \hat{\Theta}_s\right) = 1 - \alpha. \quad (2.3.1)$$

Al intervalo en el que está localizado el parámetro θ , correspondiente a una realización de la muestra $\hat{\theta}_i < \theta < \hat{\theta}_s$, se le llama intervalo de confianza al $(1 - \alpha)100\%$ de confianza, mientras que la fracción $1 - \alpha$ se le da el nombre de nivel de confianza y los extremos $\hat{\theta}_i$ y $\hat{\theta}_s$ se llaman límites de confianza, inferior y superior, respectivamente.

Para la estimación de los intervalos de confianza de los parámetros de la *DVEG* se recurre al teorema 2.1, en el cual se tiene que para un tamaño de muestra grande el estimador de máxima verosimilitud se distribuye asintóticamente normal, esto es

$$\hat{\boldsymbol{\theta}} \sim N_d(\boldsymbol{\theta}, \mathbf{I}^{-1}(\boldsymbol{\theta})) \quad (2.3.2)$$

es decir, $\mathbf{I}^{1/2}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \rightarrow N_d(\boldsymbol{\theta}, \mathbf{I}_d)$, en donde $\mathbf{I}(\boldsymbol{\theta})$ es la matriz de información de Fisher y \mathbf{I}_d es la matriz identidad de orden $d \times d$. También se tiene:

$$\begin{aligned} \hat{\boldsymbol{\theta}} &\sim N_d(\boldsymbol{\theta}, \mathbf{I}^{-1}(\hat{\boldsymbol{\theta}})) \\ \hat{\boldsymbol{\theta}} &\sim N_d(\boldsymbol{\theta}, \mathbf{J}^{-1}(\hat{\boldsymbol{\theta}})) \end{aligned}$$

donde $\mathbf{J}(\boldsymbol{\theta})$ es el valor esperado de la matriz de información de Fisher.

Entonces, el intervalo de confianza al $(1 - \alpha)100\%$ para la k -ésima componente de $\boldsymbol{\theta}$ está dado por

$$\hat{\theta}_k \pm Z_{\alpha/2} \sqrt{c_{kk}} \quad (2.3.3)$$

en donde, $Z_{\alpha/2}$ es el valor de la variable normal estándar, y c_{kk} es el k -ésimo componente de la diagonal de $\mathbf{J}^{-1}(\hat{\boldsymbol{\theta}})$ o $\mathbf{I}^{-1}(\hat{\boldsymbol{\theta}})$.

2.4. Pruebas de Bondad de Ajuste

El término “bondad de ajuste” fue introducido por Pearson (1902), mientras que por Prueba de bondad de ajuste entenderemos a un procedimiento estadístico que sirve para probar que una función de distribución de probabilidades particular ajusta o no a un conjunto de datos, su objetivo es medir la concordancia de algunos datos con una distribución teórica determinada.

Sea X_1, \dots, X_n una muestra aleatoria de una función de distribución $F_X(x) = P(X \leq x)$. El juego de hipótesis a probar está dado por:

$$H_0 : f(x) \in F^*(x) \text{ vs } H_1 : f(x) \notin F^*(x), \quad (2.4.1)$$

donde $F^*(x)$ denota una familia de densidades específica, en este caso se trata de la función de

densidad de Valores Extremos Generalizada, con función de distribución mostrada en **2.4.2** y la Gumbel con función de distribución dada en **2.4.3**.

$$F(x) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]_+^{-\frac{1}{\xi}} \right\} \quad (2.4.2)$$

$$G(x) = \exp \left\{ - \exp \left[- \left(\frac{x - \mu}{\sigma} \right) \right] \right\} \quad (2.4.3)$$

para $x \in \mathbb{R}$ en ambas distribuciones.

2.4.1. Prueba Bootstrap basada en el coeficiente de Correlación muestral

Los principios generales de las pruebas de bondad de ajuste son conocidos, pero las dificultades que se tienen en un caso particular son diversas. Una de tales dificultades se refiere a los estimadores de los parámetros de la distribución del problema. Para el caso de las distribuciones en estudio el problema puede simplificarse, porque ambas distribuciones pertenecen a la familia loc-escala, cuando se fija un valor para el parámetro de forma. Así, con base en la invarianza bajo la transformación de los parámetros de localidad y escala, será posible establecer una prueba basada en el coeficiente de correlación muestral como una estadística de prueba.

NOTA 2.1. En el estudio de las distribuciones puede observarse que existen variables aleatorias X , cuya distribución depende de los parámetros de localidad y escala denotados por μ y σ , respectivamente, pero la variable resultante de la transformación $Y = \frac{X - \mu}{\sigma}$, no depende de los parámetros. A este tipo de distribuciones se les llama de localidad y escala, o en forma simple **loc-escala**.

En la construcción de la prueba para **2.4.1**, se necesita la función de distribución empírica. Además, si H_0 de **(2.4.1)** se cumple, entonces la distribución de $f(x)$ es tal que $F(x; \mu, \sigma, \xi) = F_* \left(\frac{x - \mu}{\sigma}; \xi \right)$. Entonces para una variable aleatoria X con densidad $f(x)$ se tiene que la función $F_*(x; \xi)$ le corresponde su función inversa $F_*^{-1}(y; \xi)$.

Sean $X_{(1)}, \dots, X_{(n)}$ las observaciones ordenadas de manera creciente. La función de distribución empírica se define como la frecuencia acumulada porcentual de los datos ordenados en forma ascendente y representa la probabilidad de ocurrencia de un valor menor o igual al dato considerado.

Una expresión para la función de distribución empírica está dado por:

$$F_n(x_{(i)}) = \begin{cases} 0 & \text{si } x < x_{(1)} \\ \frac{i+0.5}{n} & \text{si } x_{(i)} \leq x < x_{(i+1)} \\ 1 & \text{si } x > x_{(n)} \end{cases} \quad (2.4.4)$$

donde $x_{(i)}$ es la i -ésima estadística de orden.

Ahora, por el teorema de Glivenko-Cantelli, se establece una convergencia casi segura, cuando $n \rightarrow \infty$, entre la distribución empírica $F_n(X)$ y la distribución teórica $F(x)$ y $G(x)$, por lo tanto, es factible establecer, para la DVEG:

$$\begin{aligned}
 G_n(x) &\approx G(x) \\
 G_n(x) &\approx \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]_+^{-\frac{1}{\xi}} \right\} \\
 \log [G_n(x)] &\approx - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \\
 [-\log [G_n(x)]]^{-\xi} &\approx 1 + \xi \left(\frac{x - \mu}{\sigma} \right) \\
 \frac{1}{\xi} [-\log [G_n(x)]]^{-\xi} - \frac{1}{\xi} &\approx \frac{1}{\sigma} x - \frac{\mu}{\sigma}
 \end{aligned}$$

en cuanto a la distribución Gumbel se establece lo siguiente:

$$\begin{aligned}
 F_n(x) &\approx F(x) \\
 F_n(x) &\approx \exp \left\{ - \exp \left[- \left(\frac{x - \mu}{\sigma} \right) \right] \right\} \\
 -\log [F_n(x)] &\approx \exp \left[- \left(\frac{x - \mu}{\sigma} \right) \right] \\
 \log \{ -\log [F_n(x)] \} &\approx - \left(\frac{x - \mu}{\sigma} \right)
 \end{aligned}$$

Ambas expresiones se pueden reescribir de la siguiente manera:

$$Y = \beta + \alpha Z$$

Para la distribución de valores extremos generalizada se tendrá:

$$\begin{aligned}
 Y_2 &= \frac{1}{\xi} (-\log [G_n(x)])^{-\xi} - \frac{1}{\xi} \\
 Z_2 &= x \\
 \beta &= -\frac{\mu}{\sigma} \\
 \alpha &= \frac{1}{\sigma}.
 \end{aligned}$$

Para la distribución Gumbel se tiene:

$$\begin{aligned} Y_1 &= \log(-\log[F_n(x)]) \\ Z_1 &= x \\ \beta &= \frac{\mu}{\sigma} \\ \alpha &= -\frac{1}{\sigma} \end{aligned}$$

Se espera que bajo H_0 exista una fuerte relación lineal entre Y y Z , aún si se sustituye al parámetro ξ por un estimador consistente ($\hat{\xi}$). Entonces para probar (2.4.1) se propone como estadística de prueba al coeficiente de correlación muestral, dado en la expresión 2.4.5:

$$\text{Corr}(\mathbf{Y}, \mathbf{Z}) = r_n(\mathbf{Y}, \mathbf{Z}) = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(Z_i - \bar{Z})}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2 \sum_{i=1}^n (Z_i - \bar{Z})^2}} \quad (2.4.5)$$

el cual se espera que bajo H_0 de (2.4.1) la distribución de $r_n(\mathbf{Y}, \mathbf{Z})$ estará concentrada muy cerca de 1. Por lo tanto, la regla de decisión para una prueba de tamaño α , con $\alpha \in (0, 1)$ un valor conocido, consiste en rechazar H_0 cuando $r_n(\mathbf{Y}, \mathbf{Z}) < r_\alpha$.

Proposición 2.4. *Sea W_1, W_2, \dots, W_n una muestra aleatoria con parámetros de localidad ω y escala τ ; y V_1, V_2, \dots, V_n otra muestra aleatoria con parámetros de localidad v y escala η , entonces el coeficiente de correlación muestral es invariante bajo transformaciones de localidad y escala.*

Demostración. Sea el coeficiente de correlación muestral dado por 2.4.5, y $W_i^* = \frac{W_i - \omega}{\tau}$ y $V_i^* = \frac{V_i - v}{\eta}$. Entonces, calcule $r_n(\mathbf{W}^*, \mathbf{V}^*)$. Para esto, se tiene que: $W_i^* - \bar{W}^* = \frac{W_i - \omega}{\tau} - \frac{\bar{W} - \omega}{\tau} = \frac{W_i - \bar{W}}{\tau}$ y $V_i^* - \bar{V}^* = \frac{V_i - v}{\eta} - \frac{\bar{V} - v}{\eta} = \frac{V_i - \bar{V}}{\eta}$. Sustituyendo las igualdades anteriores en $r_n(\mathbf{W}^*, \mathbf{V}^*)$, se tendrá

$$r_n(\mathbf{W}^*, \mathbf{V}^*) = \frac{\sum_{i=1}^n \left(\frac{W_i - \bar{W}}{\tau} \right) \left(\frac{V_i - \bar{V}}{\eta} \right)}{\sqrt{\sum_{i=1}^n \left(\frac{W_i - \bar{W}}{\tau} \right)^2 \sum_{i=1}^n \left(\frac{V_i - \bar{V}}{\eta} \right)^2}} = \frac{\sum_{i=1}^n (W_i - \bar{W})(V_i - \bar{V})}{\sqrt{\sum_{i=1}^n (W_i - \bar{W})^2 \sum_{i=1}^n (V_i - \bar{V})^2}} = r_n(\mathbf{W}, \mathbf{V}).$$

□

Se tiene que el estadístico de prueba depende únicamente del parámetro de forma ξ , el cual se sustituye por su estimador $\hat{\xi}$ obtenido por el método de máxima verosimilitud. La regla de decisión es rechazar H_0 de (2.4.1) con un nivel de significancia α si $r_n \leq C_n(\alpha)$, donde $C_n(\alpha)$ es tal que $\alpha = \max_{\xi} P(\text{Rechazar } H_0 | H_0) = \max_{\xi} P(r_n \leq C_n(\alpha))$. En caso de no conocer en forma analítica la distribución del estadístico la prueba se puede llevar a cabo mediante un algoritmo Bootstrap, como el propuesto a continuación.

Se obtiene la distribución de r_n bajo H_0 para valores fijos del parámetro de forma ξ a través de un muestreo bootstrap y con base en los estimadores de MV se realiza siguiendo los siguientes pasos:

- 1).- Dadas las observaciones x_1, \dots, x_n calcular un valor para el estimador de ξ , denotado por $\tilde{\xi}_0$.
- 2).- Con las observaciones se calcula su coeficiente de correlación muestral, r_0 .
- 3).- A partir de $\tilde{\xi}_0$ iniciar un ciclo de bootstrap.
 - a).- Generar una muestra bootstrap de tamaño n de la DVEG o Gumbel, según sea el caso.
 - b).- Con la muestra de bootstrap del inciso (3a) calcular una estimación para el parámetro de forma, denotándola por $\tilde{\xi}_1$.
 - c).- Con $\tilde{\xi}_1$ del inciso (3b) se calcula el coeficiente de correlación muestral, \hat{r}_1 .
- 4).- El ciclo de bootstrap del inciso 3 se repite m veces para calcular $\hat{r}_1, \dots, \hat{r}_m$, en donde m -cantidad de estimaciones de bootstrap, para determinar el cuantil bootstrap, generalmente se recomienda $m \geq 1000$.
- 5).- Con el ciclo terminado, $\hat{r}_1, \dots, \hat{r}_m$ son ordenados en forma no decreciente, denotándolos con \hat{r}_i . Entonces, $\hat{r}_1 \leq \hat{r}_2 \leq \dots \leq \hat{r}_m$ y obtener el cuantil α , sea éste \hat{r}_α .
- 6).- **Regla de decisión.** Comparar r_0 con el cuantil α del inciso 5.
 - a).- Si $r_0 \leq \hat{r}_\alpha$, se rechaza H_0 al nivel de significancia α .
 - b).- Si $r_0 > \hat{r}_\alpha$, no se rechaza H_0 al nivel de significancia α .

Tamaño de la prueba

Antes de iniciar con las pruebas se debe determinar si éstas tienen un tamaño de prueba adecuado, el tamaño de una prueba bootstrap se calcula por simulación con la probabilidad del error tipo I, es decir considerando que los datos provienen de una muestra de variables aleatorias DVEG para un valor de ξ dado o de una distribución Gumbel.

Se programó una función en el proyecto R para determinar el tamaño de la prueba, como se describe a continuación:

- 1).- Se genera una muestra x_1, \dots, x_n que pertenezca a la $DVEG(\mu, \sigma, \xi)$ ó a la distribución $Gumbel(\mu, \sigma)$.
- 2).- Con la muestra generada en 1 calcular un valor para el estimador de ξ , denotado por $\tilde{\xi}_0$.

- 3).- Con la muestra de 1 y $\tilde{\xi}_0$ se calcula el coeficiente de correlación muestral, \hat{r}_0 .
- 4).- A partir de $\tilde{\xi}_0$ iniciar un ciclo de bootstrap.
- a).- Generar una muestra bootstrap de tamaño n de la DVEG ó Gumbel, según sea el caso.
- b).- Con la muestra de bootstrap del inciso (4a) calcular una estimación para el parámetro de forma, denotándola por $\tilde{\xi}_1$.
- c).- Con $\tilde{\xi}_1$ del inciso (4b) se calcula el coeficiente de correlación muestral, \hat{r}_1 .
- 5).- El ciclo de bootstrap del inciso 4 se repite m veces para calcular $\hat{r}_1, \dots, \hat{r}_m$, en donde m -cantidad de estimaciones de bootstrap, para determinar el cuantil bootstrap, generalmente se recomienda $m \geq 1000$.
- 6).- Con el ciclo terminado, $\hat{r}_1, \dots, \hat{r}_m$ son ordenados en forma no decreciente, denotándolos con \hat{r}_i . Entonces, $\hat{r}_1 \leq \hat{r}_2 \leq \dots \leq \hat{r}_m$ y obtener los cuantiles requeridos. En este caso se tomaron los cuantiles 0.01, 0.025, 0.05, 0.075 y 0.1, que se denotan en general \hat{r}_{c_i} para $c_i = 0.01, 0.025, 0.05, 0.075$ y 0.1.
- 7).- Comparar \hat{r}_0 con los cuantiles del inciso 6.
- a).- Si $\hat{r}_0 \leq \hat{r}_{c_i}$, se asigna 1 en el cuantil correspondiente.
- b).- Si $\hat{r}_0 > \hat{r}_{c_i}$, se asigna 0 en el cuantil correspondiente.
- 8).- Repetir los incisos desde 1 hasta 7, M veces. Finalmente obtener los promedios de los resultados de las comparaciones hechas en (7) por cuantil de estas repeticiones, dichos promedios son los tamaños estimados de prueba buscados.

Para obtener los tamaños estimados de la prueba, con la metodología anterior se utilizaron $m = 3000$ y $M = 1000$ repeticiones y tamaños de muestra $n = 30, 50, 75, 100$ y 150, se analizó para los valores de $\xi = 0.1, 0.5$ y 1.

Para la DVEG se obtuvieron los resultados que se muestran en las tablas 2.1, 2.2 y 2.3

n	0.010	0.025	0.050	0.075	0.100
30	0.012	0.026	0.040	0.060	0.086
50	0.008	0.016	0.030	0.046	0.064
75	0.002	0.010	0.020	0.036	0.050
100	0.004	0.016	0.026	0.046	0.062
150	0.004	0.010	0.044	0.058	0.070

Tabla 2.1: Tamaños de prueba para $\xi = 0.1$

n	0.010	0.025	0.050	0.075	0.100
30	0.004	0.011	0.039	0.059	0.084
50	0.004	0.012	0.038	0.063	0.095
75	0.006	0.018	0.038	0.058	0.086
100	0.006	0.018	0.041	0.059	0.085
150	0.007	0.020	0.046	0.069	0.096

Tabla 2.2: Tamaños de prueba para $\xi = 0.5$

n	0.010	0.025	0.050	0.075	0.100
30	0.006	0.014	0.026	0.046	0.084
50	0.004	0.010	0.038	0.032	0.075
75	0.002	0.018	0.036	0.050	0.070
100	0.006	0.018	0.041	0.062	0.092
150	0.007	0.020	0.042	0.064	0.080

Tabla 2.3: Tamaños de prueba para $\xi = 1$

Para la distribución Gumbel, dado que no cuenta con parámetro de forma, se tiene únicamente la tabla 2.4 de los resultados del tamaño de prueba.

n	0.010	0.025	0.050	0.075	0.100
30	0.008	0.023	0.039	0.060	0.082
50	0.010	0.030	0.055	0.080	0.100
75	0.012	0.021	0.043	0.069	0.097
100	0.010	0.018	0.046	0.078	0.100
150	0.011	0.029	0.053	0.077	0.104

Tabla 2.4: Tamaños de prueba para la distribución Gumbel

NOTA 2.2. De los tamaños de prueba obtenidos se concluye lo siguiente:

- De las tablas 2.1, 2.2 y 2.3 se observa que para los diferentes valores de ξ la prueba bootstrap para la DVEG funciona bien, aunque al aumentar el tamaño de la muestra la prueba rebasa ligeramente el valor de α .
- En cuanto a la prueba bootstrap para la distribución Gumbel, en la tabla 2.4, los valores del tamaño de prueba son buenos, pero al igual que con la DVEG al aumentar el tamaño de muestra se rebasa el valor de α .

Potencia de la Prueba

Una vez que se ha revisado el tamaño de la prueba, para la prueba basada en el coeficiente de correlación, el paso siguiente es determinar la potencia de la prueba, determinar si los datos muestrales analizados pertenecen a la familia VEG ó Gumbel. Ya que se trata de una prueba bootstrap, se proporciona un programa en el Proyecto R que calcule la potencia de la prueba, el programa es similar al que se hace para obtener el tamaño de la prueba la diferencia que tiene es que se reemplaza el generador de números aleatorias de la distribución Gumbel o DVEG en el programa de simulación con los generadores de distribuciones alternativas:

- Función de densidad Weibull,

$$f(x; \lambda, \beta) = \beta \lambda^\beta x^{\beta-1} \exp \{ -(\lambda x)^\beta \} I_{(0, \infty)}(x), \quad \beta > 0, \quad \lambda > 1.$$

- Función de densidad normal

$$f(x; \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\}, \quad x, \mu \in \mathbb{R}, \quad \sigma > 0.$$

- Función de densidad Cauchy estándar

$$f(x; \theta, \sigma) = \frac{1}{\pi \sigma \left[1 + \left(\frac{x - \theta}{\sigma} \right)^2 \right]}, \quad x, \theta \in \mathbb{R}, \quad \sigma > 0.$$

- Función de densidad t -student

$$f(x; v) = \frac{\Gamma \left(\frac{v+1}{2} \right)}{\Gamma \left(\frac{v}{2} \right) \sqrt{v\pi} \left(1 + \left(\frac{x^2}{v} \right) \right)^{\frac{v+1}{2}}}, \quad x \in \mathbb{R}, \quad v = 1, 2, 3, \dots$$

- Función de densidad Exponencial

$$f(x; \beta) = \frac{1}{\beta} e^{-x/\beta} I_{(0, \infty)}(x), \quad \beta > 0.$$

En las tablas 2.5 a 2.9 se presenta la potencia de la prueba para ambas distribuciones, Gumbel y DVEG, con diferentes valores de significancia (0.01, 0.025, 0.05, 0.075, 0.1) y tamaños de muestra (30, 50, 75, 100 y 150), aplicados a cada una de las distribuciones mostradas anteriormente.

Las tablas 2.5 a 2.7 están dadas de la siguiente forma, por columna se muestra el nivel de significancia y por filas las potencias de la prueba. Para cada nivel de significancia la columna izquierda corresponde a la prueba para la DVEG y la columna derecha la prueba para la distribución Gumbel.

n	0.010		0.025		0.050		0.075		0.100	
30	0.658	0.838	0.764	0.898	0.824	0.928	0.862	0.950	0.882	0.960
50	0.884	0.964	0.930	0.984	0.964	0.990	0.976	0.992	0.981	0.996
75	0.960	0.992	0.986	0.998	0.996	1.000	0.998	1.000	1.000	1.000
100	0.988	0.998	0.996	1.000	0.998	1.000	0.998	1.000	1.000	1.000

Tabla 2.5: Potencia para la distribución alternativa Cauchy con $\theta = 0$ y $\sigma = 2$

n	0.010		0.025		0.050		0.075		0.100	
30	0.376	0.100	0.532	0.212	0.662	0.330	0.734	0.406	0.774	0.470
50	0.534	0.144	0.724	0.300	0.850	0.476	0.900	0.570	0.924	0.652
75	0.536	0.256	0.824	0.504	0.930	0.682	0.956	0.766	0.964	0.828
100	0.552	0.378	0.902	0.672	0.976	0.812	0.992	0.880	0.992	0.916
150	0.574	0.630	0.972	0.856	0.994	0.958	0.998	0.978	1.000	0.984

Tabla 2.6: Potencia para la distribución alternativa Normal estándar

n	0.010		0.025		0.050		0.075		0.100	
30	0.398	0.268	0.526	0.372	0.610	0.464	0.660	0.514	0.704	0.554
50	0.396	0.348	0.582	0.470	0.704	0.596	0.782	0.664	0.814	0.716
75	0.420	0.538	0.654	0.696	0.790	0.780	0.856	0.846	0.892	0.874
100	0.604	0.602	0.785	0.752	0.862	0.838	0.894	0.886	0.930	0.908
150	0.781	0.778	0.926	0.912	0.966	0.954	0.987	0.976	0.992	0.984

Tabla 2.7: Potencia para la distribución alternativa t-student con $\nu = 4$

n	0.010	0.025	0.050	0.075	0.100
30	0.084	0.142	0.226	0.270	0.306
50	0.050	0.146	0.240	0.308	0.358
75	0.028	0.128	0.268	0.342	0.392
100	0.028	0.164	0.304	0.388	0.446
150	0.020	0.296	0.516	0.610	0.644

Tabla 2.8: Potencia para la distribución alternativa Weibull con $\beta = 2$ y $\lambda = 2$

n	0.010	0.025	0.050	0.075	0.100
30	0.012	0.056	0.094	0.148	0.218
50	0.024	0.056	0.152	0.240	0.296
75	0.054	0.126	0.222	0.330	0.386
100	0.062	0.118	0.260	0.358	0.444
150	0.080	0.162	0.350	0.510	0.638

Tabla 2.9: Potencia para la distribución alternativa Exponencial con $\beta = 0.1$

NOTA 2.3. De las potencias de las pruebas anteriores se tiene que:

- En las primeras tres tablas se muestran las potencias de la prueba de ambas distribuciones, Gumbel y VEG, en 2.6 y 2.7 las potencias resultan mejor para el caso de la DVEG, mientras que en 2.5 resultan mejor los resultados de la distribución Gumbel.
- Para el caso de las distribuciones alternativas Weibull y Exponencial los resultados fueron los mismos para las dos distribuciones, Gumbel y VEG, en las tablas 2.8 y 2.9 se observa que las potencias para ambas pruebas son un poco bajas.

2.5. Niveles y Periodos de Retorno

En el análisis de eventos extremos resulta importante conocer el tiempo promedio de espera para que ocurra un evento extremo específico.

Definición 2.3. Sea A un evento y T el tiempo aleatorio entre ocurrencias sucesivas de A , entonces se dice que el valor medio de T es el **período de retorno** de A . Es decir, el período de retorno es el tiempo medio para que se repita un evento semejante.

Sea X_1, X_2, \dots una sucesión de variables aleatorias iid con función de distribución continua F y $u \in \mathbb{R}$. Considerando $\{X_i > u\}$ un evento extremo, se define el tiempo de la primera excedencia como

$$T(u) = \min\{i \geq 1 : X_i > u\} \quad (2.5.1)$$

entonces, el tiempo promedio de espera para que ocurra el evento extremo es $E[T(u)]$. Además, si $p = 1 - F(u)$, se obtiene

$$P(T(u) = k) = (1 - p)^{k-1}p, k = 1, 2, \dots, \quad (2.5.2)$$

por lo tanto, $T(u)$ es una variable aleatoria geométrica y $E[T(u)] = p^{-1}$. Se define a u como el nivel de retorno con un periodo de retorno p^{-1} para los eventos $\{X_i > u\}$. Note que u es el cuantil de F con una probabilidad $1 - p$.

2.5.1. Relación con la distribución de probabilidad acumulada

Si la probabilidad del evento $\{X_i > u\}$ es p , su período de retorno es p^{-1} bajo las siguientes condiciones:

- 1).- La probabilidad de que un evento ocurra durante un período corto de tiempo es pequeña.
- 2).- La probabilidad que más de un evento ocurra durante un período corto de tiempo es despreciable.

Luego, para el evento $\{X_i > u\}$, el periodo de retorno es

$$\tau_x = \frac{1}{1 - F(x)}. \quad (2.5.3)$$

2.5.2. Distribución geométrica

Si la probabilidad que ocurra el evento $\{X_i > u\}$ en un intervalo de tiempo determinado es p , la probabilidad que el siguiente evento ocurra en x intervalos tiene una función de probabilidad geométrica

$$P(x) = p(1 - p)^{x-1}, \quad x = 1, 2, 3, \dots \quad (2.5.4)$$

El valor esperado de intervalos entre eventos sucesivos es $\mu = p^{-1}$. Por lo tanto de acuerdo a la definición dada el período de retorno es p^{-1} .

Por lo anterior resulta de interés estudiar los cuantiles extremos superiores de la distribución *VEG*. Específicamente, el “nivel de retorno” asociado con el “periodo de retorno” de $1/p$ es el $(1 - p)$ -ésimo cuantil de la distribución *VEG* (por ejemplo, cuando modelamos máximos anuales, $p = 0.01$ correspondería a un periodo de retorno de 100 años) un concepto muy utilizado en hidrología (katz *et al.* 2002).

Se puede obtener una expresión para los cuantiles invirtiendo la función de distribución, es decir, $G^{-1}(1 - p; \mu, \sigma, \xi)$, se tiene entonces, $x_p = G^{-1}(1 - p)$.

En cuanto a la distribución de *VEG* se tiene que el cuantil $1 - p$ está dado por:

$$x_p = \begin{cases} \mu - \frac{\sigma}{\xi} \left[1 - \{-\log(1 - p)\}^{-\xi} \right] & \text{si } \xi \neq 0 \\ \mu + \sigma [-\log \{-\log(1 - p)\}] & \text{si } \xi = 0, \end{cases} \quad (2.5.5)$$

donde $G(x_p) = 1 - p$ con $0 < p < 1$. En términos de valores extremos, x_p es el nivel de retorno asociado con el periodo de retorno $1/p$. Para el cálculo de los cuantiles se emplean los estimadores de los parámetros.

Modelación del riesgo de inundaciones en Tabasco

3.1. Introducción

En este capítulo se muestra la información que será analizada para el desarrollo de cada uno de los tres componentes que influyen en el análisis del riesgo de inundación en Tabasco. Además se darán los resultados finales de los indicadores para los factores de vulnerabilidad y costo. El factor vulnerabilidad va acompañado de un mapa de vulnerabilidad por municipio.

Continuando con el factor peligro, se explicará la forma en que se obtuvieron los datos para su análisis y se mostrarán los resultados de los ajustes hechos a los máximos de precipitación para cada municipio de Tabasco, sus respectivos periodos de retorno y cuatro mapas de diferentes escenarios de peligro.

Finalmente se muestra el resultado de la unión de estos tres factores, el índice de riesgo por inundación a escala municipal con su respectivo mapa del índice de riesgo en el estado de Tabasco.

3.2. Resultados del Factor de Vulnerabilidad

Como se mencionó en el capítulo 1 el factor de vulnerabilidad se compone de dos factores que pueden clasificarse como social y natural. En lo que respecta al factor social, éste se refiere al indicador de servicios (Indserv) que se obtiene a partir de una combinación lineal de tres variables (PO_NDEE, PO_NDD y PO_NDAE), definidas en la tabla 1.1. Para obtener los pesos de la combinación lineal para cada una de las variables se realiza un análisis de componentes principales.

Por su parte la información de la disposición de servicios por municipio para calcular los pesos fue obtenida de la base de datos de INEGI que corresponde al censo de población y vivienda 2010. Las tres variables utilizadas en este indicador son conmensurables, por lo tanto

el análisis de componentes principales se desarrolla con la matriz de covarianza Σ , la cual se muestra a continuación:

$$\Sigma = \begin{bmatrix} 1.576 & 5.995 & 1.637 \\ 5.995 & 287.865 & 25.827 \\ 1.637 & 25.827 & 8.961 \end{bmatrix}$$

Con la matriz Σ se calculan los valores y vectores propios (en el anexo D se muestra una breve explicación de la matriz de covarianzas, así como de los valores y vectores propios), los cuales ayudan a obtener los valores de las *CP*; los resultados son los siguientes:

VALORES PROPIOS			
290.367	6.808	1.228	

VECTORES PROPIOS			
	C1	C2	C3
PO_NDEE	0.021	0.198	0.980
PO_NDAE	0.996	-0.094	-0.003
PO_NDD	0.091	0.976	-0.199

Tabla 3.1: Valores y Vectores Propios Correspondientes a las Variables del Índice de Servicios.

Nótese que el primer valor propio es el más grande y con éste se cubre el 97% de la varianza total, por lo tanto los vectores propios correspondientes a este valor formarán la primera componente principal y sus elementos serán las ponderaciones de las tres variables respectivamente.

Continuando con el estudio de la vulnerabilidad, se pasa al factor natural en el que se consideraron los principales recursos hidrológicos con los que cuenta cada municipio (1.2), se realizó una ponderación que toma en cuenta el número de ríos que forman parte del municipio y la capacidad que tiene cada río, ésta fue medida por el nivel de agua máximo ordinario (NAMO).

Finalmente para obtener el indicador de vulnerabilidad (InVuln) se ponderaron ambos indicadores, Inserv e Inríos, dando mayor peso al indicador que cuenta con mayor sustento de información, éste fue el indicador de servicios por lo que se le dio un peso de 0.6 y de 0.4 al indicador de ríos.

Los resultados obtenidos para el índice de vulnerabilidad se muestran en la tabla 3.2 mientras que en la figura 3.1 se muestra el mapa de vulnerabilidad.

Municipio	Inserv	Inríos	InVuln	Categoría
Balancán	0.282	0.128	0.221	Medio
Cárdenas	0.403	0.785	0.556	Muy Alto
Centla	1.000	0.161	0.664	Muy Alto
Centro	0.000	0.957	0.526	Muy Alto
Comalcalco	0.242	0.552	0.366	Alto
Cunduacán	0.531	0.785	0.632	Muy Alto
E. Zapata	0.001	0.161	0.065	Bajo
Huimanguillo	0.810	0.101	0.526	Muy Alto
Jalapa	0.123	0.827	0.405	Alto
Jalpa	0.166	0.684	0.373	Alto
Jonuta	0.123	0.161	0.138	Bajo
Macuspana	0.496	0.665	0.564	Muy Alto
Nacajuca	0.016	0.392	0.166	Bajo
Paraíso	0.112	1.000	0.467	Alto
Tacotalpa	0.367	0.000	0.220	Medio
Teapa	0.166	0.060	0.124	Bajo
Tenosique	0.256	0.161	0.218	Bajo

Tabla 3.2: Índice de Vulnerabilidad.

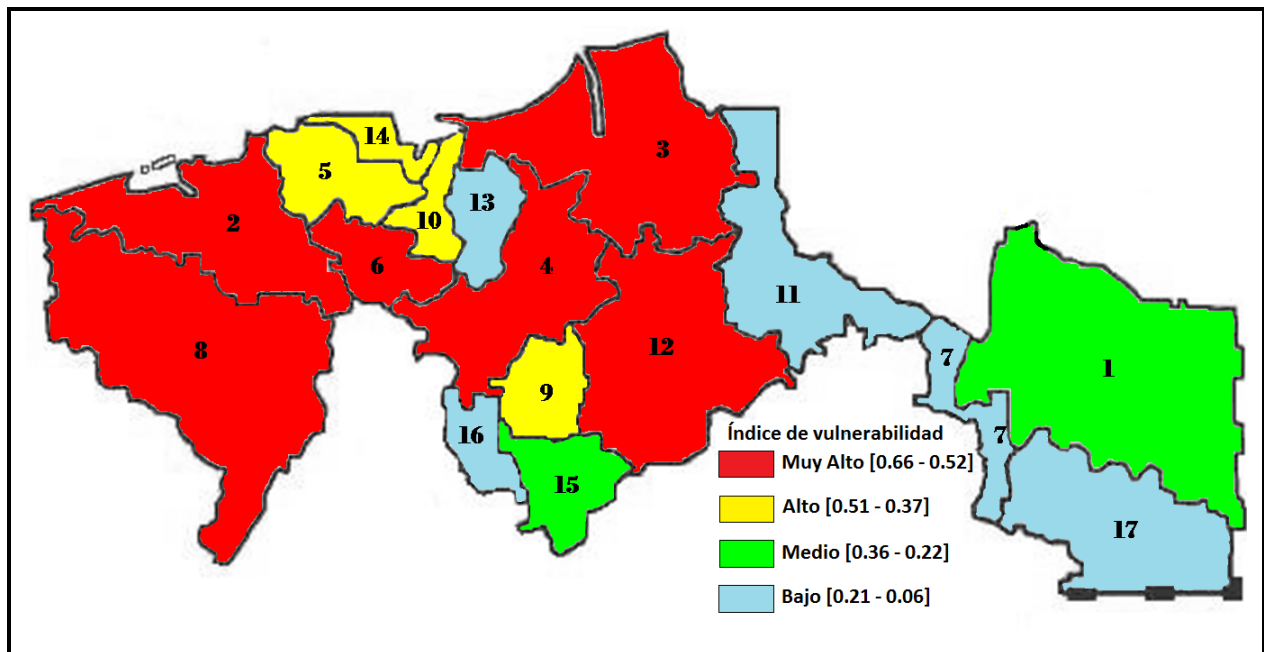


Figura 3.1: Índice de Vulnerabilidad

3.3. Factor de Costo

Para el análisis del factor de Costo intervienen dos variables de interés, en primer lugar se tiene el logaritmo del número de habitantes en cada municipio, esto basado principalmente en el supuesto de que a mayor población mayor riqueza. Se toma el logaritmo de la población (logpob) únicamente para mantener una escala más apropiada.

La segunda variable que se toma en cuenta es un indicador de los bienes (Inbienes) con los que cuenta la población del municipio en estudio. Ahora, para la construcción de este indicador se consideraron cinco bienes que estarían expuestos a sufrir cierto daño ante una inundación, éstos son: televisión (PODT), refrigerador (PODR), lavadora (PODL), computadora (PODC) y automóvil o camioneta (PODA). Entonces en el cálculo del indicador de bienes se consideraron los porcentajes de la población que cuenta con tal o cual bien.

Para obtener una combinación lineal adecuada de estas cinco variables se realiza un análisis de CP para generar los pesos adecuados para cada variable. Al igual que con los servicios, se calcula la matriz de covarianzas, ya que los datos también son conmensurables, y una vez calculada Σ se obtiene los valores y vectores propios (ver Anexo D).

$$\Sigma = \begin{bmatrix} 25.839 & 35.663 & 45.661 & 27.669 & 31.202 \\ 35.663 & 53.871 & 65.843 & 40.334 & 44.397 \\ 45.661 & 65.843 & 89.827 & 52.553 & 59.323 \\ 27.669 & 40.334 & 52.553 & 49.549 & 44.991 \\ 31.202 & 44.397 & 59.323 & 44.991 & 51.650 \end{bmatrix}$$

VALORES PROPIOS					
	243.062	17.619	5.178	3.471	1.405

VECTORES PROPIOS					
	C1	C2	C3	C4	C5
PODT	0.312	-0.218	0.071	0.258	0.885
PODR	0.452	-0.338	0.395	0.575	-0.442
PODL	0.592	-0.406	-0.163	-0.669	-0.101
PODC	0.400	0.714	0.529	-0.216	0.056
PODA	0.433	0.405	-0.729	0.329	-0.091

Tabla 3.3: Valores y Vectores Propios Correspondientes a las Variables del Índice de Bienes

En este caso también el primer valor propio es el mayor y con él se cubre el 90% de la variabilidad, por lo tanto las ponderaciones de las variables estarán dadas por los elementos de la primera componente principal.

Finalmente para obtener el indicador de costo (InCosto) se realiza una ponderación de estos dos factores que influyen en este indicador, cabe mencionar que en este caso dado que no existe alguna razón relevante para darle mayor peso a un factor en particular, se restringe a ponderar con el mismo peso a ambos factores, es decir, 0.5 para los dos. Los resultados se muestran en la tabla 3.4.

Municipio	logpob	Inbienes	InCosto
Balancán	4.754	0.274	2.514
Cárdenas	5.395	0.514	2.955
Centla	5.009	0.238	2.624
Centro	5.806	1.000	3.403
Comalcalco	5.285	0.447	2.866
Cunduacán	5.102	0.412	2.757
E. Zapata	4.470	0.672	2.571
Huimanguillo	5.254	0.286	2.770
Jalapa	4.561	0.593	2.577
Jalpa	4.921	0.546	2.734
Jonuta	4.470	0.156	2.313
Macuspana	5.185	0.428	2.806
Nacajuca	5.061	0.742	2.901
Paraíso	4.938	0.702	2.820
Tacotalpa	4.666	0.000	2.333
Teapa	4.729	0.400	2.564
Tenosique	4.771	0.421	2.596

Tabla 3.4: Índice de Costo

3.4. Factor de Peligro

Esta sección es la más extensa del capítulo ya que se muestra la forma en que se obtuvo la información a utilizar para el análisis del factor peligro, así como un breve resumen de la metodología empleada en este análisis. Finalmente se muestran los resultados obtenidos para los 17 municipios que conforman Tabasco.

3.4.1. Obtención de Datos de Precipitación

La información de precipitación se obtuvo de las mediciones de las estaciones en la base de datos que el Servicio Meteorológico Nacional maneja con la herramienta CLICOM, esto para cada uno de los días de los 12 meses, enero-diciembre, desde 1961 hasta el año 2000, en su versión interpolada en malla regular de 0.2 grados de longitud×0.2 grados de latitud (la denominada MAYA v1.0), ver figura 3.2. La información proviene de algo más de 5,000 estaciones climatológicas presentes en la base de datos. La información reciente (2001-2007)

proviene de las mediciones diarias del subconjunto de estaciones climatológicas que reporta en tiempo casi real (diariamente). Por supuesto, estos datos fueron obtenidos en base a su longitud y latitud.

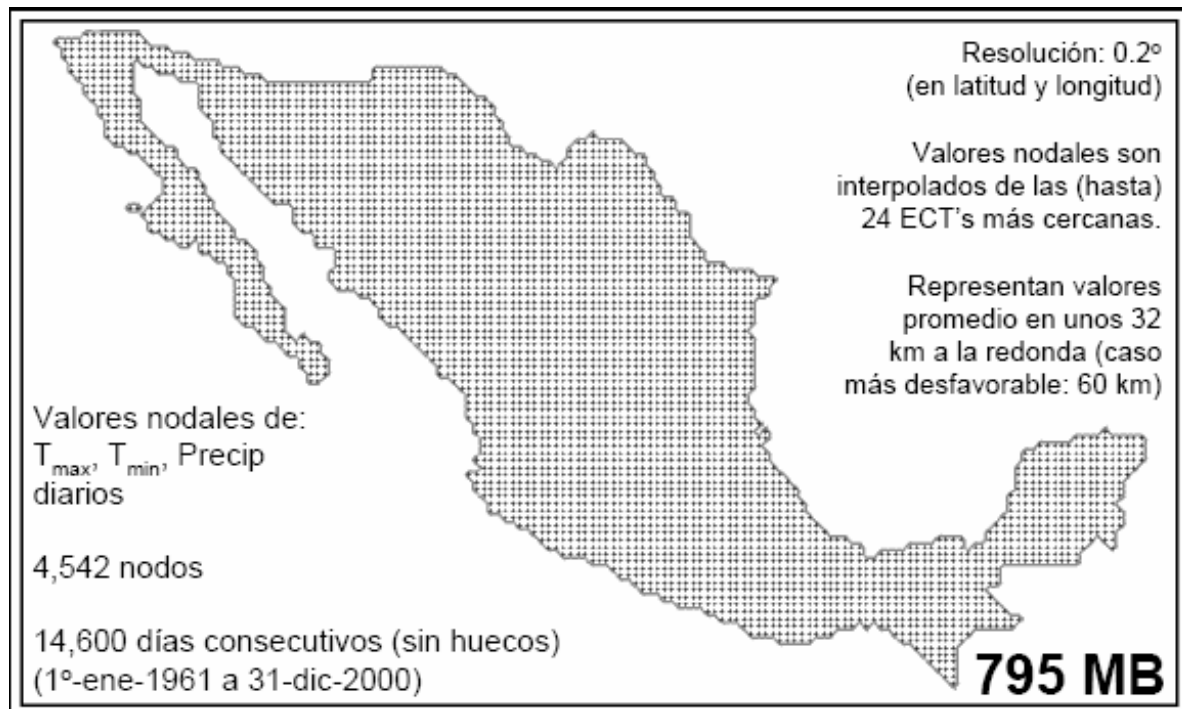


Figura 3.2: Nodos en la República Mexicana

Algunas de las características generales de MAYA v1.0 son:

- 1).- Equivalente a la base de datos climatológica nacional manejada con CLICOM, pero con valores interpolados sobre una malla regular.
- 2).- No tiene “huecos”. Para cada nodo se tiene el 100 % de los 14,600 valores (uno por día) de la variable correspondiente (aunque deberían ser 14,610; los 10 días 29 de febrero en el intervalo no fueron procesados).
- 3).- Utiliza casi la totalidad de los datos originales disponibles (se estima el 94.7% de los datos en papel).
- 4).- Usa la totalidad de las estaciones climatológicas tradicionales que tienen un dato del día en cuestión.
- 5).- Facilitará la identificación de sitios y días sospechosos de contener datos erróneos en la base de datos climatológica nacional.
- 6).- Su dominio es de -117.2 a -86.0 de longitud y 14.0 a 33.0 de latitud.

- 7).- En cuanto a resolución se tiene 0.2° en longitud y latitud, 22.2 km a lo largo de los meridianos y 18.6 a 21.6 km a lo largo de los paralelos.
- 8).- Su tamaño es de 96 renglones por 157 columnas, se tienen 15,072 nodos de los cuales solamente 4,542 caen sobre el territorio nacional, como se muestra en la figura 3.2.

Principales ventajas de MAYA v1.0 sobre CLICOM

- Principalmente evita los problemas con los datos perdidos, los errores de la toma o captura de datos que proporcionan valores demasiado grandes o demasiado pequeños y el problema con las estaciones ya que algunas desaparecieron y algunas otras fueron agregadas.
- Facilita el manejo de la base de datos, lo más complicado en este caso es obtener los nodos a utilizar.
- Si se quisiera trabajar con herramientas de geoestadística esta base de datos es mucho más apropiada.

NOTA 3.1. Se tiene que la base de datos MAYA v1.0 cuenta con tres variables: Precipitación pluvial, temperaturas máximas y temperaturas mínima, éstas vienen en dos presentaciones como campos por día (mallas) y listas nodales (nodo por nodo).

Los datos utilizados son los de precipitación pluvial y en la presentación de nodo por nodo. Lo que se tiene es una lista de 4,542 archivos de texto en donde cada uno de éstos corresponde a un nodo y cada archivo contiene una serie de datos de precipitación pluvial desde enero de 1961 hasta diciembre del 2000. Además de esta información, se conoce la precipitación del 2001-2007 de las estaciones existentes en la república mexicana. Se cuenta con un archivo por año, cada uno con los siguientes campos: latitud, longitud, clave y nombre de la estación, altura, mes y días. Para aprovechar la información extra se trabajó con ella como se explica a continuación:

- 1).- Para cada nodo que se encuentra involucrado en el análisis se calcula un valor máximo de precipitación de los años 2001-2007.
- 2).- Se obtiene la precipitación que corresponde al nodo de interés con filtros de Excel; los filtros se hacen con respecto a la ubicación de las estaciones meteorológicas (latitud y longitud), aprovechando que se conoce la ubicación de los nodos, se filtra la información de las estaciones que se encuentran a $\pm 0.2^\circ$ de la ubicación exacta del nodo.

3).- Una vez que se tienen las estaciones que rodean al nodo, se elige al valor máximo de las precipitaciones de todas las estaciones involucradas.

Ahora, para dar respuesta a ¿Cuáles son los nodos de interés?, es necesario determinar los nodos que intervienen en cada municipio. Para lo cual es importante mencionar las cuencas que conforman al estado de Tabasco. La mayor parte de la superficie del estado (75.22 %) se localiza en la Región Hidrográfica número 30, o Región del sistema Grijalva-Usumacinta, formada por las cuencas hidrográficas del Grijalva, Usumacinta y de la Laguna de Términos; que en Tabasco ocupan el 41.45 %, el 29.24 % y el 4.53 % respectivamente. Mientras que el 24.78 % restante, se encuentra dentro de la Región Hidrográfica número 29 o Región del Coatzacoalcos, formada por dos cuencas: la del Coatzacoalcos, Tonalá y lagunas del Carmen y la Machona; siendo ésta última la única que tiene representación en el estado. Ambas regiones están consideradas como las más húmedas del país, en primer y segundo lugares, respectivamente.

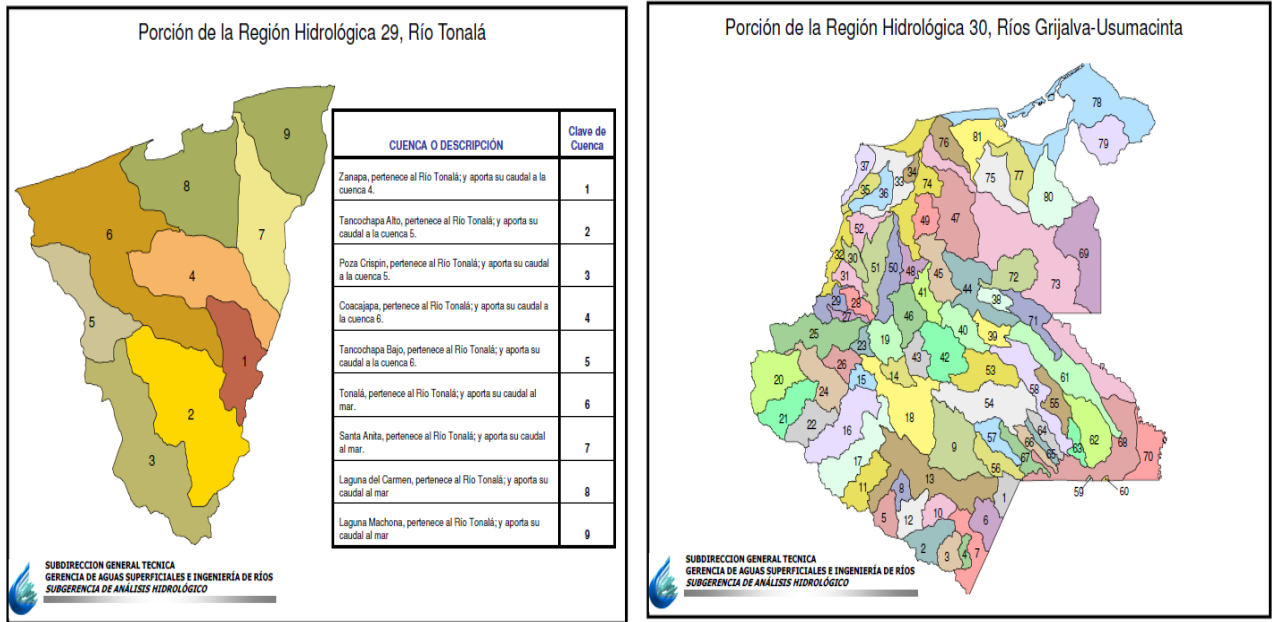


Figura 3.3: Región Hidrográfica 29 y 30

NOTA 3.2. Es importante mencionar la definición de cuenca.
Cuenca: la cuenca es una zona de la superficie terrestre donde, si fuera impermeable, las gotas de lluvia que caen sobre ella tenderían a ser drenadas por el sistema de corrientes hacia un mismo punto de salida. En realidad, el terreno no es impermeable, por lo que un porcentaje del volumen de lluvias es absorbido por el suelo, otro es atrapado en pequeñas depresiones del terreno, formando charcos, otro queda sobre la vegetación y otra parte escurre hacia los ríos y arroyos.

También se tiene que las inundaciones a consecuencia de la precipitación se presentan cuando el terreno se ha saturado y el agua de lluvia excedente comienza a acumularse, pudiendo permanecer horas o días. A estos fenómenos se suman el efecto ejercido por las cadenas montañosas (lluvia orográfica), que ocasionan tormentas de corta duración y poca extensión, pero muy intensas. Igual o más importante aún es considerar la acción conjunta de estos mecanismos de lluvia, por ejemplo, en octubre de 1999 como resultado de la interacción de la depresión tropical no. 11 y el frente frío no. 5 ocurrió una tormenta severa en el norte de Veracruz, afectando también los estados de Hidalgo y Puebla; causó inundaciones en la planicie costera del golfo de México. Otro más está dado por alguna de las inundaciones que se presentaron en septiembre 2007 en Huimanguillo que no fueron ocasionadas por la lluvia de este lugar sino por una lluvia intensa ocurrida en Coatzacoalcos, Veracruz que fue arrastrada por los ríos que comunican a dichas comunidades.

Por las razones expuestas, para obtener la base de datos de cada municipio, se habla de municipio ya que la información de vulnerabilidad y costo se tiene a este nivel, se delimitó una región para cada uno. La forma en que se obtiene esta información en la presente investigación es un tanto burda ya que no se contó con el tiempo ni con los materiales necesarios para hacerlo de la manera más apropiada, lo que probablemente se requería hacer era un análisis de isócronas y tomar subcuencas sólo hasta la isocrona correspondiente a 24 horas ya que lo que interesa es la aportación de las cuencas aguas arriba en las primeras 24 horas.

Entonces, para obtener los datos de precipitación en cada municipio se utilizó la información de las subcuencas que forman parte de cada uno de estos, ya sea que se encuentra completamente o solo parte de ella, y se toma en cuenta aquellas que la aportación de su caudal influye en el municipio de estudio. En la tabla 3.5 se ejemplifica al municipio del Centro, se observa que en este municipio convergen cuatro subcuencas y se detalla hacia donde aportan su caudal. Con esto se determina cuáles son los nodos que influyen en el municipio, la subcuenca (34) aporta su caudal a la cuenca (74) la cual tiene la mayor presencia en el Centro, por lo que se toman en cuenta tanto los nodos de la (34) que se encuentran dentro del municipio, como los que están fuera, pero cercanos a éste. En caso de que parte de alguna subcuenca se encuentre dentro del municipio, pero la aportación de su caudal no influya en éste, se toma en cuenta únicamente los nodos que se encuentren dentro del municipio, si es que éstos existen.

Centro		
Subcuenca	Aporta su caudal	N° Nodos
Tabasquillo (34)	A la cuenca 74	1
Chilapilla (49)	A la cuenca 74	1
De la sierra (50)	A la cuenca 74	1
Grijalva (74)	Al mar	3

Tabla 3.5: Subcuencas del municipio de Centro

Las tablas para la descripción de cada municipio se encuentran en el anexo F. Para delimitar la región y obtener la longitud y latitud de los nodos se utilizó el programa computacional ARCVIEW 3.2. En la figura 3.4 se muestra la región final con la que se trabajó.

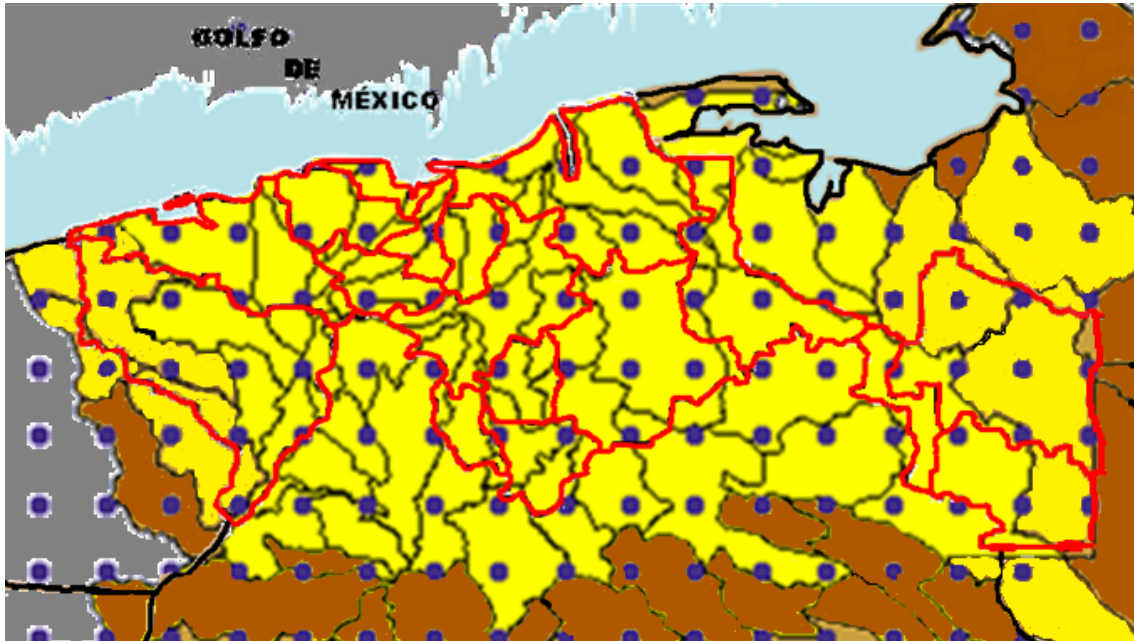


Figura 3.4: Región de Interés, Tabasco

3.4.2. Análisis de los Datos

En el análisis de los datos se requiere de una serie de observaciones para cada municipio, pero en este caso se tienen tantas series como el número de nodos que se involucren. Entonces, para obtener una sola serie se realiza una combinación lineal, que resulte adecuada, de los nodos que entran en el municipio. Para lograr esta combinación como se explicó en el capítulo 2 se utiliza la teoría de componentes principales, en este caso las variables que se consideran en el análisis son los nodos de cada municipio, los CP que se obtienen corresponden a los pesos para cada nodo con los que se obtendrá la combinación lineal deseada y la serie final de los datos de precipitación del municipio de interés.

Una vez que se cuenta con las series para cada municipio de Tabasco, se extraen las precipitaciones máximas por bloques, los bloques serán años, por lo tanto se tienen máximos anuales. Esto es, sea x_1, x_2, \dots, x_m la muestra de la variable aleatoria X : precipitaciones en el municipio l , entonces se forman los vectores:

$$\begin{aligned} \mathbf{x}^{(1)} &= (x_1^1, \dots, x_n^1) \\ \mathbf{x}^{(2)} &= (x_1^2, \dots, x_n^2) \\ &\vdots \\ \mathbf{x}^{(k)} &= (x_1^k, \dots, x_n^k) \end{aligned}$$

Una vez que se obtienen los máximos anuales, $M_{n,i} = \text{máx} \{x_1^i, \dots, x_n^i\}$, $i = 1, \dots, k$. Por el teorema de Fisher y Tippet [14] y su generalización por Gnedenko [17], se asume que tienen una *DVE* (Gumbel, Weibull y Fréchet) ó *DVEG* y se estiman los parámetros μ , σ y ξ .

En esta investigación los datos se ajustaron a dos distribuciones, Gumbel y VEG, esto con la finalidad de hacer una comparación entre la distribución Gumbel que es una de las más utilizadas en el campo de las inundaciones en México y la DVEG es una generalización de las DVE y su uso se ha ampliado en los últimos años, pero aún no muy aplicada para las precipitaciones máximas en México.

La estimación se realizó por el Método de Máxima Verosimilitud, para aprovechar que con estos estimadores se pueden realizar aproximaciones asintóticas para obtener los intervalos de confianza. El logaritmo de la función de verosimilitud para la DVEG y la Gumbel está dado por la siguiente expresión:

$$l(\mu, \sigma, \xi; \mathbf{x}) = \begin{cases} -n \log \sigma - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^n \log \left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right] - \sum_{i=1}^n \left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right]^{\frac{1}{\xi}} & \text{si } \xi \neq 0 \\ -n \log \sigma - \sum_{i=1}^n \exp \left[\frac{-x_i - \mu}{\sigma}\right] - \sum_{i=1}^n \left[\frac{-x_i - \mu}{\sigma}\right] & \text{si } \xi = 0 \end{cases}$$

En el método de máxima verosimilitud se establece que deben elegirse los estimadores $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ que maximicen l , para lograr el objetivo se realizó el cálculo de los estimadores utilizando las librerías VGAM, evir, evd del Proyecto R. Una vez que se ha obtenido la estimación puntual de los parámetros, también se calcula la estimación por Intervalos.

Por el teorema 2.1 se tiene que, para la estimación de los intervalos de confianza de los parámetros de la *DVEG*, para un tamaño de muestra grande el estimador de máxima verosimilitud se distribuye asintóticamente normal, esto es

$$\begin{aligned} \hat{\boldsymbol{\theta}} &\sim N_d(\boldsymbol{\theta}, \mathbf{I}^{-1}(\hat{\boldsymbol{\theta}})) \\ \hat{\boldsymbol{\theta}} &\sim N_d(\boldsymbol{\theta}, \mathbf{J}^{-1}(\hat{\boldsymbol{\theta}})) \end{aligned}$$

donde $\mathbf{I}(\boldsymbol{\theta})$ es la matriz de información de Fisher y $\mathbf{J}(\boldsymbol{\theta})$ es el valor esperado de la matriz de información de Fisher.

Entonces, el intervalo de confianza al $(1 - \alpha)100\%$ para $\boldsymbol{\theta}$ está dado por

$$\hat{\theta}_i \pm Z_{\alpha/2} \sqrt{c_{ii}}$$

en donde, $Z_{\alpha/2}$ es el valor de la variable normal estándar, y c_{ii} es el i -ésimo elemento de la diagonal de $\mathbf{J}^{-1}(\hat{\boldsymbol{\theta}})$ o $\mathbf{I}^{-1}(\hat{\boldsymbol{\theta}})$.

Elección del Modelo

La elección del mejor modelo se basó principalmente en el Criterio de Información de Akaike (AIC), éste se obtiene a partir de la siguiente expresión:

$$AIC(k) = -2 \log(L(\theta_k)) + 2k$$

donde, k es el número de parámetros del modelo y $\log(L(\theta_k))$ es el logaritmo de la verosimilitud del modelo. Se prefiere aquel modelo que cuente con el menor valor de AIC.

También se realiza una prueba más a los datos, en donde lo que interesa es contrastar el siguiente juego de hipótesis:

$$H_0 : f(x) = f_0(x; \mu, \sigma, \xi)$$

$$H_1 : f(x) \neq f_0(x; \mu, \sigma, \xi)$$

donde $f_0(x; \mu, \sigma, \xi)$ se refiere a la distribución Gumbel o la de Valores Extremos Generalizada. Esta prueba se basa en asumir que existe una relación lineal entre la distribución empírica y la inversa de la distribución teórica correspondiente, por lo que si se emplea el coeficiente de correlación muestral como medida de asociación.

$$r_n(\mathbf{Y}, \mathbf{Z}) = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(Z_i - \bar{Z})}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2} \sqrt{\sum_{i=1}^n (Z_i - \bar{Z})^2}}$$

Por lo tanto, si se cumple H_0 se esperaría obtener un coeficiente de correlación cercano a uno. Finalmente se prefiere aquel modelo cuyo valor del coeficiente sea más cercano a uno.

Periodos de Retorno

Para el análisis del factor peligro es de interés conocer el tiempo promedio de espera para que ocurra una lluvia extrema que pudiera causar algún problema de inundación. En la definición 2.3, se menciona que el período de retorno es el tiempo medio para que se repita un evento semejante. Entonces sea X_1, X_2, \dots una sucesión de variables aleatorias *iid* con función de distribución continua F y $u \in \mathbb{R}$. Considerando $\{X_i > u\}$ un evento extremo, se define el tiempo de la primera excedencia como

$$T(u) = \min \{i \geq 1 : X_i > u\}$$

entonces, el tiempo promedio de espera para que ocurra un evento extremo es $E[T(u)]$.

Si la probabilidad del evento $\{X_i > u\}$ es p , su período de retorno es p^{-1} . Luego, usando

la distribución de probabilidad acumulada, para el evento $\{X_i > u\}$, el periodo de retorno es

$$\tau_x = \frac{1}{1 - F(x)}$$

Por lo anterior resulta de interés estudiar los cuantiles extremos superiores de la distribución Gumbel y la DVEG. Específicamente, el nivel de retorno asociado con el periodo de retorno de $1/p$, es el $(1 - p)$ -ésimo cuantil de la distribución VEG. Se puede obtener una expresión para los cuantiles invirtiendo la función de distribución, es decir, $G^{-1}(1 - p; \mu, \sigma, \xi)$, se tiene entonces, $x_p = G^{-1}(1 - p)$.

Por lo tanto, para a la distribución Gumbel y de VEG se tiene que el cuantil $1 - p$ está dado por:

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

Los cuantiles se calculan a partir de las estimaciones de los parámetros, que se obtuvieron como se mencionó en la sección anterior.

3.4.3. Resultados del Factor Peligro

Como se mencionó para el cálculo del factor peligro se utilizó una serie de datos históricos de las precipitaciones registradas de 1961 al 2007 para cada uno de los 17 municipios de Tabasco. A estos datos se les extrajo los máximos anuales mismos que sirvieron para el análisis del componente peligro. A estos datos se les ajustaron dos distribuciones de valores extremos, la distribución Gumbel y la distribución de valores extremos generalizada (DVEG), para ambas distribuciones se realizó una estimación puntual y otra por intervalos. Para elegir cuál de los dos modelos era el mejor se utilizaron tres criterios, el logaritmo de la verosimilitud (L-veros), el criterio de Akaike (AIC) y la correlación entre la distribución empírica y el modelo propuesto (Corr). Finalmente se eligió aquel modelo en el cual dos o más criterios le favorecieron. El ajuste se hizo para los 17 municipios de interés.

De la tabla 3.6 al 3.22 se muestran los resultados de los ajustes hechos a cada uno de los municipios, así como en las figuras de las 3.5 a la 3.21 se muestran de manera gráfica la dispersión de los datos, con la que es posible observar la independencia de los datos, el ajuste de las dos distribuciones que se estudian y dos gráficas más que muestran la correlación entre la distribución empírica y el modelo dado por la distribución Gumbel y por la DVEG.

Empezando con el municipio de Balancán (tabla 3.6 y figura 3.5) nótese que las diferencias de estimación entre los dos modelos son muy pocas, sin embargo se observa que la L-veros y la Corr son mejores en el modelo dado por la DVEG y por tanto es con éste con el cual se calculan los periodos de retorno.

Estimación de Parámetros, Balancán						
Dist.	Escala (σ)	Forma (ξ)	Loc (μ)	L-veros	AIC	Corr
VEG	27.840	0.129	117.294	-234.053	474.106	0.974
Gumbel	29.368	.	119.298	-234.831	473.662	0.972

IC	VEG	Gumbel
σ	(21.002, 34.678)	(27.172, 31.564)
ξ	(-0.092, 0.351)	.
μ	(108.315, 126.274)	(110.456, 128.139)

Tabla 3.6: Estimación del municipio de Balancán

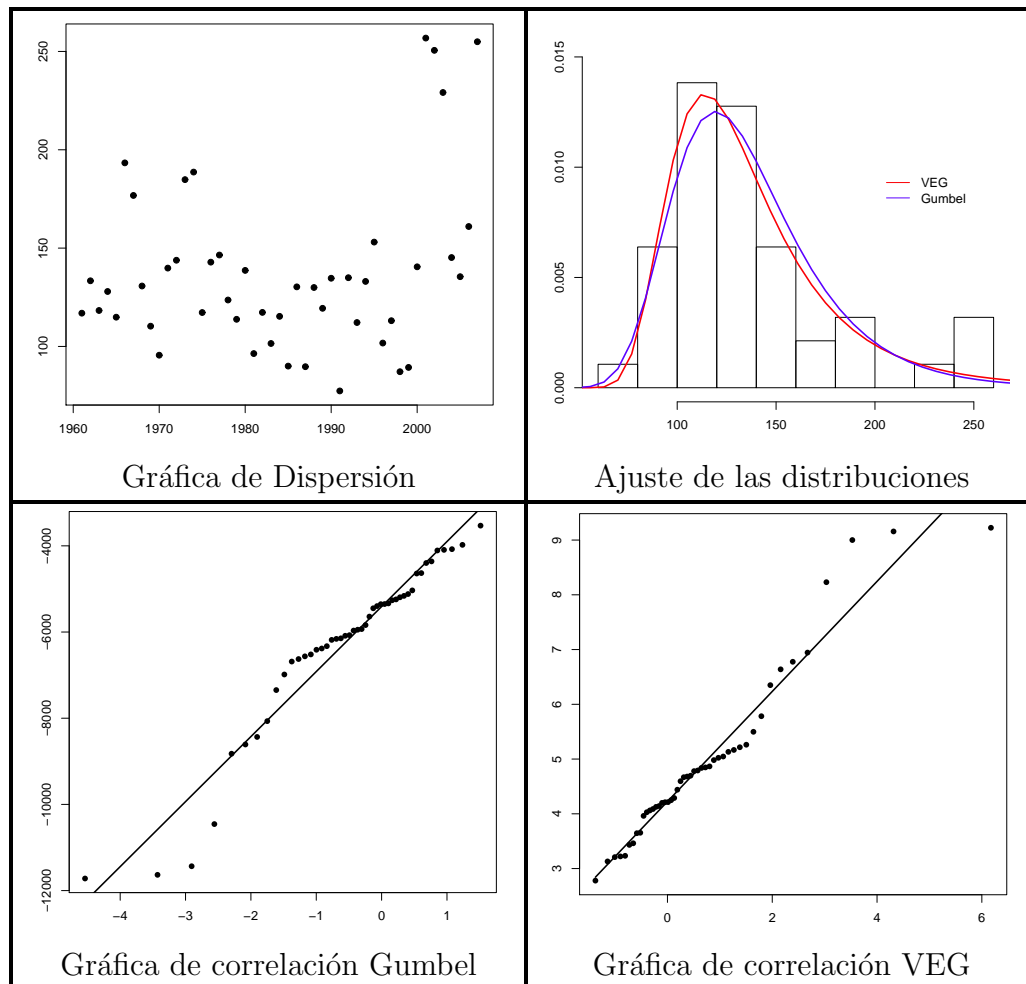


Figura 3.5: Gráficas del municipio de Balancán

La tabla 3.7 muestra la estimación del municipio de Cárdenas y la figura 3.6 muestra sus gráficas respectivas. En el análisis de este municipio se observa que el AIC y la Corr son

mejores en la distribución Gumbel y por tanto es el modelo elegido.

Estimación de Parámetros, Cárdenas						
Dist.	Escala (σ)	Forma (ξ)	Loc (μ)	L-veros	AIC	Corr
VEG	30.113	0.075	116.396	-236.240	478.480	0.987
Gumbel	30.990	.	117.649	-236.491	476.982	0.989

IC	VEG	Gumbel
σ	(22.851, 37.375)	(28.793, 33.186)
ξ	(-0.145, 0.295)	.
μ	(106.664, 126.128)	(108.319, 126.979)

Tabla 3.7: Estimación del municipio de Cárdenas

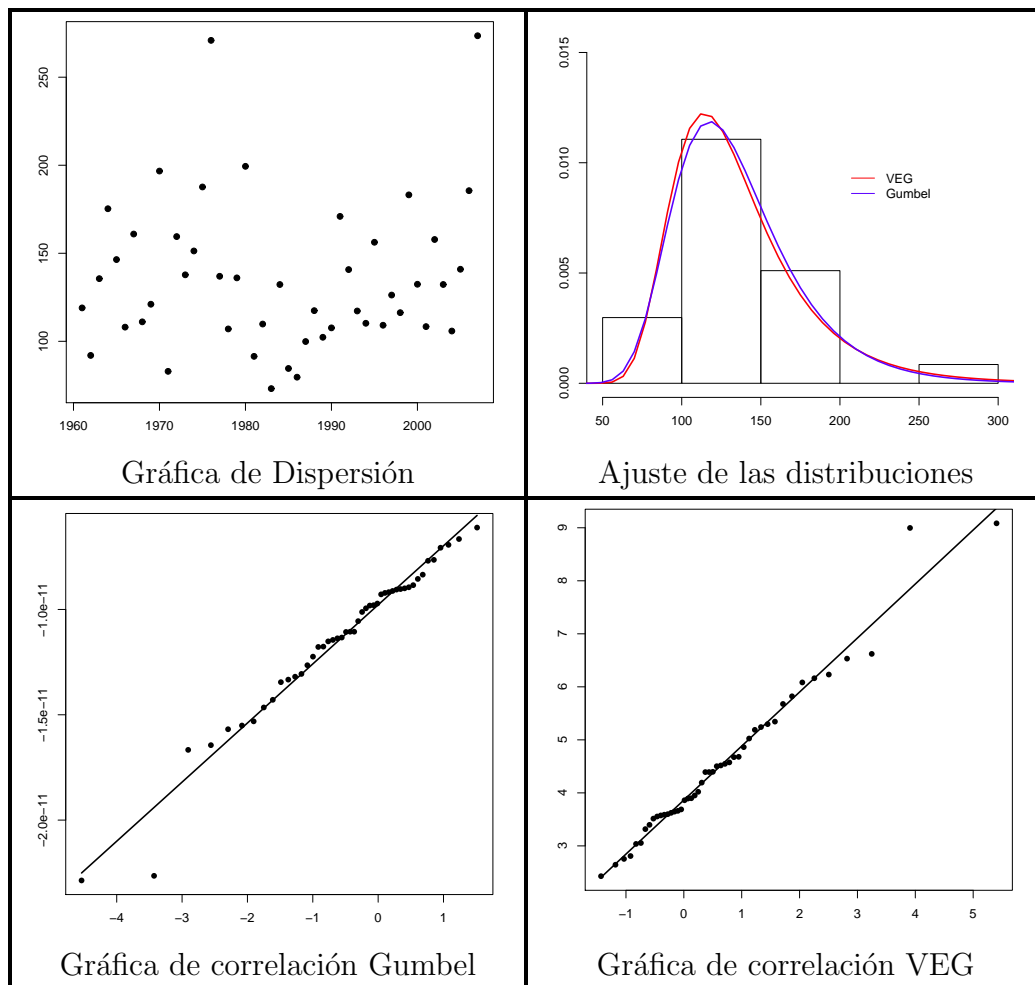


Figura 3.6: Gráficas del municipio de Cárdenas

En el municipio de Centla hay dos criterios que son mejores, la L-veros y la Corr, en el

modelo de la DVEG. En la tabla 3.8 y en la figura 3.7 se da la estimación y las gráficas. Por lo tanto para este municipio el mejor modelo es el de DVEG.

Estimación de Parámetros, Centla						
Dist.	Escala (σ)	Forma (ξ)	Loc (μ)	L-veros	AIC	Corr
VEG	30.429	0.129	125.749	-238.245	482.491	0.983
Gumbel	32.087	.	127.929	-239.042	482.084	0.980

IC	VEG	Gumbel
σ	(22.994, 37.863)	(29.891, 34.283)
ξ	(-0.087, 0.346)	.
μ	(115.975, 135.524)	(118.269, 137.589)

Tabla 3.8: Estimación del municipio de Centla

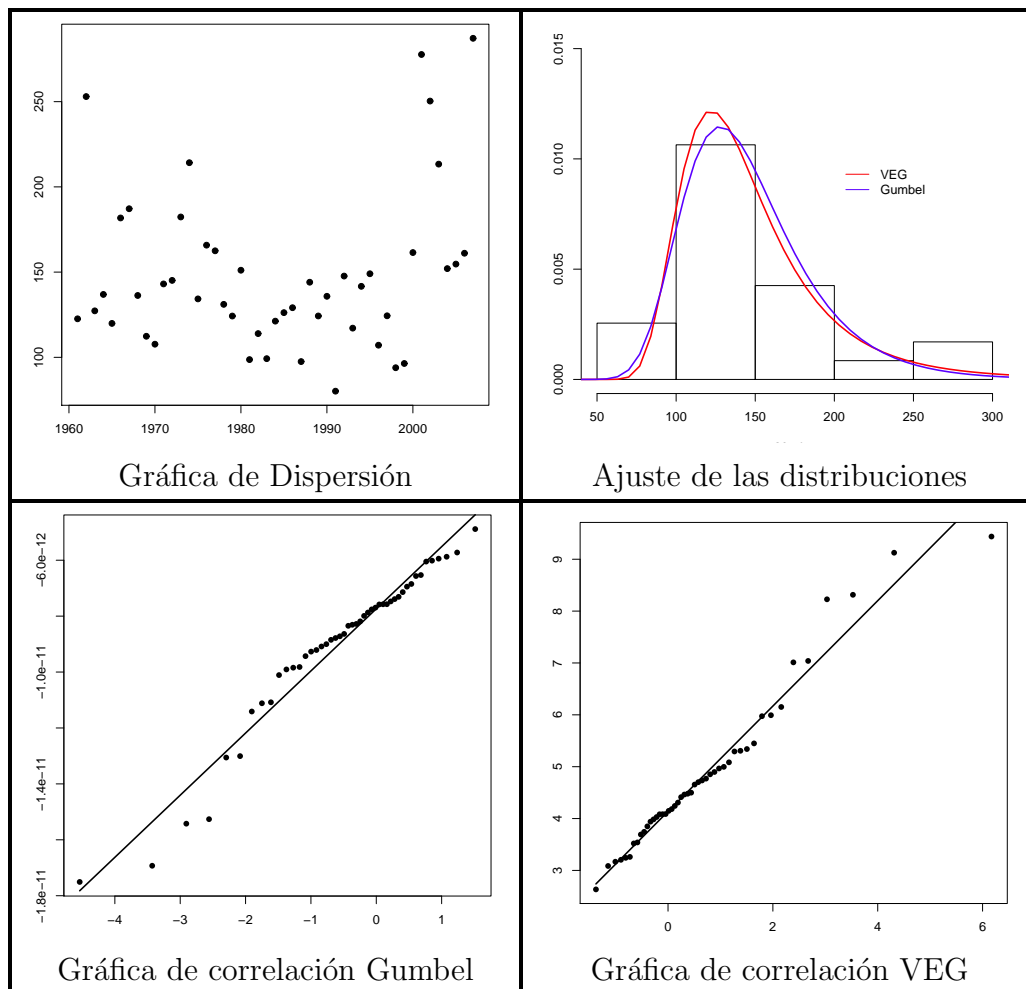


Figura 3.7: Gráficas del municipio de Centla

Ahora, en la tabla 3.9 y en la figura 3.8 se tienen la estimación del municipio Centro, éste es uno de los más importantes ya que en él se encuentra la capital del estado Villahermosa, y la cual ha sufrido severos problemas de inundación además de ser el municipio con mayor número de habitantes. El mejor modelo para este municipio es el de la distribución Gumbel ya que el AIC y la Corr resultan ser favorables para este ajuste.

Estimación de Parámetros, Centro						
Dist.	Escala (σ)	Forma (ξ)	Loc (μ)	L-veros	AIC	Corr
VEG	29.428	0.074	140.899	-235.300	476.601	0.989
Gumbel	30.375	.	142.057	-235.433	474.866	0.993

IC	VEG	Gumbel
σ	(21.711, 37.144)	(28.179, 32.572)
μ	(-0.219, 0.367)	.
ξ	(130.895, 150.903)	(132.908, 151.205)

Tabla 3.9: Estimación del municipio de Centro

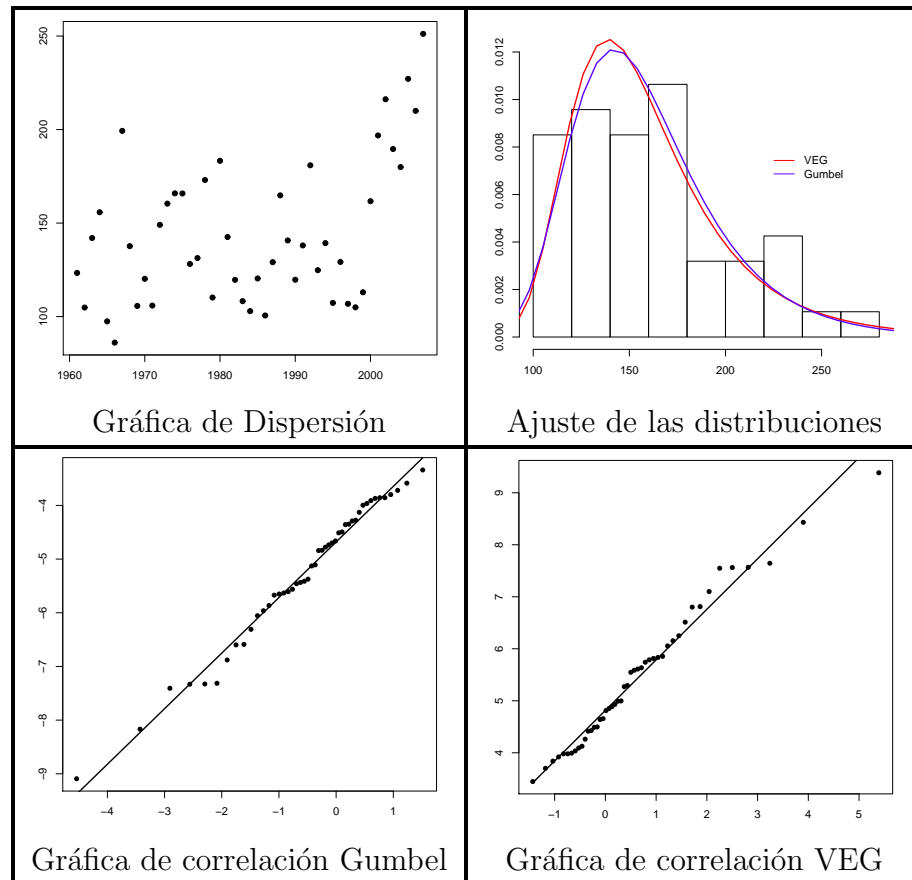


Figura 3.8: Gráficas del municipio de Centro

Para el municipio de Comalcalco se tiene la tabla 3.10 y la figura 3.9 que muestran los resultados de la estimación de dicho municipio, con los que se elige la DVEG como mejor modelo ya que tanto la Corr como la L-veros son mejores en este modelo.

Estimación de Parámetros, Comalcalco						
Dist.	Escala (σ)	Forma (ξ)	Loc (μ)	L-veros	AIC	Corr
VEG	32.020	0.115	105.520	-240.263	486.526	0.993
Gumbel	33.631	.	107.576	-240.783	485.566	0.990

IC	VEG	Gumbel
σ	(24.02, 40.019)	(31.435, 35.827)
ξ	(-0.124, 0.355)	.
μ	(95.059, 115.981)	(97.451, 117.701)

Tabla 3.10: Estimación del municipio de Comalcalco

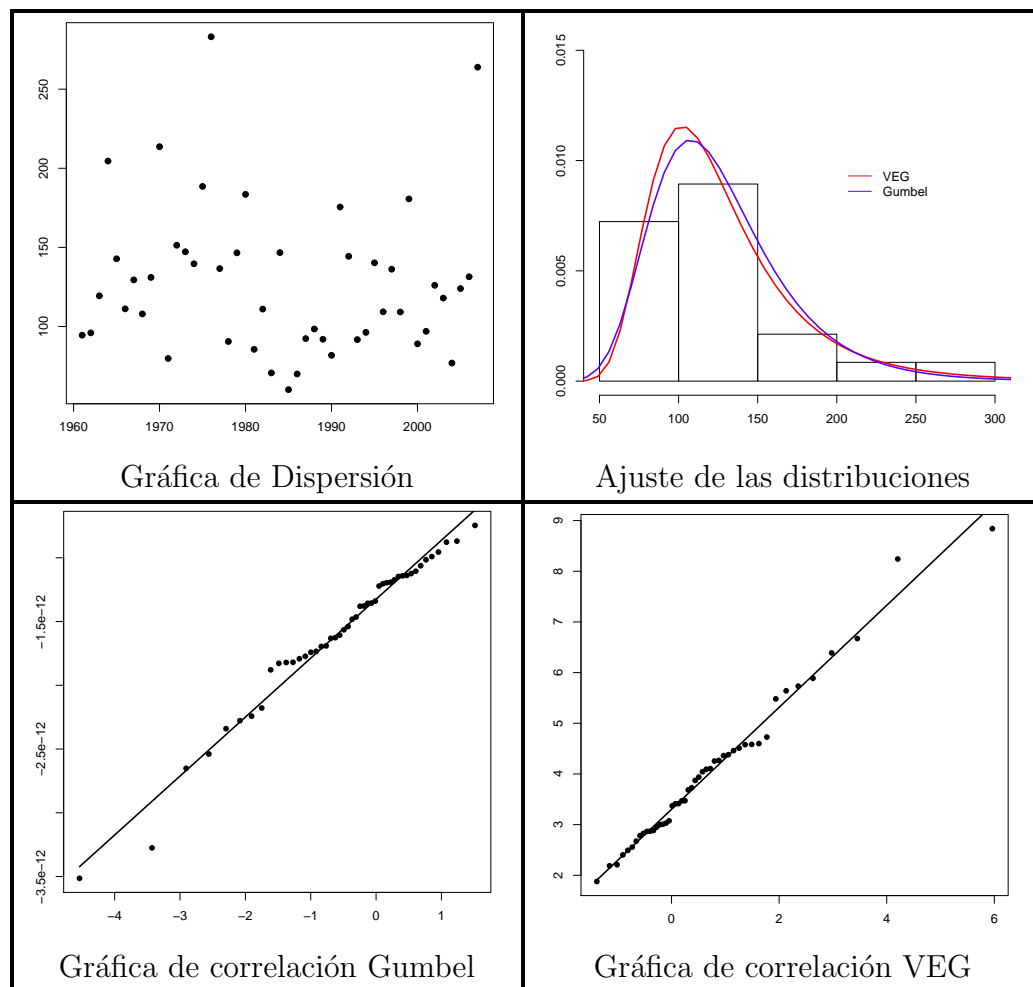


Figura 3.9: Gráficas del municipio de Comalcalco

Para el municipio de Cunduacán se elige como mejor modelo el Gumbel ya que según la tabla 3.11 este modelo tiene dos criterios a su favor, la Corr y el AIC. Además observando las gráficas de la figura 3.10 se reafirma que éste es el mejor modelo.

Estimación de Parámetros, Cunduacán						
Dist.	Escala (σ)	Forma (ξ)	Loc (μ)	L-veros	AIC	Corr
VEG	33.485	0.177	114.592	-243.959	493.919	0.972
Gumbel	36.307	.	117.920	-244.885	493.771	0.985

IC	VEG	Gumbel
σ	(24.637, 42.333)	(34.111, 38.503)
ξ	(-0.093, 0.447)	.
μ	(103.509, 125.674)	(106.986, 128.853)

Tabla 3.11: Estimación del municipio de Cunduacán

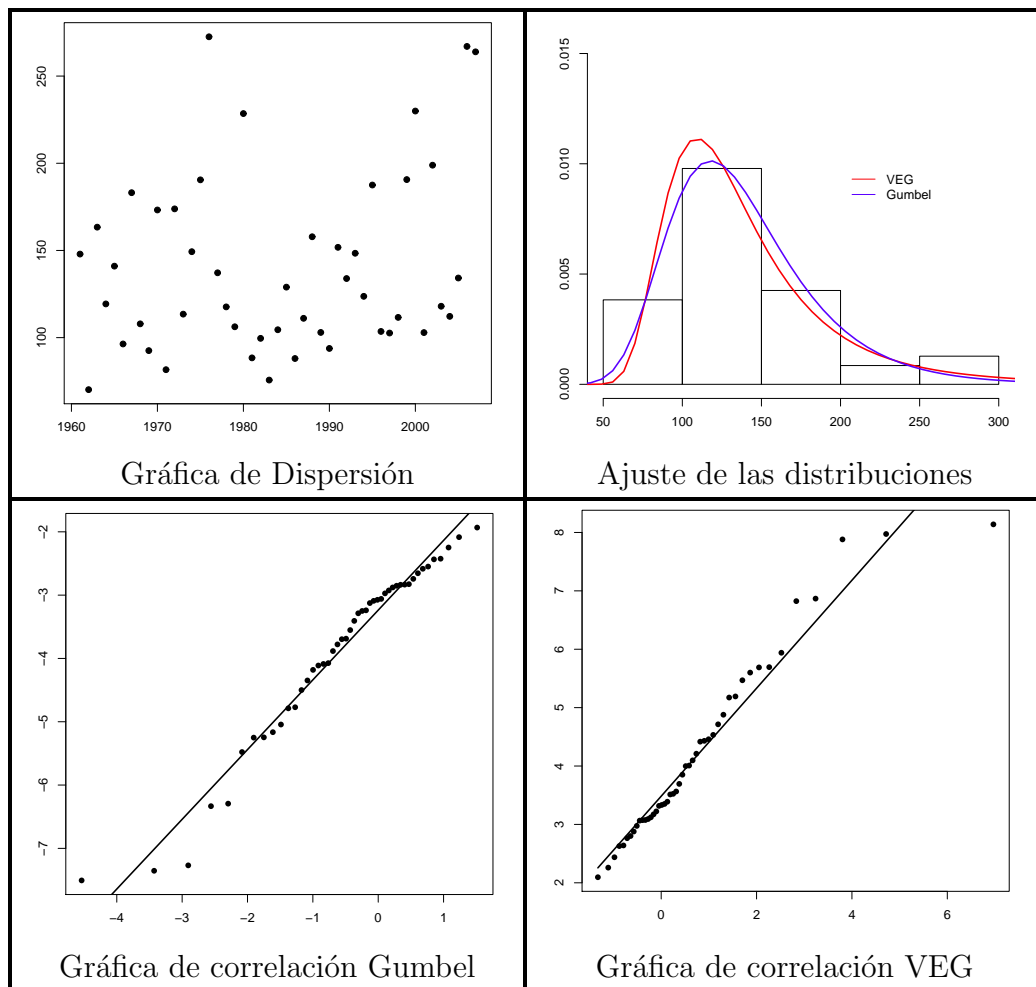


Figura 3.10: Gráficas del municipio de Cunduacan

En cuanto al municipio de Emiliano Zapata el modelo a elegir es el de la DVEG, ya que como se observa en la tabla 3.12 este modelo tiene dos criterios a su favor, la L-veros y la Corr, del mismo modo al observar el ajuste de la DVEG en la figura 3.11 se nota que es mejor.

Estimación de Parámetros, Emiliano Zapata						
Dist.	Escala (σ)	Forma (ξ)	Loc (μ)	L-veros	AIC	Corr
VEG	32.384	0.121	136.028	-240.865	487.730	0.985
Gumbel	34.002	.	138.187	-241.583	487.166	0.978

IC	VEG	Gumbel
σ	(24.493, 40.275)	(31.806, 36.198)
ξ	(-0.095, 0.336)	.
μ	(125.613, 146.443)	(127.95, 148.423)

Tabla 3.12: Estimación del municipio de Emiliano Zapata

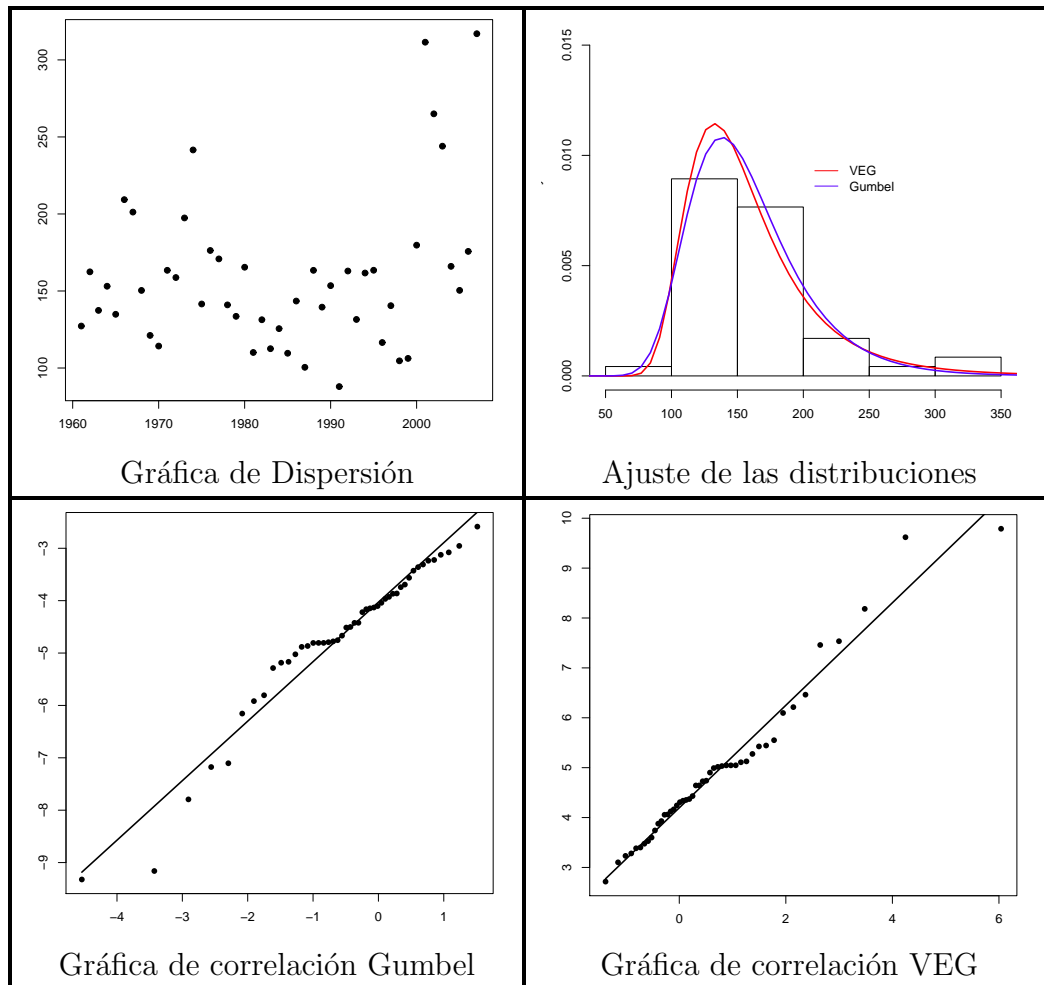


Figura 3.11: Gráficas del municipio de Emiliano Zapata

Otro municipio importante es Huimanguillo quien ha presentado problemas fuertes de inundación, además de ser el municipio más marginado de Tabasco, las pérdidas económicas son menores, pero los daños ocasionados a la población son más trascendentes. En la tabla 3.13 y la figura 3.12 se tiene la estimación y las gráficas del municipio, el modelo es Gumbel.

Estimación de Parámetros, Huimanguillo						
Dist.	Escala (σ)	Forma (ξ)	Loc (μ)	L-veros	AIC	Corr
VEG	50.647	0.125	161.799	-262.104	530.209	0.982
Gumbel	53.403	.	165.261	-262.577	529.155	0.989

IC	VEG	Gumbel
σ	(37.606, 63.689)	(51.207, 55.599)
ξ	(-0.139, 0.39)	.
μ	(145.042, 178.555)	(149.179, 181.344)

Tabla 3.13: Estimación del municipio de Huimanguillo

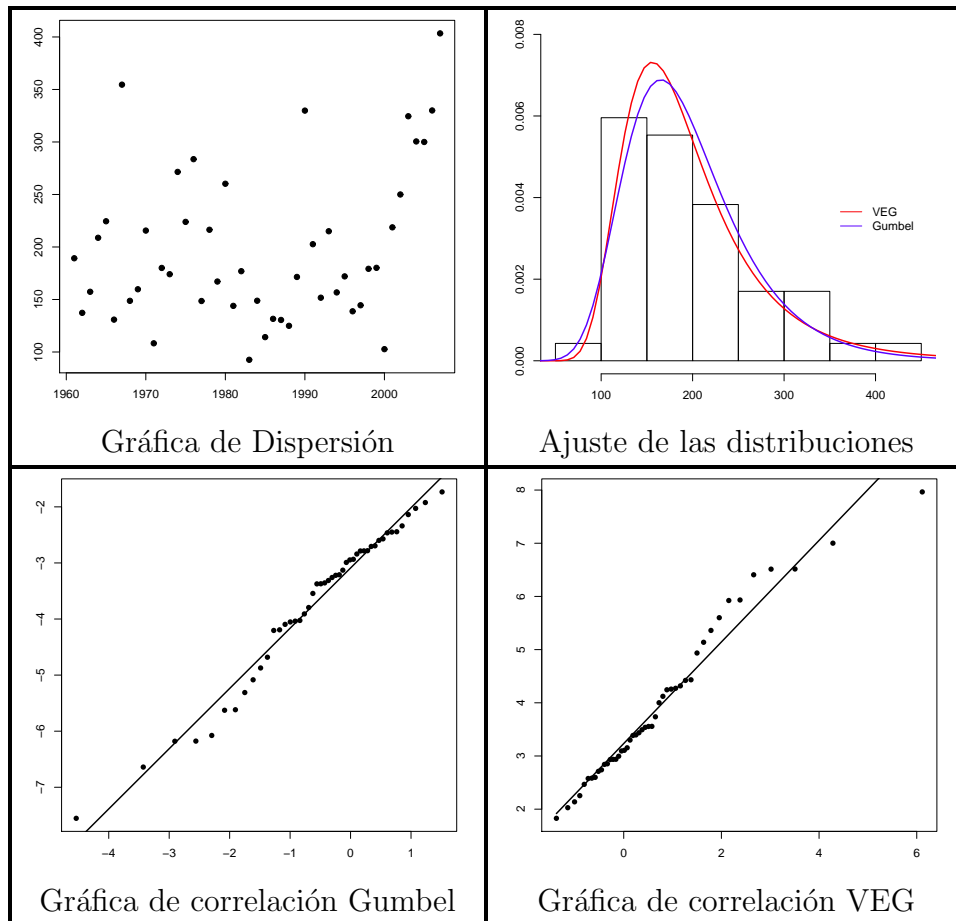


Figura 3.12: Gráficas del municipio de Huimanguillo

En la tabla 3.14 se muestra la estimación del municipio de Jalapa, en este caso el criterio de Corr es igual en ambos modelos, por lo que la elección del mejor modelo se basa en el AIC ya que éste toma en cuenta el número de parámetros y discrimina a un modelo más sencillo. Entonces el modelo Gumbel es elegido para Jalapa. Sus gráficas se muestran en la figura 3.13.

Estimación de Parámetros, Jalapa						
Dist.	Escala (σ)	Forma (ξ)	Loc (μ)	L-veros	AIC	Corr
VEG	30.926	0.082	143.438	-237.786	481.572	0.996
Gumbel	31.992	.	144.847	-238.018	480.036	0.996

IC	VEG	Gumbel
σ	(23.205, 38.647)	(29.796, 34.188)
ξ	(-0.167, 0.332)	.
μ	(133.252, 153.624)	(135.214, 154.479)

Tabla 3.14: Estimación del municipio de Jalapa

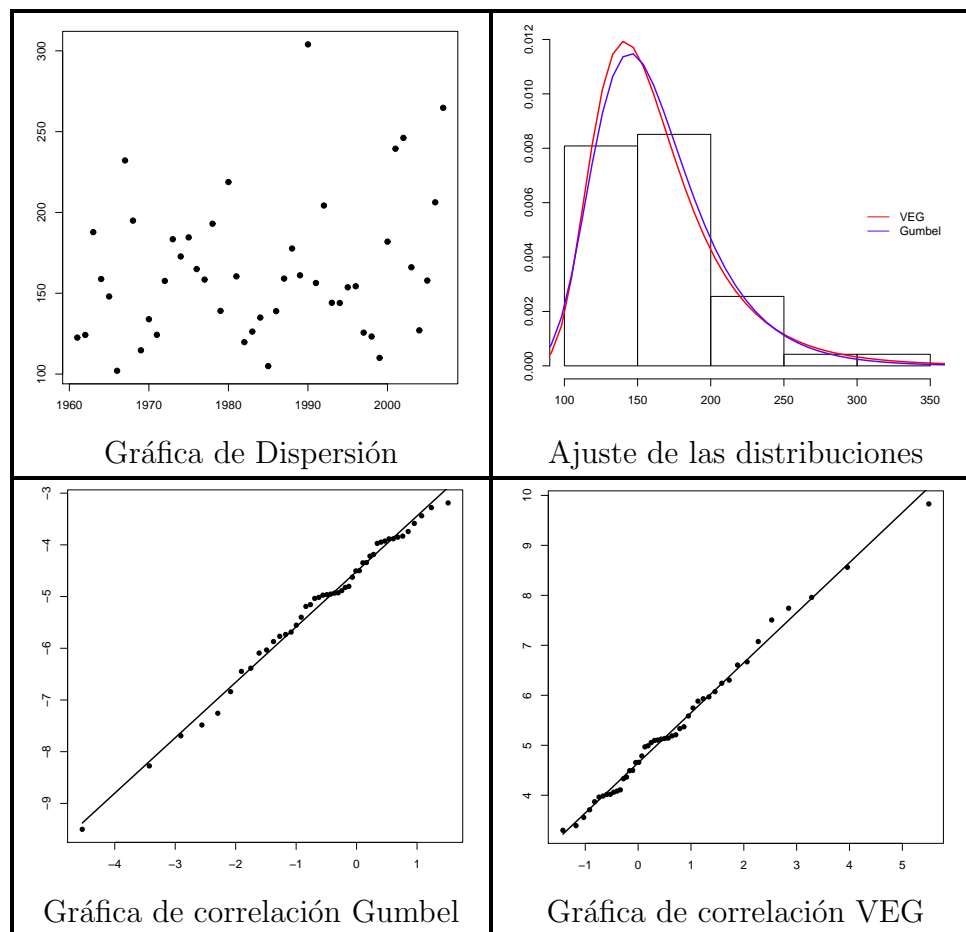


Figura 3.13: Gráficas del municipio de Jalapa

El modelo elegido para Jalpa de Méndez es el Gumbel puesto que tiene dos criterios con valores mejores que el modelo VEG, ver la tabla 3.15 y el AIC es notablemente mejor; aunque la diferencia de la Corr entre los dos modelos es muy pequeña, el modelo Gumbel tiene el mejor valor. Las gráficas correspondientes a esta estimación se muestran en la figura 3.14.

Estimación de Parámetros, Jalpa de Méndez						
Dist.	Escala (σ)	Forma (ξ)	Loc (μ)	L-veros	AIC	Corr
VEG	34.660	0.050	128.697	-242.323	490.646	0.982
Gumbel	35.255	.	129.636	-242.427	488.853	0.984

IC	VEG	Gumbel
σ	(26.434, 42.887)	(33.059, 37.451)
ξ	(-0.168, 0.267)	.
μ	(117.549, 139.846)	(119.021, 140.251)

Tabla 3.15: Estimación del municipio de Jalpa de Méndez

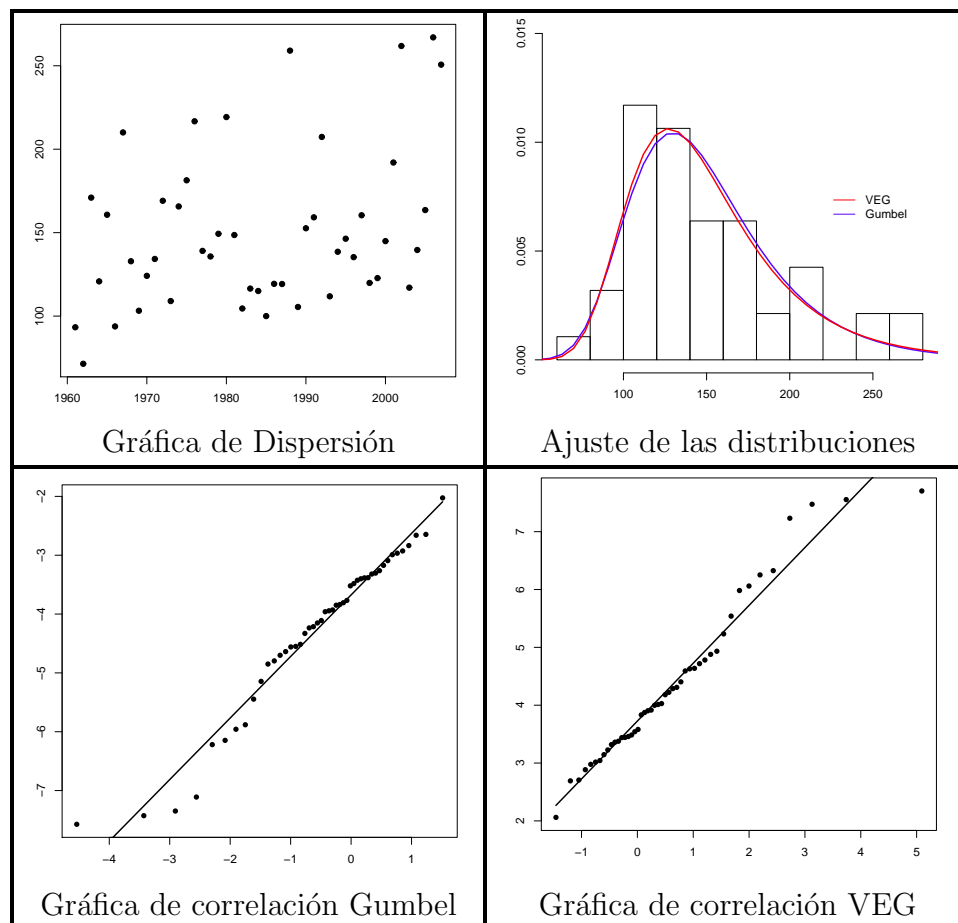


Figura 3.14: Gráficas del municipio de Jalpa de Méndez

Los municipios de Jonuta y Macuspana tienen como mejor modelo el DVEG, en ambos casos los dos criterios a favor son la Corr y L-veros. La estimación de Jonuta se encuentra en la tabla 3.16 y figura 3.15 y en cuanto a Macuspana los resultados se muestran en la tabla 3.17 y figura 3.16.

Estimación de Parámetros, Jonuta						
Dist.	Escala (σ)	Forma (ξ)	Loc (μ)	L-veros	AIC	Corr
VEG	27.860	0.135	121.934	-234.250	474.500	0.990
Gumbel	29.547	.	124.042	-234.984	473.968	0.986

IC	VEG	Gumbel
σ	(20.849, 34.871)	(27.351, 31.743)
ξ	(-0.105, 0.376)	.
μ	(112.830, 131.037)	(115.146, 132.937)

Tabla 3.16: Estimación del municipio de Jonuta

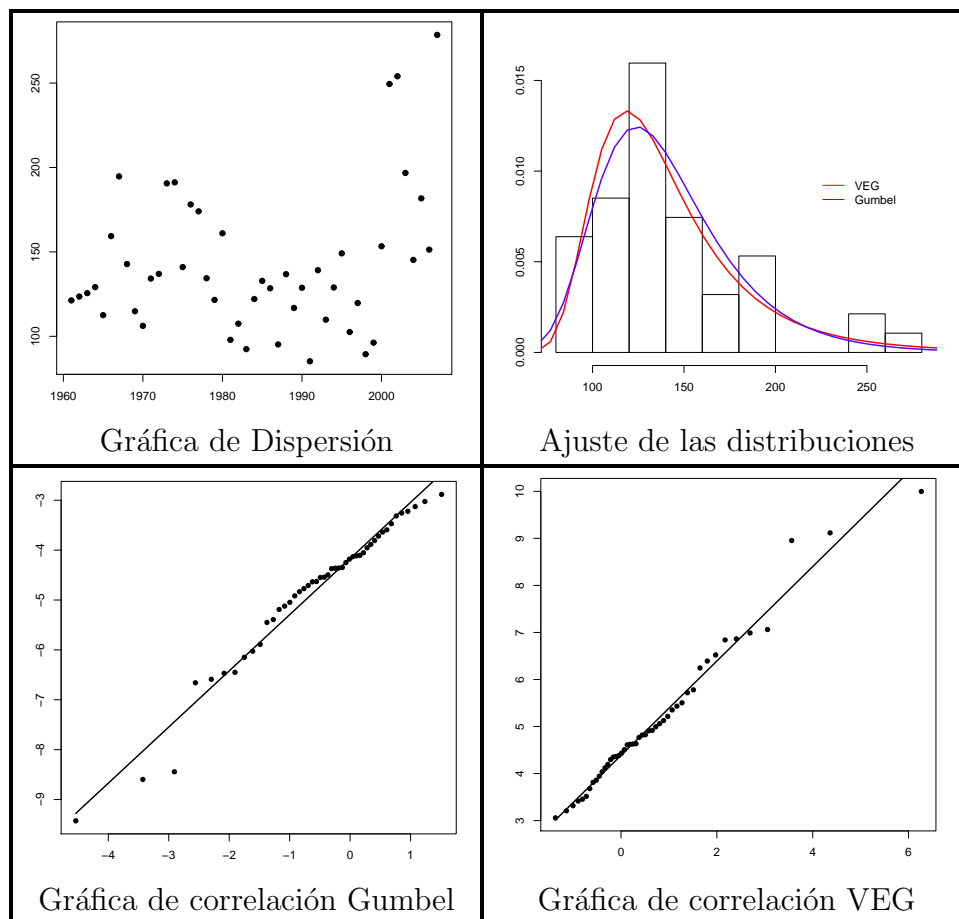


Figura 3.15: Gráficas del municipio de Jonuta

Estimación de Parámetros, Macuspána						
Dist.	Escala (σ)	Forma (ξ)	Loc (μ)	L-veros	AIC	Corr
VEG	30.329	0.086	116.947	-236.968	479.935	0.995
Gumbel	31.367	.	118.384	-237.267	478.534	0.994

IC	VEG	Gumbel
σ	(22.946, 37.712)	(29.171, 33.563)
ξ	(-0.141, 0.313)	.
μ	(107.127, 126.767)	(108.94, 127.828)

Tabla 3.17: Estimación del municipio de Macuspána

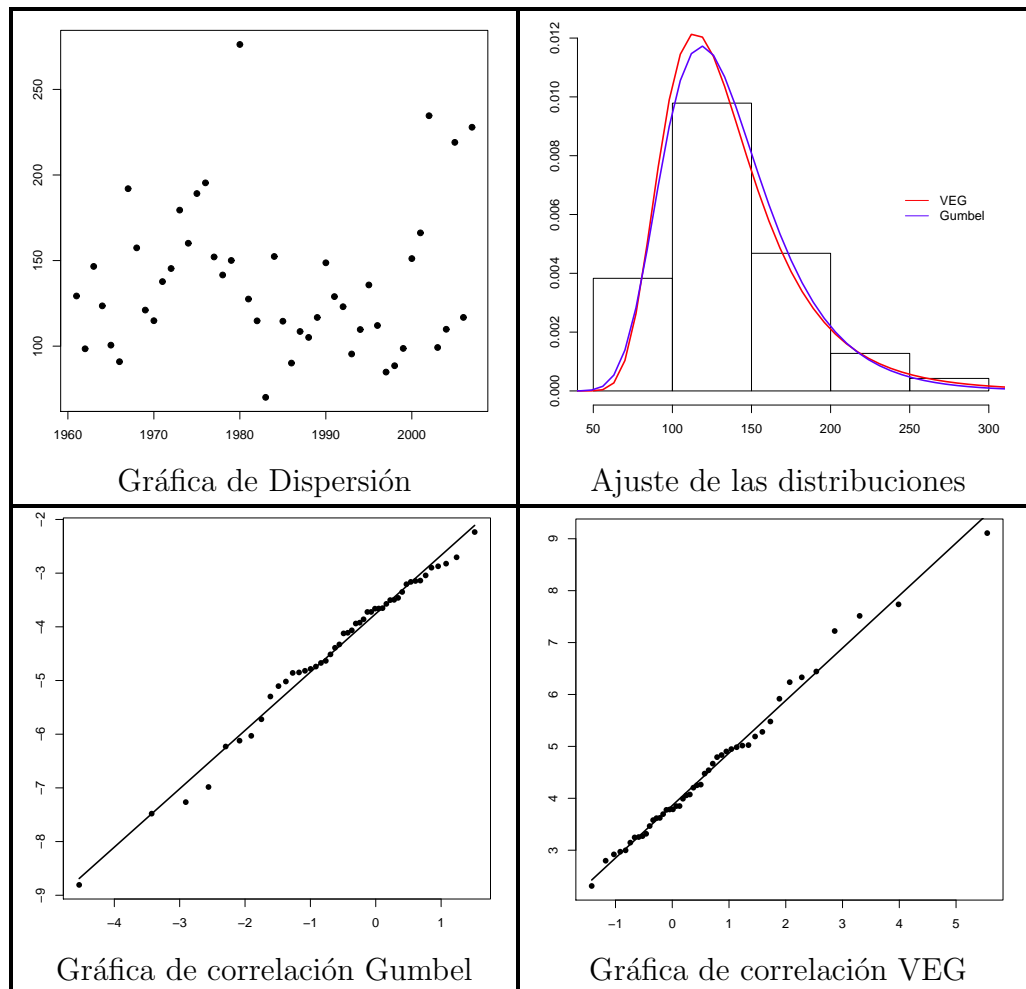


Figura 3.16: Gráficas del municipio de Macuspána

En la tabla 3.18 y figura 3.17 se encuentra la estimación correspondiente al municipio de Nacajuca en el que se consideró como mejor modelo el dado por la DVEG, cuenta con la Corr

y el AIC a su favor, aunque el valor de la Corr no es significativamente diferente si resulta ser el mejor además de que el valor de AIC resulta ser notoriamente mejor.

Estimación de Parámetros, Nacajuca						
Dist.	Escala (σ)	Forma (ξ)	Loc (μ)	L-veros	AIC	Corr
VEG	35.250	0.049	128.885	-243.087	492.173	0.981
Gumbel	35.850	.	129.821	-243.182	490.364	0.983

IC	VEG	Gumbel
σ	(26.848, 43.652)	(33.654, 38.046)
ξ	(-0.173, 0.27)	.
μ	(117.516, 140.255)	(119.026, 140.615)

Tabla 3.18: Estimación del municipio de Nacajuca

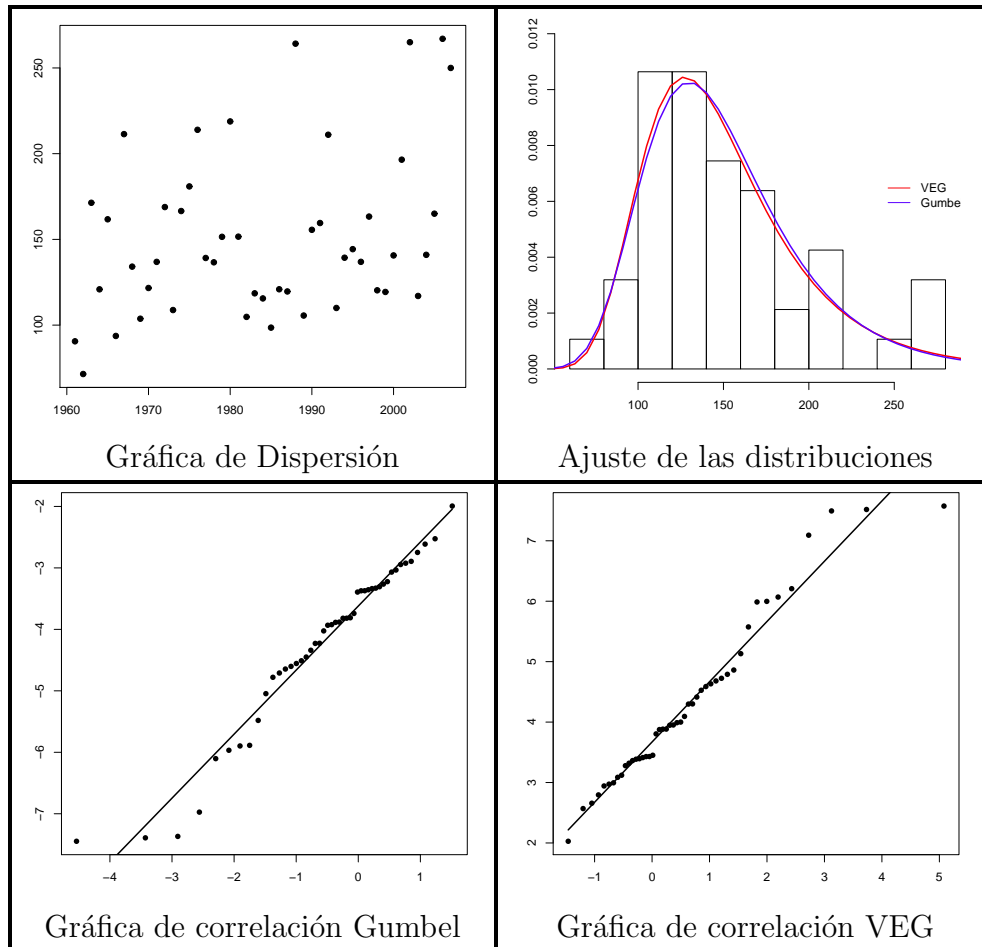


Figura 3.17: Gráficas del municipio de Nacajuca

En la tabla 3.19 y figura 3.18 se muestran los resultados de la estimación hecha al

municipio de Paraíso. En cuanto a los criterios resulta que para la Corr y el AIC la diferencia resulta ser muy pequeña pero mejor para el modelo de DVEG que para el modelo Gumbel.

Estimación de Parámetros, Paraíso						
Dist.	Escala (σ)	Forma (ξ)	Loc (μ)	L-veros	AIC	Corr
VEG	23.678	0.165	126.404	-227.475	460.950	0.989
Gumbel	25.585	.	128.613	-228.340	460.680	0.990

IC	VEG	Gumbel
σ	(17.465, 29.891)	(23.389, 27.781)
ξ	(-0.104, 0.435)	.
μ	(118.522, 134.286)	(120.909, 136.317)

Tabla 3.19: Estimación del municipio de Paraíso

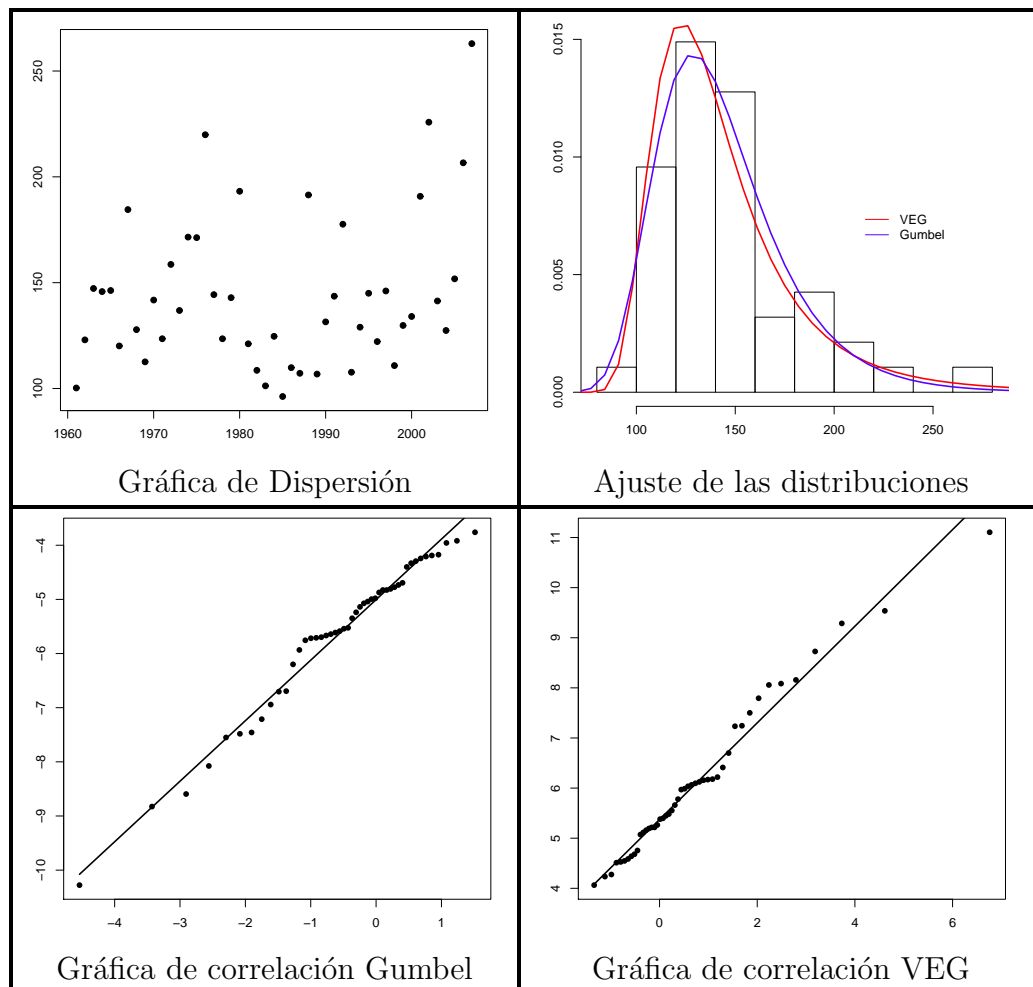


Figura 3.18: Gráficas del municipio de Paraíso

La estimación para el municipio de Tacotalpa se encuentra en la tabla 3.20 y en la figura 3.19 en las que se aprecia que el mejor modelo para las precipitaciones máximas es el de DVEG, en este modelo tanto la Corr como el criterio de AIC resultan ser mejores que los valores del modelo Gumbel.

Estimación de Parámetros, Tacotalpa						
Dist.	Escala (σ)	Forma (ξ)	Loc (μ)	L-veros	AIC	Corr
VEG	57.358	-0.101	162.237	-262.015	530.030	0.994
Gumbel	55.749	.	159.350	-262.358	528.716	0.988

IC	VEG	Gumbel
σ	(43.931, 70.785)	(53.553, 57.945)
ξ	(-0.33, 0.129)	.
μ	(143.593, 180.881)	(142.563, 176.138)

Tabla 3.20: Estimación del municipio de Tacotalpa

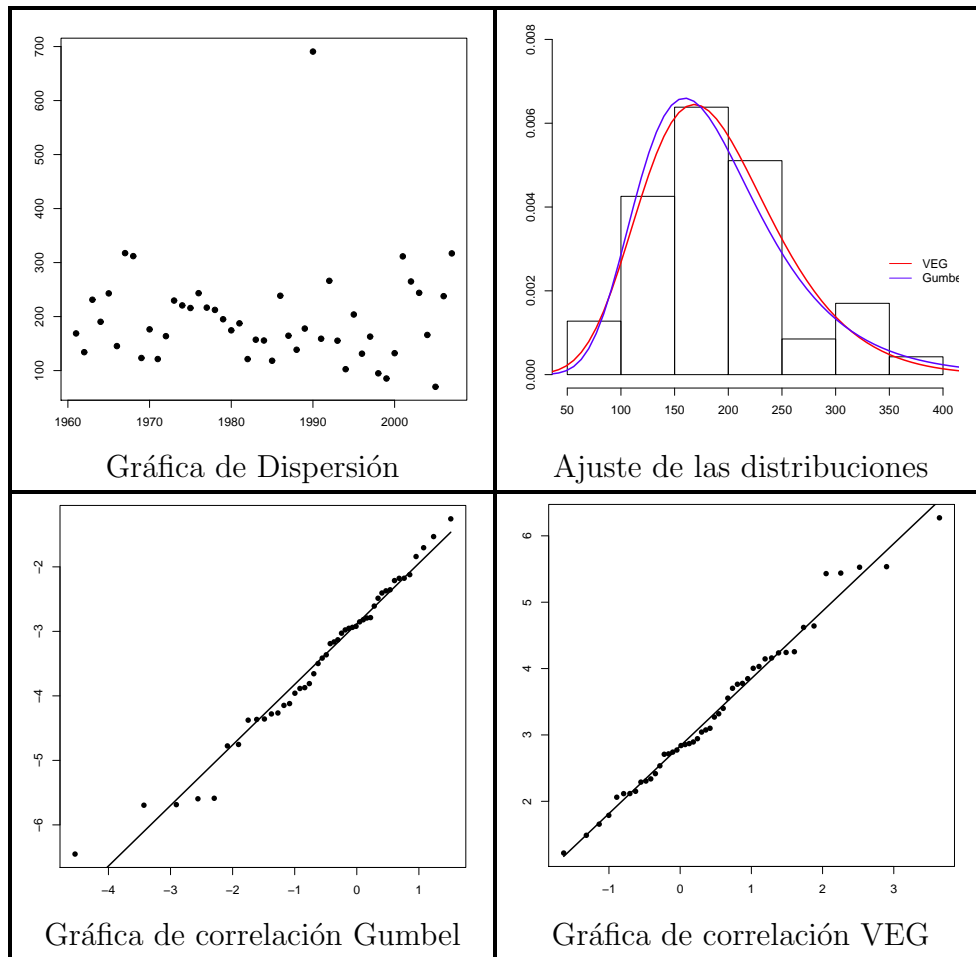


Figura 3.19: Gráficas del municipio de Tacotalpa

En el municipio de Teapa el criterio de Corr es igual para el ajuste de los dos modelos, por lo que la decisión del mejor modelo se basará en el criterio del AIC ya que proporciona un modelo con menor número de parámetros y por tanto más sencillo. Por lo que el modelo elegido es el Gumbel. Las estimaciones para Teapa se muestran en la tabla 3.21 y figura 3.20.

Estimación de Parámetros, Teapa						
Dist.	Escala (σ)	Forma (ξ)	Loc (μ)	L-veros	AIC	Corr
VEG	54.389	0.023	159.143	-262.578	531.157	0.992
Gumbel	54.816	.	159.779	-262.601	529.202	0.992

IC	VEG	Gumbel
σ	(41.489, 67.289)	(52.62, 57.012)
ξ	(-0.192, 0.237)	.
μ	(141.575, 176.711)	(143.277, 176.282)

Tabla 3.21: Estimación del municipio de Teapa

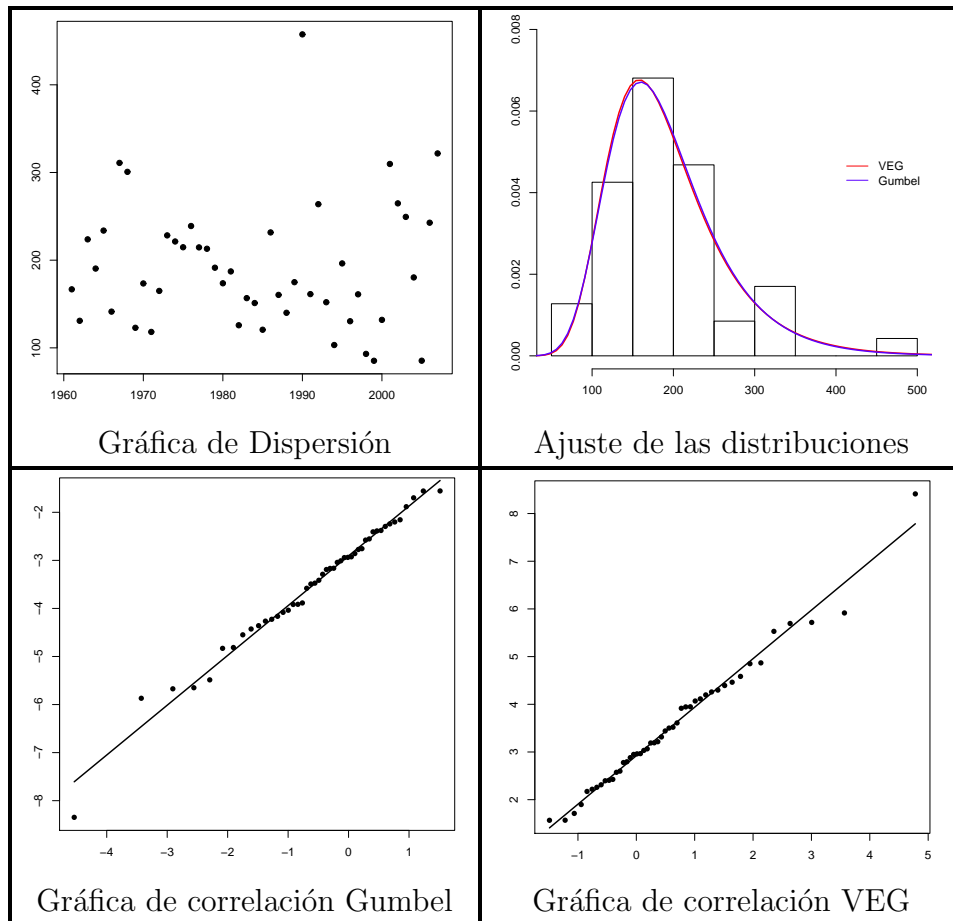


Figura 3.20: Gráficas del municipio de Teapa

Por último los resultados del municipio de Tenosique se muestran en la tabla 3.22 y en la figura 3.21. Para las lluvias máximas de Tenosique se elige como mejor modelo el de DVEG, éste es el único caso en el que los tres criterios (Corr, AIC y L-veros), resultan ser favorables para este modelo.

Estimación de Parámetros, Tenosique						
Dist.	Escala (σ)	Forma (ξ)	Loc (μ)	L-veros	AIC	Corr
VEG	29.447	0.282	91.188	-240.581	487.163	0.981
Gumbel	34.217	.	96.056	-243.058	490.117	0.973

IC	VEG	Gumbel
σ	(20.999, 37.894)	(32.021, 36.413)
ξ	(-0.028, 0.593)	.
μ	(81.143, 101.234)	(85.754, 106.357)

Tabla 3.22: Estimación del municipio de Tenosique

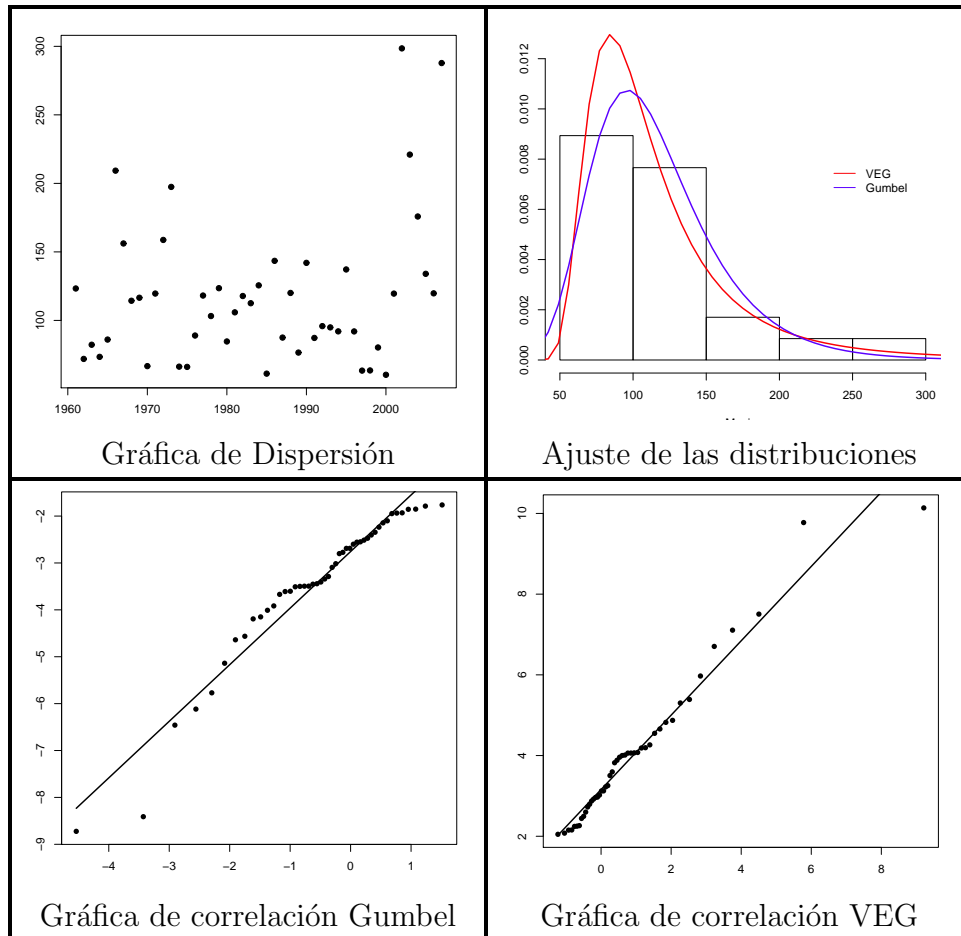


Figura 3.21: Gráficas del municipio de Tenosique

Municipios	Mejor Modelo
Balancán	$F(x) = \exp \left\{ - \left[1 + 0.129 \left(\frac{x - 117.294}{27.84} \right) \right]^{-\frac{1}{0.129}} \right\}$
Cárdenas	$G(x) = \exp \left\{ - \exp \left[- \left(\frac{x - 117.649}{30.99} \right) \right] \right\}$
Centla	$F(x) = \exp \left\{ - \left[1 + 0.129 \left(\frac{x - 125.749}{30.429} \right) \right]^{-\frac{1}{0.129}} \right\}$
Centro	$G(x) = \exp \left\{ - \exp \left[- \left(\frac{x - 142.057}{30.375} \right) \right] \right\}$
Comalcalco	$F(x) = \exp \left\{ - \left[1 + 0.115 \left(\frac{x - 105.52}{32.02} \right) \right]^{-\frac{1}{0.115}} \right\}$
Cunduacán	$G(x) = \exp \left\{ - \exp \left[- \left(\frac{x - 117.92}{36.307} \right) \right] \right\}$
Emiliano Z.	$F(x) = \exp \left\{ - \left[1 + 0.121 \left(\frac{x - 136.028}{32.384} \right) \right]^{-\frac{1}{0.121}} \right\}$
Huimanguillo	$G(x) = \exp \left\{ - \exp \left[- \left(\frac{x - 165.261}{53.403} \right) \right] \right\}$
Jalapa	$G(x) = \exp \left\{ - \exp \left[- \left(\frac{x - 144.847}{31.992} \right) \right] \right\}$
Jalpa	$G(x) = \exp \left\{ - \exp \left[- \left(\frac{x - 129.636}{35.255} \right) \right] \right\}$
Jonuta	$F(x) = \exp \left\{ - \left[1 + 0.135 \left(\frac{x - 121.934}{27.86} \right) \right]^{-\frac{1}{0.135}} \right\}$
Macuspana	$F(x) = \exp \left\{ - \left[1 + 0.086 \left(\frac{x - 116.947}{30.329} \right) \right]^{-\frac{1}{0.086}} \right\}$
Nacajuca	$G(x) = \exp \left\{ - \exp \left[- \left(\frac{x - 129.821}{35.85} \right) \right] \right\}$
Paraíso	$G(x) = \exp \left\{ - \exp \left[- \left(\frac{x - 128.613}{25.585} \right) \right] \right\}$
Tacotalpa	$F(x) = \exp \left\{ - \left[1 - 0.101 \left(\frac{x - 162.237}{57.358} \right) \right]^{-\frac{1}{-0.101}} \right\}$
Teapa	$G(x) = \exp \left\{ - \exp \left[- \left(\frac{x - 159.779}{54.816} \right) \right] \right\}$
Tenosique	$F(x) = \exp \left\{ - \left[1 + 0.282 \left(\frac{x - 91.188}{29.447} \right) \right]^{-\frac{1}{0.282}} \right\}$

Tabla 3.23: Modelos elegidos por municipio

En la tabla 3.23 se muestran los modelos elegidos como mejores para los 17 municipios de Tabasco. $F(x)$ se refiere a la DVEG y $G(x)$ a la distribución Gumbel. En la tabla 3.24 se muestran los resultados de la prueba de bondad de ajuste bootstrap basada en el coeficiente de correlación muestral como estadístico de prueba, para decidir sobre la distribución en cada municipio.

El juego de hipótesis a probar es el siguiente:

$$\begin{aligned} H_0 &: f(x) \in F^*(x) \\ H_1 &: f(x) \notin F^*(x) \end{aligned} \quad (3.4.1)$$

donde $F^*(x)$ denota una familia de densidades específica, en este caso se trata de la función de densidad Gumbel y la DVEG según sea el caso de cada municipio.

Municipio	Corr-muest	Estad-corr	Decisión
Balancán	0.9744	0.9446	No Rechazar H_0
Cárdenas	0.9894	0.9626	No Rechazar H_0
Centla	0.9835	0.9433	No Rechazar H_0
Centro	0.9933	0.9667	No Rechazar H_0
Comalcalco	0.9931	0.9469	No Rechazar H_0
Cunduacán	0.9849	0.9631	No Rechazar H_0
Emiliano Z.	0.9853	0.9464	No Rechazar H_0
Huimanguillo	0.9891	0.9634	No Rechazar H_0
Jalapa	0.9958	0.9658	No Rechazar H_0
Jalpa	0.9838	0.9652	No Rechazar H_0
Jonuta	0.9896	0.9398	No Rechazar H_0
Macuspana	0.9952	0.9418	No Rechazar H_0
Nacajuca	0.9828	0.9636	No Rechazar H_0
Paraíso	0.9905	0.9656	No Rechazar H_0
Tacotalpa	0.9936	0.8339	No Rechazar H_0
Teapa	0.9916	0.9643	No Rechazar H_0
Tenosique	0.9812	0.9158	No Rechazar H_0

Tabla 3.24: Resultados de las Pruebas de Bondad de Ajuste

Periodos de Retorno por Municipio

Una vez que se cuenta con el ajuste hecho a las precipitaciones máximas de los 17 municipios de Tabasco, se calculan sus respectivos periodos y niveles de retorno. Entonces, para los periodos se tiene que

$$T_x = \frac{1}{1 - F(x)}.$$

donde:

- T_x es el tiempo promedio para que ocurra una precipitación mayor o igual a x .
- $F(x)$ se refiere a la función de distribución de probabilidad acumulada.

En cuanto a los niveles de retorno, para la distribución de VEG el cuantil $1 - p$ está dado por:

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

con $0 < p < 1$. En términos de valores extremos, x_p es el nivel de retorno asociado con el periodo de retorno $1/p$.

En la literatura es usual calcular los niveles de retorno, pero en este caso es de interés escoger diferentes precipitaciones de lluvia y a éstos asignarles un tiempo promedio (el índice de riesgo se considera la probabilidad de lluvia), esto es $p = \frac{1}{T_x}$, para x igual a 100mm, 150mm, 200mm y 250mm con $F(x)$ la distribución Gumbel ó DVEG según el modelo de cada municipio. En la tabla 3.25 se muestran los resultados de los periodos de retorno.

Municipio	Periodos de Retorno			
	100 mm	150 mm	200 mm	250 mm
Balancán	1.174	3.513	12.860	41.492
Cárdenas	1.206	3.370	14.765	72.082
Centla	1.094	2.674	8.852	27.072
Centro	1.019	1.862	7.249	35.443
Comalcalco	1.437	4.150	13.180	38.315
Cunduacán	1.241	2.954	10.099	38.515
Emiliano	1.038	2.077	6.397	19.303
Huimanguillo	1.035	1.359	2.460	5.405
Jalapa	1.018	1.745	6.122	27.262
Jalpa	1.109	2.328	7.870	30.891
Jonuta	1.112	3.104	11.265	36.163
Macuspana	1.205	3.364	12.205	41.889
Nacajuca	1.112	2.303	7.594	29.071
Paraíso	1.049	2.843	16.789	115.446
Tacotalpa	1.065	1.410	2.518	5.785
Teapa	1.054	1.434	2.623	5.702
Tenosique	1.295	2.191	13.073	27.014

Tabla 3.25: Diferentes escenarios para los periodos de retorno

Por ejemplo para el municipio de Centro resulta que el tiempo promedio para que ocurra una precipitación mayor o igual a 150mm es 2.7 años aproximadamente (2 años nueve meses), así sucesivamente para cada x y cada uno de los 17 municipios.

Ahora, para la tabla 3.26 referente a los niveles de retorno se consideran tres escenarios diferentes: 5, 10 y 20 años. Entonces, para el municipio de Balancán se espera que se tenga una lluvia de 193.15 mm en 5 años, una de 222.82 mm en 10 años y una de 254.12 mm en 20 años, por mencionar un ejemplo.

Municipio	Nivel de Retorno		
	5 años	10 años	20 años
Balancán	193.145	222.816	254.115
Cárdenas	149.061	159.161	168.849
Centla	208.628	241.036	275.216
Centro	172.846	182.746	192.242
Comalcalco	191.281	224.135	258.437
Cunduacán	154.721	166.554	177.904
E. Zapata	223.313	257.011	292.327
Huimanguillo	219.393	236.797	253.492
Jalapa	177.275	187.701	197.702
Jalpa	165.372	176.862	187.883
Jonuta	198.404	228.587	260.567
Macuspana	195.327	224.054	253.405
Nacajuca	166.159	177.843	189.050
Paraíso	154.547	162.885	170.883
Tacotalpa	281.519	314.279	343.457
Teapa	215.343	233.208	250.344
Tenosique	188.471	236.037	292.179

Tabla 3.26: Diferentes niveles de retorno

Finalmente en las figuras 3.22-3.25 se muestran los mapas con cuatro diferentes escenarios de peligro (100mm, 150mm, 200mm y 250mm), se eligieron estos cuatro escenarios dado que los registros históricos muestran que se han presentado algunos problemas de inundación con una precipitación mayor a 100mm y los tres restantes se dan para mostrar que la ocurrencia de lluvias mayores a 100mm en Tabasco disminuye el peligro más no el riesgo.

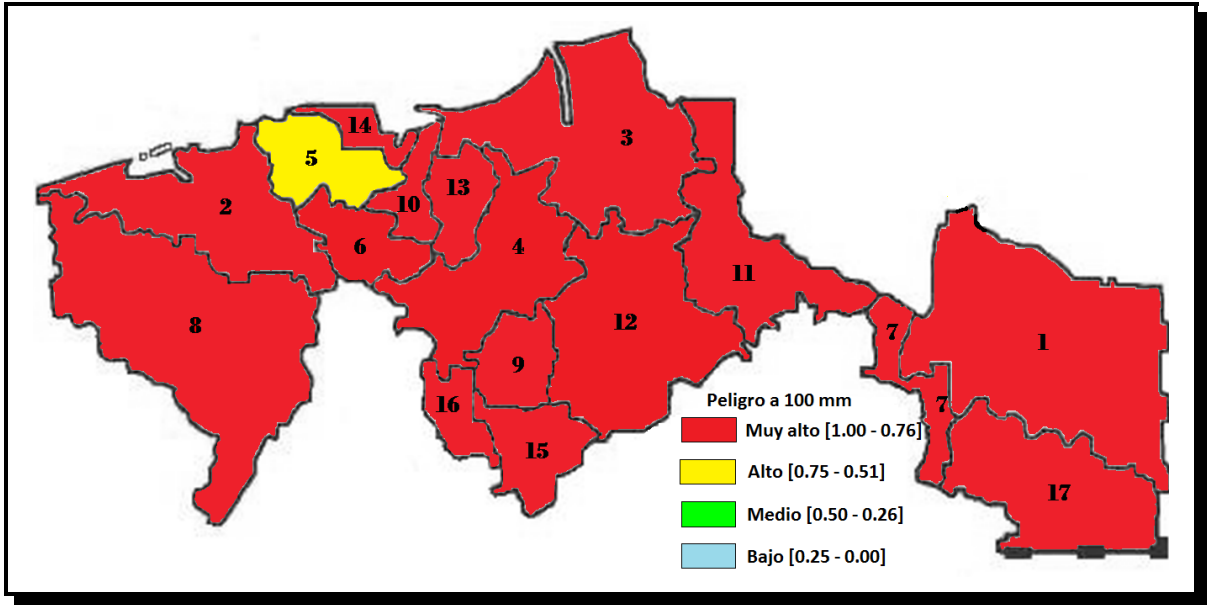


Figura 3.22: Mapa de peligro a un nivel de retorno de 100mm.

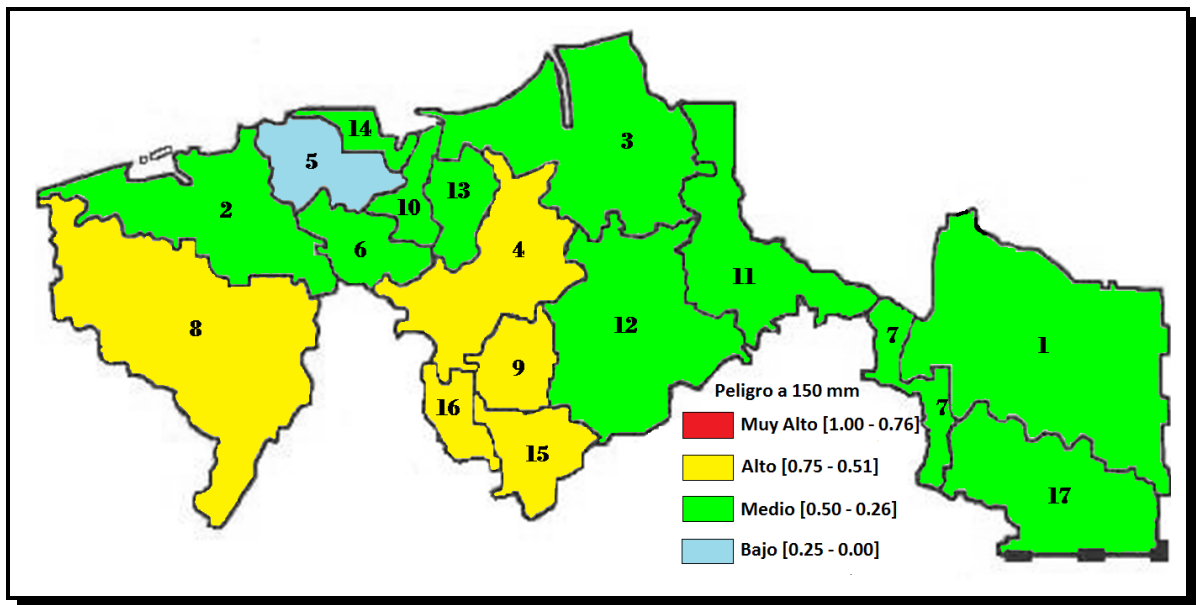


Figura 3.23: Mapa de peligro a un nivel de retorno de 150mm.

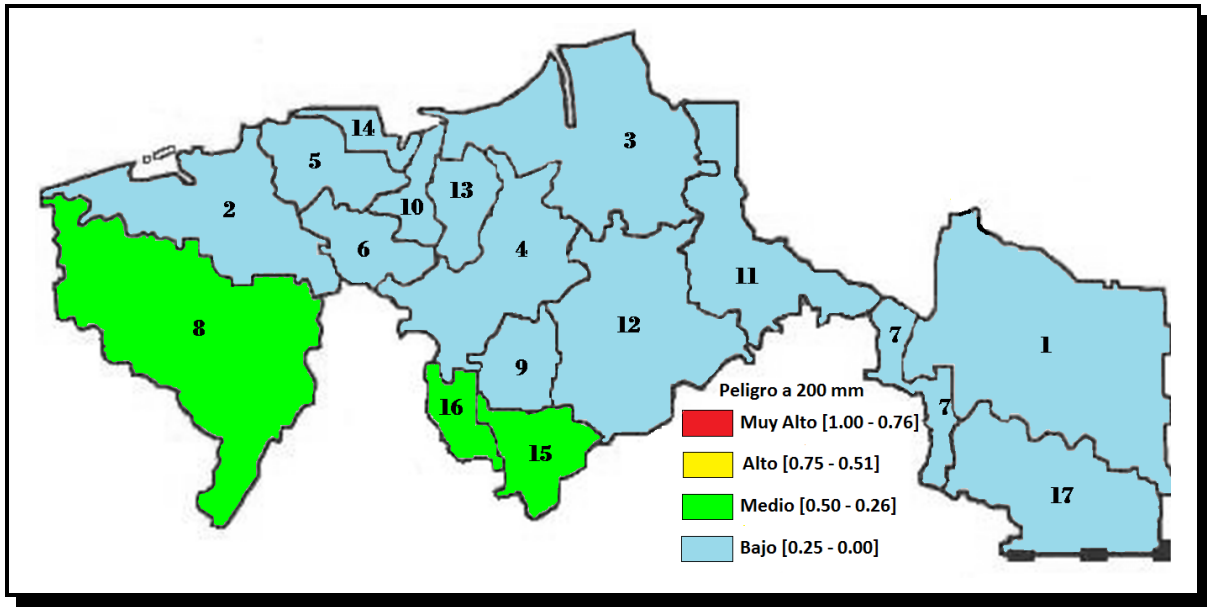


Figura 3.24: Mapa de peligro a un nivel de retorno de 200mm.

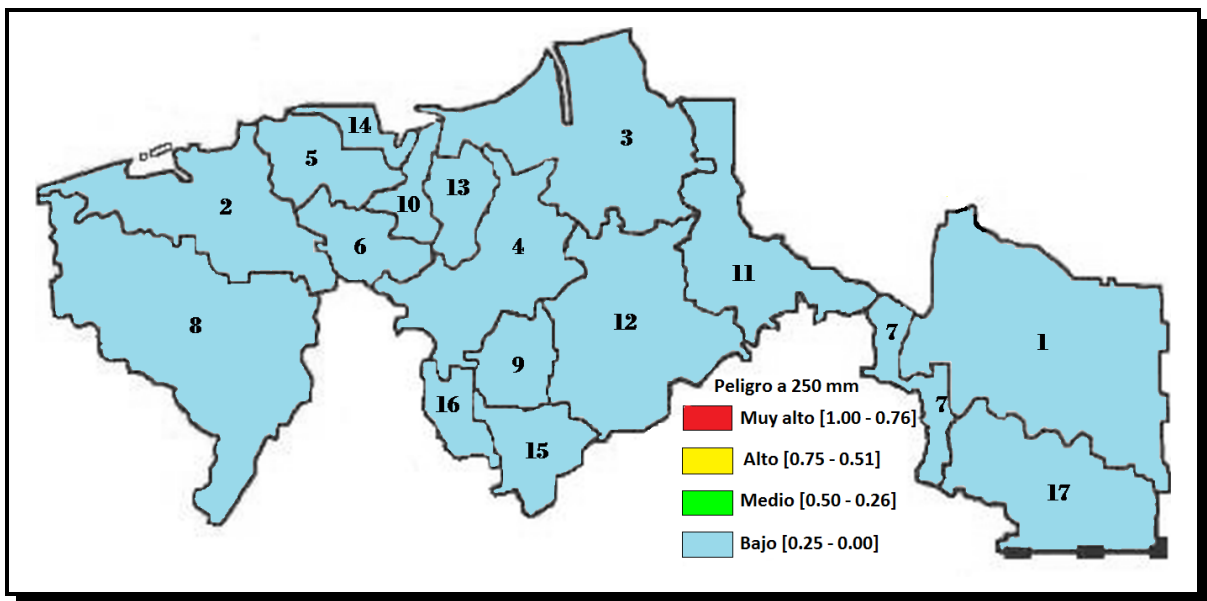


Figura 3.25: Mapa de peligro a un nivel de retorno de 250mm.

3.5. Resultados del Índice de Riesgo

Finalmente se ha llegado a los resultados del Índice de Riesgo propuesto, para la construcción de este indicador se utilizó la fórmula

$$R = C \times V \times P,$$

donde C es el factor costo, V es el factor vulnerabilidad y P el factor peligro.

El factor de vulnerabilidad se toma de la tabla 3.2, se refiere a la columna de InVuln, para el factor costo se toma la columna InCosto de la tabla 3.4 y en cuanto al factor peligro se toma la probabilidad p del evento $\{X_i > x\}$, para este índice de riesgo se toma a $x = 150mm$ ya que la experiencia sugiere que a partir de este valor se han presentado problemas de inundación.

En la tabla 3.27 se muestra el resultado final de aplicar la fórmula de Riesgo, se da el valor del indicador y la categoría en la que se ubica, el índice se clasificó en cuatro categorías: Bajo, Medio, Alto y Muy alto.

Núm.	Municipios	C	V	P150	IR	Categoría
1	Balancán	2.514	0.221	0.285	0.158	Bajo
2	Cárdenas	2.955	0.556	0.297	0.487	Medio
3	Centla	2.624	0.664	0.374	0.652	Alto
4	Centro	3.403	0.526	0.537	0.962	Muy Alto
5	Comalcalco	2.866	0.366	0.241	0.253	Medio
6	Cunduacán	2.757	0.632	0.339	0.590	Alto
7	E. Zapata	2.571	0.065	0.481	0.080	Bajo
8	Huimanguillo	2.770	0.526	0.736	1.000	Muy Alto
9	Jalapa	2.577	0.405	0.573	0.598	Alto
10	Jalpa	2.734	0.373	0.429	0.438	Medio
11	Jonuta	2.313	0.138	0.322	0.103	Bajo
12	Macuspana	2.806	0.564	0.297	0.470	Medio
13	Nacajuca	2.901	0.166	0.434	0.210	Bajo
14	Paraíso	2.820	0.467	0.352	0.463	Medio
15	Tacotalpa	2.333	0.220	0.709	0.364	Medio
16	Teapa	2.564	0.124	0.697	0.221	Bajo
17	Tenosique	2.596	0.218	0.456	0.258	Medio

Tabla 3.27: Índice de Riesgo

Finalmente en la figura 3.26 se muestra el mapa del Índice de Riesgo propuesto.

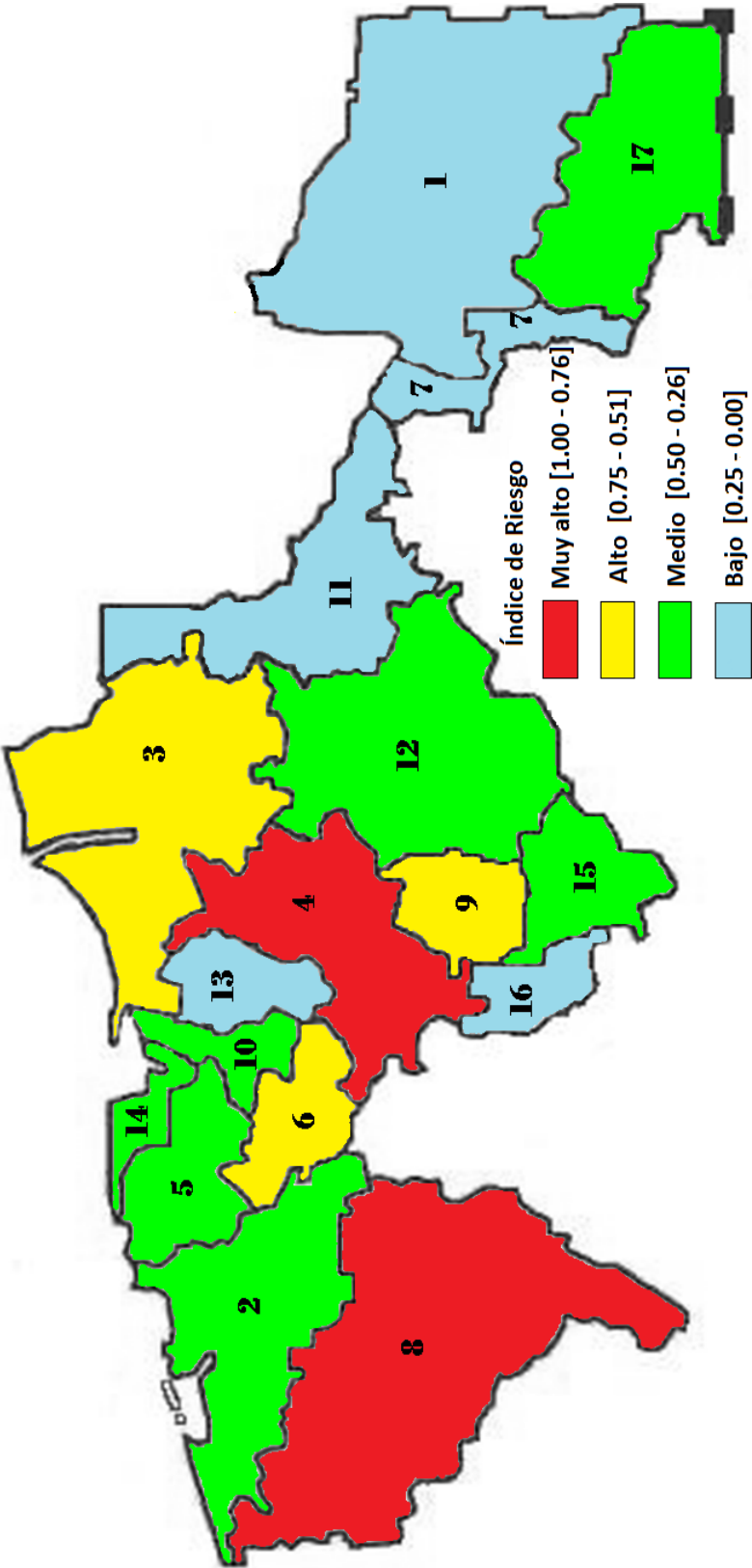


Figura 3.26: Índice de Riesgo de Inundaciones

Conclusiones

Como se ha mencionado los mapas de riesgo ayudan a prevenir los problemas que se pueden presentar ante una inundación. Si se hace una comparación entre el índice de riesgo propuesto con los registros históricos sobre desastres por inundaciones en los 17 municipios de Tabasco en los años 1982, 1989, 1995, 1998, 2002, 2003, 2005-2009, se puede concluir que los resultados obtenidos fueron acordes a los registros históricos. El comparativo se muestra en la tabla C.28

En algunos municipios no coinciden los resultados, esto se puede deber a diferentes situaciones, por ejemplo: no se encontraron suficientes registros históricos de las inundaciones que se han presentado en todos los municipios y en particular para el caso del Centro el índice de servicios obtenido fue el más bajo pero no se pudo tomar en cuenta algunos otros factores que aumentan su vulnerabilidad. En general se puede concluir que los factores que fueron incluidos en la construcción de los índices fueron correctos, así como el uso de los componentes principales para obtener las ponderaciones adecuadas para cada uno de los factores considerados.

Mun.	Balancán	Cárdenas	Centla	Centro	Comalcalco	Cunduacán
IH	Bajo	Alto	Alto	Muy Alto	Alto	Alto
IRP	Bajo	Medio	Alto	Alto	Medio	Alto
Mun.	E. Zapata	Huimang.	Jalapa	Jalpa	Jonuta	Macuspana
IH	Bajo	Alto	Alto	Medio	Bajo	Medio
IRP	Bajo	Muy alto	Alto	Medio	Bajo	Medio
Mun.	Nacajuca	Paraíso	Tacotalpa	Teapa	Tenosique	
IH	Medio	Medio	Bajo	Bajo	Medio	
IRP	Bajo	Medio	Medio	Bajo	Medio	

Tabla C.28: Comparativo de inundaciones históricas e IRP por municipio (Mun.).

Es importante mencionar que el factor que tiene mayor influencia en el índice de riesgo propuesto es la vulnerabilidad. Recuérdese que en este estudio para obtener un indicador de vulnerabilidad se incluyeron factores naturales y sociales, y al menos en uno de ellos se puede trabajar para prevenir las pérdidas ocasionadas por inundaciones. Ahora, en cuanto al peligro se sabe que aún no se puede evitar, pero con un alto grado de confiabilidad se puede predecir y con esto preparar a la población para minimizar las pérdidas tanto económicas como sociales.

Bibliografía

- [1] Casson E. and Coles S. (2000). Simulation and Extremal Analysis of Hurricane Events. *Journal of the Royal Statistical Society*, 49(1), 227-245.
- [2] Bayliss, A. C., Reed, D. W. (2001). The use of historical data in flood frequency estimation. Walingford: Report to MAFF. Center for Ecology and Hidrology.
- [3] Begum, S., Merz, B., Thiekm, A. H., Gocht, M. . (2007). Flood risk Management in Europe. Springer , 231-251.
- [4] Blaikie P., Wisner B., Davis I., Cannon T. (1994). *At Risk. Natural Hazards, people´s vulnerability and disasters*. New York: Routledge, London.
- [5] Brillinger, D. R. (2003). Three Environmental Probabilistic Risk Problems. *Statistical Science*, 18(4), 412-421.
- [6] Castillo, E., Hadi, A. S., Balakrishnan, N. and Sarabia, J. M. (2005). *Extreme Value and Related Models with Applications in Engineering and Science*. New York: Wiley.
- [7] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer-Verlag.
- [8] Coles, S. and Dixon, M. (1999). Likelihood-Based Inference for Extreme Values Models Extremes. *Journal of the Royal Statistical Society*, 2(1), 5-23.
- [9] Coles, S. and Perichi, L. (2003). Anticipating Catastrophes through Extreme Value Modelling. *Journal of the Royal Statistical Society*, 52(4), 405-416.
- [10] Coles, S. G. and Tawn, J. A. (1996). Modelling Extremes of the Areal Rainfall Process. *Journal of the Royal Sattistical Society*, 58(2), 329-347.
- [11] Comisión Nacional del Agua. (2007). *Estadísticas del agua en México*. México.

-
- [12] Centro Nacional de Prevención de Desastres. (2006). Guía Básica para la Elaboración de Atlas Estatales y Municipales de Peligros y Riesgos. México.
- [13] Duncan, W. R. (2002). Reinforcing Flood-Risk Estimation. *Philosophical Transaction: Mathematical, Physical and Engineering Sciences*, 1373-1387.
- [14] Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24, 180-190.
- [15] Fuentes Mariles, O. F. (1996). Escurrimiento en ríos y volúmenes de inundación por desbordamiento. *CENAPRED: Cuadernos de Investigación* 26.
- [16] Fuentes Mariles, O. F. (1997a). Modelo matemático de áreas de inundación. *CENAPRED: Cuadernos de Investigación* 41.
- [17] Gnedenko, B. (1943). Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *The Annals of Mathematics*, 44(3), 423-453.
- [18] Green, C. H., Penning-Rowsell, E. C. & Parker, D. J. (1987). Estimating the risk from flooding and evaluating worry. In *uncertainty in risk assessment, risk management and decision making*. New York: Plenum.
- [19] Green, C. H., Van der Veen, A., Wiertra, E. & Penning-Rowsell, E. (1994). *Vulnerability refined: analysing full flood impacts*. London: Middlesex University Press.
- [20] Gumbel, E. J. (1934). Les moments des distribution finales de la première et de la dernière valeur. *Comptes Rendus de l'Académie des Sciences*, 198, 141-143.
- [21] Gumbel, E. J. (1935a). Les valeurs extremes des distributions statistiques. *Annales de l'Institut Henri Poincaré*, 5, 115-158.
- [22] Gumbel, E. J. (1958). *Statistics of extremes*. New York: Columbia Univ. Press.
- [23] Gumbel, E. J. (1941b). The return period of flood flows. *Annals Mathematics Statistics*, XII(2), 1-3.
- [24] Guzmán, M. (2011). Modelación de valores máximo de ozono. Tesis de Maestría. Colegio de Postgraduados.
- [25] Hunt, J. C. (2002). Floods in a Changing Climate: A Review. *Philosophical Transaction: Mathematical, Physical and Engineering Sciences*, 1531-1543.
- [26] Jenkinson, A. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quart. J. Roy. Meteor. Soc.*, 81, 158-171.

- [27] Lehner, B. and Döll, P. (2001). Europe's flood today and in the future. Model-based assessment of european water resources and hidrology in the face of global change. Center for Enviromental Sistems Research, University of Kassel, 5, 6.1-6.14.
- [28] Mileti, D. E. (1999). Disasters by design. A reassessment of natural hazard in the United States. Washington, DC: Joseph Henry Press.
- [29] Ordaz, M. (1996). algunos Conceptos del Analisis de Riesgo. CENAPRED, 14, 6-12.
- [30] Penning-Rowsell, E., Fordham, M. (1994). Floods across Europe: flood hazard assessment, modelling and management. London: Middlesex University Press.
- [31] Rosbjerg, D. and Madsen, H. (1996). The role of regional information in estimation of extreme point rainfall. Atmos. Res., 42, 113-122.
- [32] Sexto, M. (2010). Modelación de Eventos Extremos Usando la Distribución Dagum. Tesis de Maestría. Colegio de Postgraduados.
- [33] Smith, R. (2001). Flood damage estimation a review of urban stage-damage curves and loss function. Water S. A., 20, 231-238.
- [34] Smolka, A. (2006). Natural Disaster and the Challenge of Extreme Events: Risk Management from an Insurance Perspective. Philosophical Transactions: Mathematical, Physical and Engineering Sciences , 364 (1845), 2147-2165.
- [35] Tapsell, S. M., Penning-Rowsell, S. M., Tunstall, T. L. (2002). Vulnerability to Flooding: Health and Social Dimensions. Philosophical Transactions: Mathematical, Physical an Engineering Sciences, 360, 1511-1525.
- [36] Tawatchi, T., Mohammed, F. K. (2005). Flood Hazard and risk analysis in the southwest region of Bangladesh. Hydrological Processes, 19, 2055-2069.
- [37] Tawn, J. A. (1992). Estimating Probabilitis of Extreme Sea-Levels. Journal of the Royal Statistical Society, 41(1), 77-93.
- [38] Von Mises, R. (1936). La distribution de la plus grande de n valeurs. Revue Mathématique de l'Union Interbalkanique (Athens), 1, 141-160.
- [39] Young, P. C. (2002). Advances in real time flood forecasting. Philosophical Transaction: Mathematical, Physical an Engineering Sciences, 1433-1450.

Anexos

En esta parte se muestra la información utilizada en el cálculo de los indicadores de servicios, de bienes y los datos de la precipitación máxima anual por municipio. Se revisarán los programas para el cálculo de las componentes principales en SAS, así como los programas hechos en R para la estimación de los parámetros con sus gráficas y los programas para la estimación de tamaño y potencia de las pruebas dadas en el capítulo 2.

A.1. Anexo A: Programas para la estimación de CP en SAS

Este programa fue realizado en SAS para el cálculo de Componentes Principales que se utilizan en la obtención de los indicadores de servicios y de bienes y además en la obtención de la serie de datos por municipio a la cual se extrajo los máximos anuales. El programa es el mismo para los diferentes cálculos, lo único que se modifica son las variables.

```
DATA CP;
INPUT Mun $ y1-y5;
CARDS;
Municipio PODT PODR PODL PODC PODA
Balancán 85.94 72.320 58.71 11.43 20.26
Cárdenas 88.78 79.310 69.37 17.57 25.01
Centla 84.10 73.265 58.29 9.76 11.75
...
Tenosique 89.03 79.630 62.99 16.29 19.98;
PROC PRINCOMP OUT=VOYCP COV PREFIX=compr;
VAR y1 y2 y3 y4 y5;

PROC PRINT DATA=VOYCP;
PROC CORR;
VAR y1-y5 compr1-compr5;
RUN;
```

A.2. Anexo B: Información de disposición de servicios en los municipios de Tabasco

La información que se muestra en la tabla A.29 fue tomada de INEGI 2010.

Municipio	PO_NDEE	PO_NDAE	PO_NDD
Balancán	4.96	16.29	5.30
Cárdenas	1.28	26.14	4.61
Centla	2.15	59.66	9.93
Centro	0.26	3.46	1.16
Comalcalco	0.89	16.17	3.97
Cunduacán	0.90	33.77	5.49
Emiliano Zapata	0.92	3.82	0.77
Huimanguillo	3.01	51.13	6.22
Jalapa	0.87	8.70	3.57
Jalpa de Méndez	0.68	11.15	3.95
Jonuta	2.63	6.23	5.21
Macuspana	1.49	30.77	5.92
Nacajuca	0.42	2.71	2.63
Paraíso	0.53	9.50	2.31
Tacotalpa	2.16	13.47	12.89
Teapa	1.30	12.59	2.64
Tenosique	2.70	15.40	5.05

Tabla A.29: Datos sobre disponibilidad de servicios por municipio

PO_NDEE: Porcentaje de Ocupantes que no Disponen de Energía Eléctrica.

PO_NDAE: Porcentaje de Ocupantes que no Disponen de Agua Entubada.

PO_NDD: Porcentaje de Ocupantes que no Disponen de Drenaje.

A.3. Anexo C: Información de bienes con los que cuenta la población por municipio

La información que se muestra en la tabla A.30 fue tomada de INEGI 2010.

Municipio	PODT	PODR	PODL	PODC	PODA
Balancán	85.94	72.320	58.71	11.43	20.26
Cárdenas	88.78	79.310	69.37	17.57	25.01
Centla	84.10	73.265	58.29	9.76	11.75
Centro	95.56	91.921	81.10	35.43	37.64
Comalcalco	88.10	77.904	64.40	15.30	23.95
Cunduacán	88.33	79.128	64.24	11.62	21.52
Emiliano Zapata	92.71	87.992	71.25	19.30	25.38
Huimanguillo	84.34	73.550	60.14	10.78	20.28
Jalapa	91.05	87.183	69.16	11.78	23.00
Jalpa de Méndez	90.36	82.720	68.48	14.67	23.07
Jonuta	80.62	71.455	55.84	5.59	9.76
Macuspana	87.09	79.961	61.37	14.72	20.82
Nacajuca	93.52	88.222	76.72	21.00	24.73
Paraíso	92.19	86.898	74.37	20.87	31.27
Tacotalpa	75.41	64.647	42.27	9.67	11.87
Teapa	89.32	79.286	59.75	16.10	19.30
Tenosique	89.03	79.630	62.99	16.29	19.98

Tabla A.30: Datos sobre disponibilidad de bienes por municipio

PODT: Porcentaje de Ocupantes que Disponen de Televisión.

PODR: Porcentaje de Ocupantes que Disponen de Refrigerador.

PODL: Porcentaje de Ocupantes que Disponen de Lavadora.

PODC: Porcentaje de Ocupantes que Disponen de Computadora.

PODA: Porcentaje de Ocupantes que Disponen de Automóvil o camioneta.

A.4. Anexo D: Breve explicación de la matriz de covarianzas, valores y vectores propios

Matriz de varianzas y covarianzas

La variabilidad respecto a la media se mide habitualmente por la varianza, pero la relación lineal entre dos variables X y Y se puede medir por la covarianza, si el valor de la covarianza es alto, entonces se considerará que las variables son significativamente dependientes. A diferencia de la varianza, la cual siempre deberá tomar valores positivos, la covarianza puede ser negativa con lo que se infiere que las variables son inversamente dependiente.

La información anterior en el caso multivariado puede presentarse de forma compacta en la matriz de varianzas y covarianzas. En la diagonal principal se colocan las varianzas, mientras que en los otros componentes de la matriz se ponen los valores de las covarianzas que corresponderán a las variables identificadas por la fila y columna del componente de la matriz. En forma general se define como:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

Así como la varianza es siempre un número no negativo, la matriz de varianzas y covarianzas tiene una propiedad similar: es semidefinida positiva. Esta propiedad asegura que si \mathbf{x} es cualquier vector, $\mathbf{x}'\Sigma\mathbf{x} \geq 0$. También la traza, el determinante y los valores propios de esta matriz son no negativos.

Valores y vectores propios

Los valores y vectores propios representan a los vectores (diferentes del vector cero) que bajo cierta transformación lineal T del vector da como resultado otro vector paralelo al primero con valor de proporcionalidad igual al valor propio. Así cuando se asocia la matriz de transformación A se tiene que $\mathbf{u} \neq 0$ es un vector propio de la matriz \mathbf{A} si existe un escalar λ denominado valor propio de la matriz, tal que

$$\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$$

Para calcular el vector propio se puede reescribir la ecuación anterior como:

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = 0$$

y éste es un sistema homogéneo de ecuaciones que tendrá solución no nula si y sólo si la matriz del sistema, $(\mathbf{A} - \lambda\mathbf{I})$, es singular. Esta ecuación es polinómica en λ de orden n y sus n raíces se denominan valores propios de la matriz. A cada valor propio distinto de una matriz se puede asociar al menos un vector propio y a lo más k vectores propios. En donde, k es la multiplicidad del valor propio en la ecuación polinómica.

A.5. Anexo E: Precipitación Máxima Anual en mm, por municipio

En este anexo se muestran los datos de precipitación máxima en 24 horas de cada año utilizados para el análisis del componente peligro para los 17 municipios de Tabasco.

A.5. ANEXO E: PRECIPITACIÓN MÁXIMA ANUAL EN MM, POR MUNICIPIO

Año	Balancán	Cárdenas	Centla	Centro	Comalcalco	Cunduacán
1961	116.92	118.98	122.60	130.30	94.53	147.92
1962	133.38	91.99	252.99	115.38	96.01	70.17
1963	118.28	135.59	127.27	171.07	119.39	163.36
1964	127.94	175.30	136.90	171.65	204.58	119.33
1965	114.88	146.43	119.90	132.13	142.82	141.00
1966	193.34	108.05	181.77	101.45	111.18	96.37
1967	176.78	160.92	187.18	222.69	129.47	183.12
1968	130.73	111.06	136.29	164.40	107.95	107.91
1969	110.31	121.06	112.36	113.00	131.05	92.55
1970	95.55	196.70	107.70	136.52	213.65	173.21
1971	139.84	82.94	143.06	117.14	79.80	81.62
1972	143.84	159.48	145.14	168.96	151.37	173.84
1973	184.82	137.76	182.34	155.77	147.26	113.47
1974	188.66	151.31	214.19	178.18	139.67	149.29
1975	117.24	187.66	134.29	165.83	188.55	190.49
1976	142.86	270.92	165.81	163.25	283.10	272.53
1977	146.48	136.94	162.50	145.13	136.62	137.18
1978	123.61	107.03	131.08	170.31	90.55	117.63
1979	113.80	136.03	124.23	132.36	146.59	106.23
1980	138.67	199.39	151.10	208.96	183.54	228.50
1981	96.37	91.46	98.67	165.10	85.60	88.40
1982	117.31	109.81	113.93	117.53	111.02	99.59
1983	101.50	73.14	99.22	108.91	70.71	75.67
1984	115.30	132.26	121.23	118.78	146.74	104.52
1985	90.00	84.58	126.22	137.02	60.12	128.96
1986	130.30	79.63	129.11	117.13	70.04	88.03
1987	89.70	99.88	97.51	147.05	92.37	111.08
1988	129.97	117.42	144.07	191.63	98.44	157.85
1989	119.39	102.29	124.27	143.85	91.95	102.98
1990	134.68	107.63	135.81	172.29	81.83	93.80
1991	77.38	170.95	80.15	142.75	175.52	151.79
1992	134.95	140.74	147.67	222.54	144.37	133.89
1993	112.17	117.23	117.13	146.90	91.74	148.39
1994	133.06	110.24	141.66	155.15	96.31	123.69
1995	153.06	156.25	149.04	129.93	140.27	187.52
1996	101.73	109.12	107.12	141.62	109.31	103.55
1997	113.09	126.27	124.37	125.41	136.23	102.64
1998	87.15	116.27	93.90	120.42	109.18	111.61
1999	89.34	183.18	96.39	121.41	180.64	190.64
2000	140.49	132.41	161.47	181.11	89.07	229.95
2001	256.80	108.35	277.68	200.44	96.99	102.90
2002	250.57	157.83	250.35	248.13	126.07	198.90
2003	229.13	132.27	213.35	200.15	118.00	118.00
2004	145.21	105.84	152.05	183.93	76.93	112.20
2005	135.47	140.93	154.70	222.15	124.00	134.20
2006	160.98	185.56	161.07	224.97	131.45	267.00
2007	254.94	273.54	287.17	276.17	263.90	263.90

A.5. ANEXO E: PRECIPITACIÓN MÁXIMA ANUAL EN MM, POR MUNICIPIO

Año	Emiliano Z.	Huimanguillo	Jalapa	Jalpa	Jonuta	Macuspana
1961	127.25	189.25	122.56	93.30	121.26	129.30
1962	162.44	137.28	124.21	71.42	123.55	98.45
1963	137.41	157.34	187.82	170.98	125.58	146.58
1964	153.09	208.76	158.81	120.80	129.13	123.54
1965	134.87	224.50	148.00	160.73	112.51	100.58
1966	209.29	130.83	102.02	93.81	159.32	90.91
1967	201.22	354.59	232.16	210.05	194.70	192.00
1968	150.37	148.72	194.94	132.86	142.80	157.42
1969	121.16	159.74	114.69	103.20	114.83	121.07
1970	114.25	215.63	133.92	124.12	106.21	114.86
1971	163.43	108.27	124.28	134.24	134.21	137.71
1972	158.73	180.04	157.63	169.10	136.97	145.36
1973	197.42	174.07	183.41	109.01	190.58	179.48
1974	241.55	271.50	172.75	165.71	191.19	160.12
1975	141.56	223.90	184.56	181.36	140.98	189.16
1976	176.30	283.62	164.95	216.72	178.10	195.40
1977	170.83	148.55	158.44	139.07	174.00	152.11
1978	140.98	216.41	193.02	135.72	134.40	141.57
1979	133.55	167.10	139.09	149.33	121.58	150.06
1980	165.43	260.22	218.81	219.26	161.02	276.26
1981	110.09	143.93	160.44	148.58	97.93	127.52
1982	131.34	176.93	119.73	104.55	107.50	114.78
1983	112.54	92.58	126.28	116.49	92.39	70.09
1984	125.57	148.81	134.95	115.04	122.09	152.39
1985	109.59	114.16	104.92	99.99	132.84	114.56
1986	143.47	131.57	138.91	119.33	128.44	90.10
1987	100.47	130.51	159.07	119.22	95.18	108.56
1988	163.41	124.93	177.68	259.07	136.78	105.10
1989	139.51	171.47	161.10	105.46	116.80	116.76
1990	153.47	329.87	303.98	152.63	128.77	148.69
1991	87.95	202.65	156.34	159.18	85.20	128.95
1992	162.97	151.71	204.28	207.36	139.15	123.04
1993	131.53	214.99	144.09	111.84	109.86	95.42
1994	161.66	156.73	143.96	138.56	128.93	109.70
1995	163.46	172.00	153.66	146.37	149.13	135.75
1996	116.55	138.75	154.33	135.31	102.54	112.06
1997	140.47	144.49	125.63	160.41	119.72	84.84
1998	104.66	179.23	123.22	119.88	89.39	88.55
1999	106.24	180.15	109.99	122.76	96.23	98.67
2000	179.75	102.70	181.93	144.92	153.33	151.16
2001	311.50	218.70	239.43	192.02	249.46	166.18
2002	265.00	250.00	246.13	261.84	254.02	234.63
2003	244.00	324.50	166.07	117.05	196.75	99.21
2004	166.00	300.50	127.04	139.62	145.27	109.86
2005	150.40	300.00	157.85	163.53	181.70	219.06
2006	175.70	330.00	206.25	267.00	151.35	116.79
2007	317.00	403.40	264.74	250.67	278.52	227.90

A.5. ANEXO E: PRECIPITACIÓN MÁXIMA ANUAL EN MM, POR MUNICIPIO

Año	Nacajuca	Paraíso	Tacotalpa	Teapa	Tenosique
1961	90.55	100.25	168.82	166.75	123.30
1962	71.48	122.98	134.09	130.87	71.90
1963	171.36	147.25	231.23	223.75	82.25
1964	120.87	145.79	190.37	190.44	73.41
1965	161.72	146.29	242.98	233.76	86.02
1966	93.68	120.11	145.38	141.33	209.29
1967	211.40	184.53	317.50	310.94	156.14
1968	134.11	127.80	311.95	300.75	114.35
1969	103.74	112.59	123.35	122.89	116.55
1970	121.65	141.78	176.37	173.45	66.64
1971	136.89	123.50	121.41	118.17	119.60
1972	168.86	158.62	163.78	164.95	158.73
1973	108.79	136.85	229.67	228.27	197.42
1974	166.54	171.53	220.71	221.32	66.26
1975	180.90	171.28	215.88	214.74	66.08
1976	213.92	219.89	243.36	238.96	88.97
1977	139.16	144.34	216.61	214.59	118.15
1978	136.63	123.50	212.40	213.02	103.19
1979	151.50	142.91	195.05	191.41	123.51
1980	218.80	193.18	174.54	173.65	84.63
1981	151.60	121.11	187.52	187.16	105.89
1982	104.80	108.58	121.34	125.77	117.80
1983	118.54	101.24	157.23	156.71	112.54
1984	115.57	124.65	155.71	151.11	125.57
1985	98.53	96.22	118.24	120.67	61.16
1986	120.90	109.85	238.60	231.65	143.47
1987	119.63	107.18	164.61	160.40	87.48
1988	264.16	191.44	138.62	140.00	120.04
1989	105.58	106.82	177.83	174.90	76.53
1990	155.59	131.47	359.71	457.52	142.00
1991	159.55	143.60	159.03	161.29	87.22
1992	211.05	177.61	266.25	263.85	95.83
1993	110.00	107.68	155.37	152.03	94.90
1994	139.31	128.96	102.58	103.24	92.00
1995	144.30	145.00	203.87	196.30	137.18
1996	136.91	122.16	131.31	130.37	91.96
1997	163.31	146.08	162.89	161.05	63.37
1998	120.30	110.82	94.96	93.11	63.49
1999	119.35	129.79	85.34	85.27	80.25
2000	140.65	134.03	132.22	131.92	60.32
2001	196.50	190.80	311.50	309.73	119.57
2002	265.00	225.82	265.00	264.79	298.50
2003	117.00	141.35	244.00	249.37	221.00
2004	141.00	127.40	166.00	180.31	175.83
2005	165.00	151.81	70.00	85.35	134.00
2006	267.00	206.63	237.80	242.73	119.75
2007	250.00	262.94	317.00	321.74	287.80

A.6. Anexo F: Subcuencas para cada municipio

En el siguiente anexo se muestran las subcuencas que influyen en el análisis de cada municipio de Tabasco, y sobre todo los nodos que entran en el estudio de dicho municipio. La figura A.27 tiene el propósito de ejemplificar la información que se encuentra contenida en cada una de las tablas, las subcuencas que se encuentran en verde son las que conforman al municipio y los nodos en color rosa son los que se consideraron.

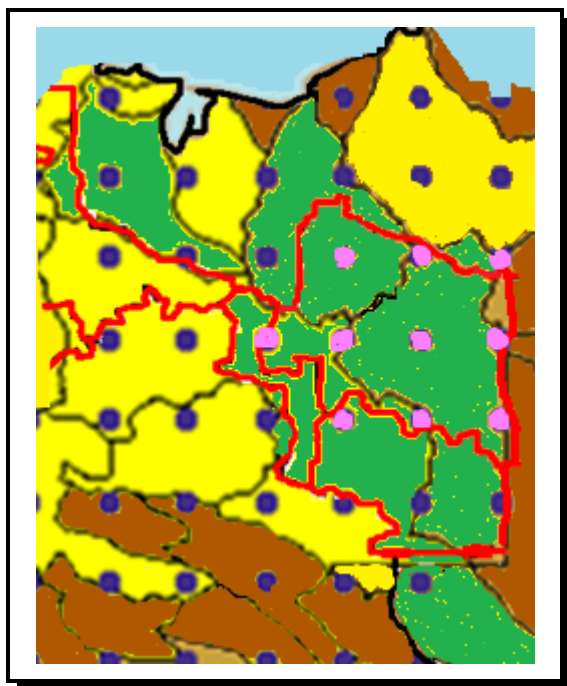


Figura A.27: Nodos de interés en el municipio de Balancán.

Balancán		
Subcuenca	Aporta su caudal	Nº Nodos
San Pedro (69)	A la cuenca 73	5
Usumacinta (73)	A la cuenca 74, 75 y 76	4
Cumpan (80)	Al mar	1

Cárdenas		
Subcuenca	Aporta su caudal	Nº Nodos
Santa Anita (7)	Al mar	1
Laguna del Carmen (8)	Al mar	3
Laguna Machona (9)	Al mar	1
Caxcuchapa (37)	Al mar	0

Centla		
Subcuenca	Aporta su caudal	Nº Nodos
Chilapa (47)	A la cuenca 74	1
Usumacinta (73)	A la cuenca 74, 75 y 76	2
Grijalva (74)	Al mar	1
San Pedro y San Pablo (76)	Al mar	1

Centro		
Subcuenca	Aporta su caudal	N° Nodos
Tabasquillo (34)	A la cuenca 74	1
Chilapilla (49)	A la cuenca 74	1
De la sierra (50)	A la cuenca 74	1
Grijalva (74)	Al mar	3

Comalcalco		
Subcuenca	Aporta su caudal	N° Nodos
Laguna Machona (9)	Al mar	1
Caxcuchapa (37)	Al mar	1

Cunduacán		
Subcuenca	Aporta su caudal	N° Nodos
Cunduacán (35)	A la cuenca 36	0
Samaria (36)	Al mar	2
Caxcuchapa (37)	Al mar	0

Emiliano Zapata		
Subcuenca	Aporta su caudal	N° Nodos
Chacamax (72)	A la cuenca 73	1
Usumacinta (73)	A la cuenca 74, 75 y 76	2

Huimanguillo		
Subcuenca	Aporta su caudal	N° Nodos
Zanapa (1)	A la cuenca 4	2
Tancochapa Alto (2)	A la cuenca 5	2
Coacajapa (4)	A la cuenca 6	2
Tancochapa Bajo (5)	A la cuenca 6	2
Tonalá (6)	Al mar	3
Paredón (30)	A la cuenca 32	2
Mezcalapa (32)	A las cuencas 33 y 36	2

Jalapa		
Subcuenca	Aporta su caudal	N° Nodos
Macuxpana (45)	A la cuenca 47	1
Tacotalpa (48)	A la cuenca 74	2
Chilapilla (49)	A la cuenca 74	0
Grijalva (74)	Al mar	1

Jalpa de Méndez		
Subcuenca	Aporta su caudal	N° Nodos
Cunduacán (35)	A la cuenca 36	0
Samaria (36)	Al mar	2

Jonuta		
Subcuenca	Aporta su caudal	N° Nodos
Chilapa (47)	A la cuenca 74	1
Usumacinta (73)	A la cuenca 74, 75 y 76	4
Laguna del Pom y Atasta (81)	Al mar	1

Macuspana		
Subcuenca	Aporta su caudal	N° Nodos
Tulija (44)	A la cuenca 47	1
Macuspana (45)	A la cuenca 47	1
Chilapa (47)	A la cuenca 74	4
Chilapilla (49)	A la cuenca 74	2

Nacajuca		
Subcuenca	Aporta su caudal	N° Nodos
El carrizal (33)	A la cuenca 34	1
Cunduacán (36)	Al mar	1
Grijalva (74)	Al mar	1

Paraíso		
Subcuenca	Aporta su caudal	N° Nodos
Caxcuchapa (37)	Al mar	2

Tacotalpa		
Subcuenca	Aporta su caudal	N° Nodos
Puxcatán (41)	A la cuenca 45	1
Almendro (46)	A la cuenca 48	1
Tacotalpa (48)	A la cuenca 74	2

Teapa		
Subcuenca	Aporta su caudal	N° Nodos
Tacotalpa (48)	A la cuenca 74	1
De la sierra (50)	A la cuenca 74	2
Pichucalco (51)	A la cuenca 74	0

Tenosique		
Subcuenca	Aporta su caudal	N° Nodos
San Pedro (69)	A la cuenca 73	3
Usumacinta (73)	A la cuenca 74, 75 y 76	3

A.7. Anexo G: Programas para la estimación del tamaño y potencia de la prueba

En este anexo se muestran los programas para la estimación del tamaño de la prueba bootstrap y la potencia para las tres pruebas dadas en el capítulo 2. Para que este programa función es importante cargar las siguientes dos librerías: `evir` y `evd`, para generar la muestra aleatoria de VEG, fréchet y gumbel, además de que la librería `evir` ayuda en la estimación de los parámetros de la `dveg`.

```
# Función para obtener el coeficiente de correlación muestral #
corm <- function(n,l,ee,f)
{
  #m0 <- rgev(n,l,ee,f)
  m0 <- Maxim
  p0 <- fgev(m0, std.err = FALSE)$estimate
  z0 <- sort(m0)
  f0 <- ((1:n)-0.5)/n
  y0 <- (p0[2]/p0[3])*((-log(f0))^-p0[3]-1)+p0[1]
  r0 <- cor(y0,z0)
  return(c(r0,p0))
}

# Función para obtener el estadístico de la prueba
# bootstrap basada en el coeficiente de correlación #
correst <- function(n,B,p,l,ee)
{
  rb <- rep(NA,B)
  i <- 1
  while (i<=B)
  {
    mb <- rgev(n,l,ee,p)
    pb <- fgev(mb,std.err = FALSE)$estimate
    z <- sort(mb)
    f1 <- ((1:n)-0.5)/n
    y <- (pb[2]/pb[3])*((-log(f1))^-pb[3]-1)+pb[1]
    rb[i] <- cor(y,z)
    i <- i+1
  }
  orb <- sort(rb)
  return(orb)
}

# Función para obtener el tamaño y potencia de la prueba bootstrap #
tamprueba <- function(n,f,B,M,l,ee)
{
  alfa <- c(0.01,0.025,0.05,0.075,0.10)
  contador <- rep(0,length(alfa))
  j <- 1
  while(j<=M)
```

```

{
  d <- corm(n,l,ee,f) # para el tamaño de prueba se usa rgev y para
  d1 <- d[4]          # la potencia se cambia por la distribución alternativa.
  r0 <- d[1]
  rc <- correst(n,B,d1,l,ee)
  k <- 1
  while(k<=length(alfa))
  {
    cuantil <- quantile(rc,alfa[k])
    if (r0<=cuantil){ contador[k] <- contador[k]+1 }
    else { contador[k] <- contador[k]}
    k<-k+1
  }
  print(j)
  j<-j+1
}
return(contador/M)
}

```

A.8. Anexo H: Programas para la estimación de parámetros y gráficas

El siguiente programa ayuda a obtener la estimación puntual y por intervalos de los parámetros de la distribución de valores extremos generalizada y la obtención de los periodos de retorno. Es necesaria la librería `evir`.

```

##### Cargar los datos #####
años<-read.table("C://Users/Ana Laura
/Desktop/Proyecto/Datos-subcuencas/maxparaiso.txt",header=T)
colnames(años) <- c("Año","Maximos")
Maxim <- años[,2]
fig1 <- plot(años,pch=19,xlab="Año",ylab="Lluvia máxima por año (mm)")
hist(Maxim,freq=FALSE,ylim=c(0,.015))

#####
# ESTIMACIÓN DE PARAMETROS E INTERVALOS DE CONFIANZA DE LA DVEG #

estim.veg<- function(x)
{
  n <- length(x)
  veg<-fgev(x)
  p<-veg$estimate ## estimadores de la dveg
  k<-p[3]        ##parámetro de forma
  a <- p[2]      ## parámetro de escala
  mu <- p[1]     ## parámetro de localidad

  veroveg<- -n*log(a)-(1+1/k)*

```

```

sum(log(1+k*((x-mu)/a)))-sum((1+k*((x-mu)/a))^(1/k))
aic<- -2*veroveg+2*3
valor <- (c(k,a,mu,veroveg,aic))
aux12 <- function(x){k<-p[3]; a <- p[2];
mu <- p[1]; (1/a)*(1+k*((x-mu)/a))^(1/k-1)*exp(-(1+k*((x-mu)/a))^(1/k)) }
plot(aux12,type = "l",col = 2,lty = 1,lwd = 2,main=" ",
xlab="", ylab=" ",xlim=c(0,700),ylim=c(0,1),add=TRUE)

### Estimación IC #####
ee<-veg$std.err      #Errores estandar
z<-qnorm(0.975)
a1 <- p[2]-(z*ee[2])
a2 <- p[2]+(z*ee[2]) # ESCALA
k1 <- p[3]-(z*ee[3])
k2 <- p[3]+(z*ee[3]) # FORMA
mu1 <- p[1]-(z*ee[1])
mu2 <- p[1]+(z*ee[1]) # LOCALIDAD
cat("Escala=",a,"IC=",a1,a2, "\n")
cat("Forma=",k,"IC=",k1,k2, "\n")
cat("Localidad=", mu,"IC=",mu1,mu2, "\n")
cat("logverosimilitud=",veroveg,"AIC=",aic, "\n")
}

##### PERIODOS DE RETORNO #####

prveg <- function(x1)
{
  x <-Maxim
  veg <-fgev(x)
  p <-veg$estimate
  a <- p[2] #p[2]  ## ESCALA
  b <- p[3] #p[1]  ## FORMA
  mu <- p[1] #p[3]  ## LOCALIDAD
  pr <- pgev(x1,mu,a,b)
  1/(1-pr)
}

salidafinal <- matrix(NA,1,4)
salidafinal[1,] <-c(prveg(100),prveg(150),prveg(200),prveg(250))
salidafinal

```

Finalmente con este programa se obtiene la estimación puntual y por intervalos de los parámetros de la distribución Gumbel y la obtención de los periodos de retorno. Es necesaria la librería VGAM.

```

estim.gumbel<- function(x)
{
  n <- length(x)

```

```

gumbel<-vglm(x~1,gumbel)
p<-coef(gumbel, mat=T)    ## estimadores de la dgumbel
a <- exp(p[2])           ## parámetro de escala
mu <- p[1]                ## parámetro de localidad
verogumbel<- -n*log(a)-sum(x-mu)/a -sum( exp(-(x-mu)/a))
aic<- -2*verogumbel+2*2
aux12 <- function(x){a <- exp(p[2]);
mu <- p[1]; exp(-(x-mu)/a - exp(- (x-mu)/a ))/a }
plot(aux12,type = "l",col = 4,lty = 1,lwd = 2,main=" ",
xlab="", ylab=" ",xlim=c(0,700),ylim=c(0,1),add=TRUE)

##### Estimación de IC #####
ee<-sqrt(diag(vcov(summary(gumbel)))) #Errores estandar
z<-qnorm(0.975)
a1 <- p[2]-(z*exp(ee[2]))
a2 <- p[2]+(z*exp(ee[2]))    ## ESCALA
mu1 <- p[1]-(z*ee[1])
mu2 <- p[1]+(z*ee[1])      ## LOCALIDAD

cat("Escala=",a,"IC=",a1,a2, "\n")
cat("Localidad=", mu,"IC=",mu1,mu2, "\n")
cat("logverosimilitud=",verogumbel,"AIC=",aic,"\n")
}

### PERIODO DE RETORNO DGUMBEL #####
prgumbel <- function(x1)
{ x<-Maxim
gumbel<-vglm(x~1,gumbel)
p<-coef(gumbel)
a <- exp(p[2])    ## ESCALA
mu <- p[1]    ## LOCALIDAD
pr <- pgumbel(x1,mu,a)
1/(1-pr)
}
salidafinal <- matrix(NA,1,4)
salidafinal[1,] <-c(prgumbel(100),prgumbel(150),prgumbel(200),prgumbel(250))
salidafinal

```