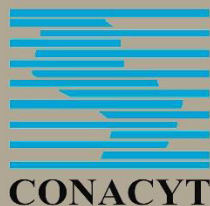




ATLAS DE SEQUÍAS EN PARAGUAY BASADO EN EL ANÁLISIS REGIONAL DE L-MOMENTOS

Ana Violeta González Santander



Atlas de sequias en Paraguay basado en el Análisis Regional de L-Momentos

Investigación para el Desarrollo

Julio del 2016

www.proyectoclima.com.py

Consejo Nacional de Ciencia y Tecnología - CONACYT

Programa Paraguayo para el Desarrollo de la Ciencia y Tecnología – PROCIENCIA

Proyecto de iniciación científica: 14-INV-338

Asesor del proyecto: Rossana Scribano

Tesista: Ana Violeta González Santander

Asesor Informático: Alcides Domínguez

Investigación para el Desarrollo

Presidente Cesar Cabello

Director Ejecutivo Bruno Osmar Martínez

Gerente de proyectos: Karina Godoy

Directora de Área: Rossana Scribano

Agradecimientos

Quiero agradecer a las instituciones que hicieron posible realizar este trabajo; Prociencia; Programa paraguayo para el desarrollo de la ciencia y tecnología del CONACYT, Investigación para el Desarrollo; que me abrió sus puertas y me permitió desarrollar mi tesis, Dirección de Meteorología e Hidrología de la Dirección de Aeronáutica Civil; que me facilitó todos los datos meteorológicos históricos que fueron el insumo para la realización de este trabajo.

También un reconocimiento a la Facultad Politécnica de la Universidad Nacional de Asunción; que fue la institución donde me formo como profesional de la Ciencias de la Atmósfera.

Por último un agradecimiento a todos mis profesores, ya que ellos me enseñaron a valorar los estudios y a superarme cada día, depositando su fe en mí.

ÍNDICE

Resumen.....	7
1. Introducción	8
1.1. Planteamiento del Problema	10
1.2. Objetivos	10
1.2.1. Objetivo General	10
1.2.2. Objetivos Específicos.....	11
1.3. Justificación	11
1.4. Importancia	12
2. Marco Teórico	13
2.1. Aspectos Geográficos del Paraguay	13
2.1.1. Régimen de Precipitación en Paraguay.....	13
2.2. Sequía.....	13
2.3. Hidrología Probabilística	17
2.4. Introducción a los L-momentos.....	20
2.5. Análisis Regional de Frecuencia	22
2.6. Los Cuantiles.....	24
2.7. Región Homogénea	26
3. Control de calidad de los datos de precipitación	27
3.1. Grupos de tests de control de calidad	27
3.1.1. Implementación de los tests de control de calidad	29
3.2. RCLIMDEX_EXTRAQC.R.	30
3.3. ANDESQC.....	32
3.4. RClimTool	34
4. Metodología	38
4.1. Población y muestra.....	38
4.2. Ajuste de los datos de precipitación	40
4.3. Identificación regiones homogéneas	43
4.4. Determinación del modelo de probabilidad de mejor ajuste.....	44
4.5. Determinación de cuantiles	45
4.6. Mapeo	45
5. Resultados.....	47

5.1.	Procedimientos de regionalización	47
5.2.	Determinación del modelo de probabilidad de mejor ajuste	51
5.3.	Cuantiles.....	52
5.4.	Mapeo	54
5.4.1.	Período de retorno de eventos de sequía.....	55
6.	Conclusiones.....	64
6.1.	Recomendaciones para trabajos futuros	69
7.	ANEXO	70
7.1.	ANEXO I: Formato De Entrada De Los Datos.....	70
7.1.1.	Control de calidad	70
7.1.2.	Análisis Regional de Frecuencia basados en L-momentos.....	73
7.2.	ANEXO II: GRÁFICAS De Los Resultados	76
7.2.1.	Control de calidad	76
7.2.2.	Análisis Regional de Frecuencia basados en L-momento	87
8.	Referencias.....	98
9.	Glosario	100

Resumen

La sequía es un evento extremo que afecta a Paraguay, tiene efectos perjudiciales para las personas, animales y cultivos, ya que el agua es el elemento vital para el desarrollo de las actividades más destacadas de la humanidad.

Se obtuvo un atlas de los eventos de sequía que se produjeron en Paraguay a lo largo de los datos históricos de precipitación acumulada diaria, desde el inicio de cada registro hasta el año 2013; de los datos de las estaciones meteorológicas, de la Dirección de Meteorología e Hidrología. Se realizó previamente un control de calidad de los datos, estos se desarrollaron en R (lenguaje y entorno de programación para análisis estadístico y gráfico). Luego se procedió al análisis de los datos de dos maneras consecutivas: la identificación de zonas homogéneas y la determinación de cuantiles, así también se analizó la probabilidad de ocurrencia de sequía. Por último se procedió al análisis de la precipitación para la realización del mapeo para El Atlas de Sequía; esto se realizó por los métodos estadísticos ARF (Análisis Regional de Frecuencias) y L-momentos. El Análisis Regional de Frecuencias se utilizó para la homogenización de las zonas de estudio, ya que esto permite la utilización de los mismos sin la necesidad de rellenar los datos faltantes en el periodo de estudio. Todas las rutinas requeridas por el Análisis Regional de Frecuencia basado en L-momentos se desarrollaron en R.

1. Introducción

Un evento extremo que afecta a la mayoría de las regiones del mundo es la sequía, que pueden perturbar la vida cotidiana de las personas, ya que tiene un gran impacto en la economía de la región afectada y, por ende se extiende a nivel social.

Para poder mitigar los efectos provocados por este evento es necesario su estudio y análisis. Es muy importante el estudio sobre los eventos de sequía, para conocer su comportamiento temporal y su duración. Hay que tener en cuenta que podemos conocer o tener una idea, pero no certeza, de cuándo comienza una sequía, y de la misma forma cuándo culmina. No solamente la precipitación es la variable que podemos utilizar para caracterizar a la sequía, según el lugar podemos tener en cuenta algunas variables más, como por ejemplo: la temperatura, el viento y la humedad relativa.

Una definición de la sequía según la National Oceanic and Atmospheric Administration (NOAA, 2006), es que “La sequía es una deficiencia en la precipitación sobre un período prolongado, por lo general una temporada o más, lo que resulta en una escasez de agua causando impactos adversos sobre la vegetación, los animales y/o personas; es una aberración temporal de las condiciones climáticas normales, por lo que puede variar significativamente de una región a otra”.

El Vocabulario Meteorológico Internacional de la Organización Meteorológica Mundial (OMM, 1966) dice que la sequía es la “ausencia prolongada o déficit notable de precipitación”.

También la Organización Meteorológica Mundial (WCP, 1986) a petición del Programa de las Naciones Unidas para el Desarrollo, proporcionó el siguiente criterio “se dice que hay sequía en una región si la precipitación anual es inferior al 60% de la normal durante más de dos años consecutivos en más de 50% de la superficie de la región”.

Para el ajuste de los modelos de probabilidad de los datos observados en este caso la sequía, se han utilizado por muchos años los métodos convencionales, por ejemplo los

más conocidos son el método de los momentos y el método de máxima verosimilitud, pero en lugares con cortos registros de datos y valores atípicos, ellos presentan problemas de sesgo a la hora de estimar los periodos de retorno y los cuantiles.

Por eso es importante poder estudiar otras metodologías como el Análisis Regional de Frecuencia basado en L-Momentos para su validación y su difusión en la toma de decisiones en eventos extremos como la sequía.

Un evento de sequía puede ser devastador dependiendo de su duración y su intensidad, provocando daños tanto a nivel social como económico. Es difícil saber cuándo exactamente comienza un evento de sequía, ya que nos percatamos de su presencia cuando los efectos son evidentes. Los sectores más afectados son aquellos que se dedican al cultivo y la ganadería. Paraguay se caracteriza por su producción agrícola, que se ve perjudicada por eventos, como la sequía.

Existen varios métodos para calcular la ocurrencia de la sequía, los más conocidos son los índices. En el Paraguay, la investigación sobre sequía es aún incipiente. Existen sin embargo, trabajos que abordan el tema de manera muy superficial.

El método del Análisis Regional de Frecuencia basado en L-momentos ha sido desarrollado en otros países, como por ejemplo México. En Sudamérica también se está implementando, como es el caso de Chile donde dio buenos resultados.

La sequía es uno de los eventos extremos que afecta a Paraguay, tiene efectos perjudiciales para diversos sectores del país, ya que el agua es el elemento vital para el desarrollo de todas las actividades de la humanidad.

Por ese motivo se pretende conocer con mayor detalle el comportamiento de las sequías; ya sea en su duración y severidad, para un posterior análisis de las zonas de ocurrencia del evento; con esto se podría prever y mitigar los daños que ocasionan las sequías en las diferentes actividades relacionadas con el agua.

Por otro lado, con el resultado que nos proporcione la base de datos analizada, se pretende transferir los conocimientos adquiridos con rigor científico especialmente para

el sector agropecuario, que podrían servir como base para futuras investigaciones en el área.

Los L-momentos son un sistema alternativo para describir las “formas” de las distribuciones de probabilidad. Desde el punto de vista estadístico, son una combinación lineal de los denominados “momentos ponderados por probabilidad” o “momentos de probabilidad pesada” que fueron sugeridos por Greenwood et al (1979).

El Análisis Regional de Frecuencias es un procedimiento consistente en agregar la información proveniente de varias muestras, en una sola gran muestra, asumiendo previamente que todas éstas provienen de un mismo modelo de probabilidad, excepto por una diferencia entre ellas debida a un factor de escala.

1.1. Planteamiento del Problema

- ¿Cuál es el nivel de los datos meteorológicos pertenecientes a la Dirección de Meteorología e Higrología del Paraguay?
- ¿Podría usarse el método de Análisis Regional de Frecuencia, basados en L-momentos, para la evaluación de la frecuencia de eventos de sequía en Paraguay?
- ¿Cuál sería la frecuencia de los eventos de sequía y los periodos de retorno para diferentes umbrales en el Paraguay?

1.2. Objetivos

1.2.1. Objetivo General

- Analizar la sequía en Paraguay mediante el Análisis Regional de Frecuencia basados en L-momentos para los datos históricos de precipitación.

1.2.2. Objetivos Específicos

- ✓ Verificar que el método de Análisis Regional de Frecuencia basados en L-momentos cumple con las necesidades señaladas, en su aplicación en la evaluación de frecuencia de eventos de sequía en Paraguay.
- ✓ Realizar un control de calidad de los datos meteorológicos pertenecientes a la Dirección de Meteorología e Higrología.
- ✓ Obtener mapas donde se reflejen los periodos de ocurrencia y frecuencia de los eventos de sequía.
- ✓ Relacionar la magnitud y frecuencia de las sequías que ocurrieron en el periodo de análisis con el método estadístico seleccionado.

1.3. Justificación

Un evento de sequía puede ser devastador dependiendo de su duración y su intensidad, provocando daños tanto a nivel social como económico. Es difícil saber cuándo exactamente comienza un evento de sequía, ya que nos percatamos de su presencia cuando los efectos son evidentes. Los sectores más afectados son aquellos que se dedican al cultivo y la ganadería. Paraguay se caracteriza por su producción agropecuaria, que es perjudicado por eventos como la sequía.

Existen varios métodos para calcular la ocurrencia de la sequía, los más conocidos son los índices. En el Paraguay, la investigación sobre sequía es aún incipiente. Existen sin embargo, varios trabajos que abordan el problema pero muy superficialmente.

El método del Análisis Regional de Frecuencia basado en L-momentos ha sido desarrollado en otros países, como por ejemplo México. En Sudamérica también se está implementando, como es el caso de Chile donde dio buenos resultados.

La sequía es uno de los eventos extremos que afecta a Paraguay, tiene efectos perjudiciales para diversos sectores del país, ya que el agua es el elemento vital para el desarrollo de todas las actividades humanas.

Por ese motivo se pretende conocer con mayor detalle el comportamiento de las sequías; ya sea en su duración y severidad, para un posterior análisis de las zonas de

ocurrencia del evento; con esto se podría prever y mitigar los daños que ocasionan las sequías en las diferentes actividades relacionadas con el agua.

Por otro lado, con el resultado que nos proporcione la base de datos analizados, se pretende transferir los conocimientos adquiridos especialmente para el sector agropecuario, que podrían servir como base para futuras investigaciones en el área.

1.4. Importancia

El clima en el Paraguay, tiene una característica bien diferenciada entre las dos regiones que la componen; teniendo en la Región Occidental una zona más árida, que presenta una estación más seca; y una zona más húmeda como lo es la Región Oriental. Esta condición se ve reflejada en los mapas de frecuencia y de periodo de retorno.

Con los resultados obtenidos con el Análisis Regional de Frecuencia basados en L-momentos, podremos adquirir un panorama más claro para la evaluación, estimación y tiempo de ocurrencia de eventos de sequía en el Paraguay. Teniendo en cuenta que es de los primeros trabajos realizados con el método Análisis Regional de Frecuencia Basado En L-Momentos en el país; abriendo un espacio para el análisis de las zonas de ocurrencia de este fenómeno poco estudiado localmente.

2. Marco Teórico

2.1. Aspectos Geográficos del Paraguay

Como se menciona en el trabajo *Un análisis del comportamiento de la precipitación en el Paraguay*: “Por su situación geográfica, el Paraguay se encuentra ubicado en el borde occidental del anticiclón subtropical del atlántico y este sistema de presión es en gran medida el responsable del transporte de aire cálido y húmedo desde el océano Atlántico tropical, pasando por el pantanal del Mato Grosso, hacia el Paraguay, por consiguiente el régimen de viento dominante es del cuadrante nordeste; siendo frecuentemente perturbado por vientos del sur producto de la incursión de frentes fríos procedentes del extremo austral de Sudamérica, estos frentes fríos de origen polar, irrumpen con frecuencia durante el invierno, por ello se observa una onda simple en la temperatura del aire, con mínimo en invierno y máximo en verano” (Grassi 2005).

2.1.1. Régimen de Precipitación en Paraguay

En el mismo trabajo se aclara que: “La precipitación en el Paraguay presenta un ciclo anual con máximos en verano y mínimos en invierno, empieza a incrementarse desde los inicios de la primavera y declina hacia finales del otoño. En la región Occidental, el mínimo de precipitación es claramente observado durante los meses de julio y agosto, cuando la precipitación se vuelve muy escasa, y el máximo es claramente observado entre los meses de diciembre y enero. En la región Oriental, especialmente en el sudeste, si bien se observan también la misma variación estacional, no es tan pronunciada como en la región Occidental” (Grassi 2005).

2.2. Sequía

Como lo describió Wilhite (2000) los grandes sectores afectados por la sequía, su distribución temporal, su diversidad geográfica, y el incremento de la demanda de la reserva del agua por el sistema hecho por el hombre hace difícil de desarrollar una simple y universal definición de la sequía.

De acuerdo a la OMM (Organización Meteorológica Mundial), en su informe OMM-N° 1006, la sequía puede ser catalogada en cuatro formas: meteorológica, agrícola, hidrológica y socioeconómica.

La sequía meteorológica es detectada a partir de un umbral de déficit de precipitación que se alcanza durante un período de tiempo previamente determinado. El umbral escogido (por ejemplo, un 75% de la precipitación normal) y el período de duración (por ejemplo, seis meses) variarán según el lugar y en función de las necesidades de los usuarios y de sus actividades. La sequía meteorológica es un fenómeno natural que responde a muy diversas causas según la región.

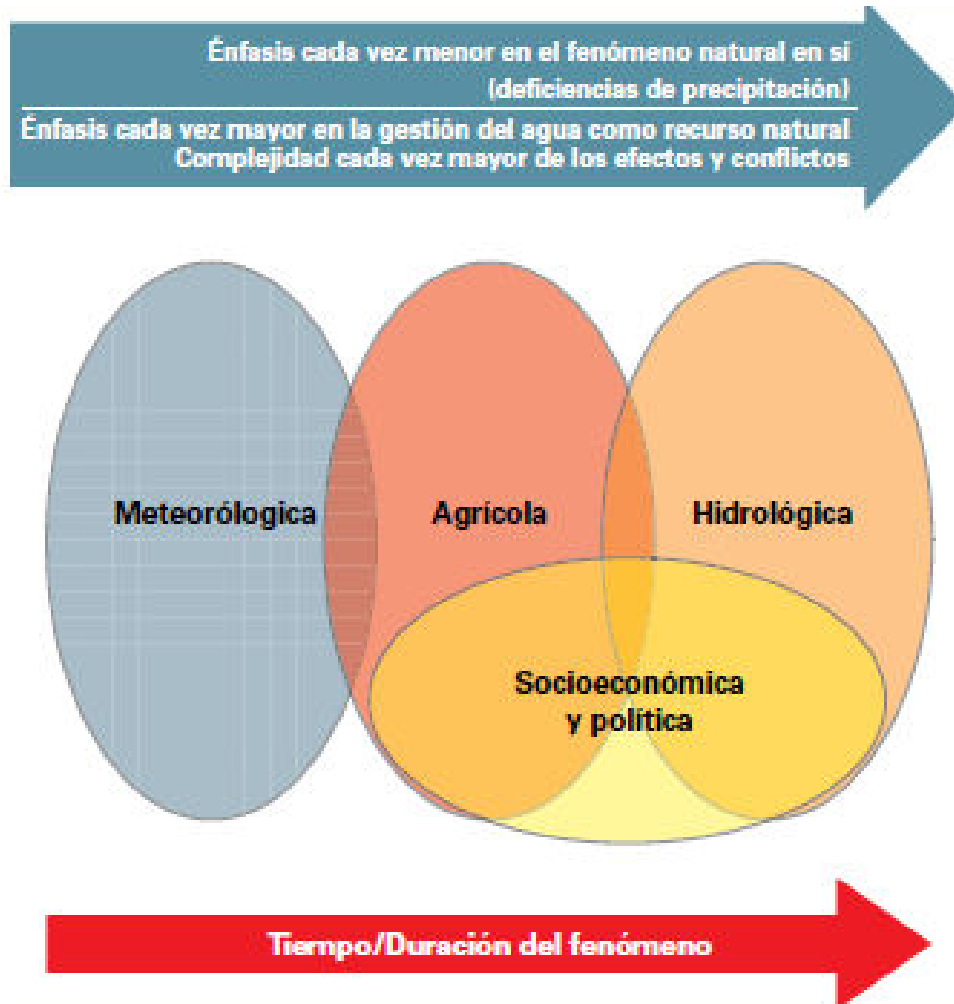
La sequía agrícola se define habitualmente en términos de disponibilidad de agua en los suelos para el sostenimiento de los cultivos y para su crecimiento y, menos habitualmente, como una desviación de los regímenes de precipitaciones normales durante cierto período de tiempo.

La sequía hidrológica se produce cuando las reservas de agua disponibles en fuentes como acuíferos, ríos, lagos y presas caen por debajo de la media estadística. La sequía hidrológica tiende a aparecer más lentamente porque se trata de agua almacenada que se utiliza pero no se repone. Aunque una sequía hidrológica suele ser provocada por una precipitación deficiente, también pueden tener otras causas.

La sequía socioeconómica se diferencia notablemente de los demás tipos de sequía porque refleja la relación entre la oferta y la demanda de mercancías básicas, como lo son el agua, los alimentos (para consumo humano como animal) o la energía hidroeléctrica, que dependen de las precipitaciones.

Las Sequías agrícolas, hidrológica y socioeconómica se dan en menor medida que la sequía meteorológica. Regularmente, transcurren varias semanas antes de que las deficiencias de precipitación comiencen a producir un déficit de humedad en el suelo que, a su vez, afectará negativamente a los cultivos y a los pastizales. La persistencia de un tiempo seco durante meses reduce el caudal de los ríos y el nivel de los embalses y lagos

y, potencialmente, el nivel de las aguas freáticas. Cuando la sequía se prolonga sobrevienen las sequías agrícola, hidrológica y socioeconómica, con sus correspondientes efectos.

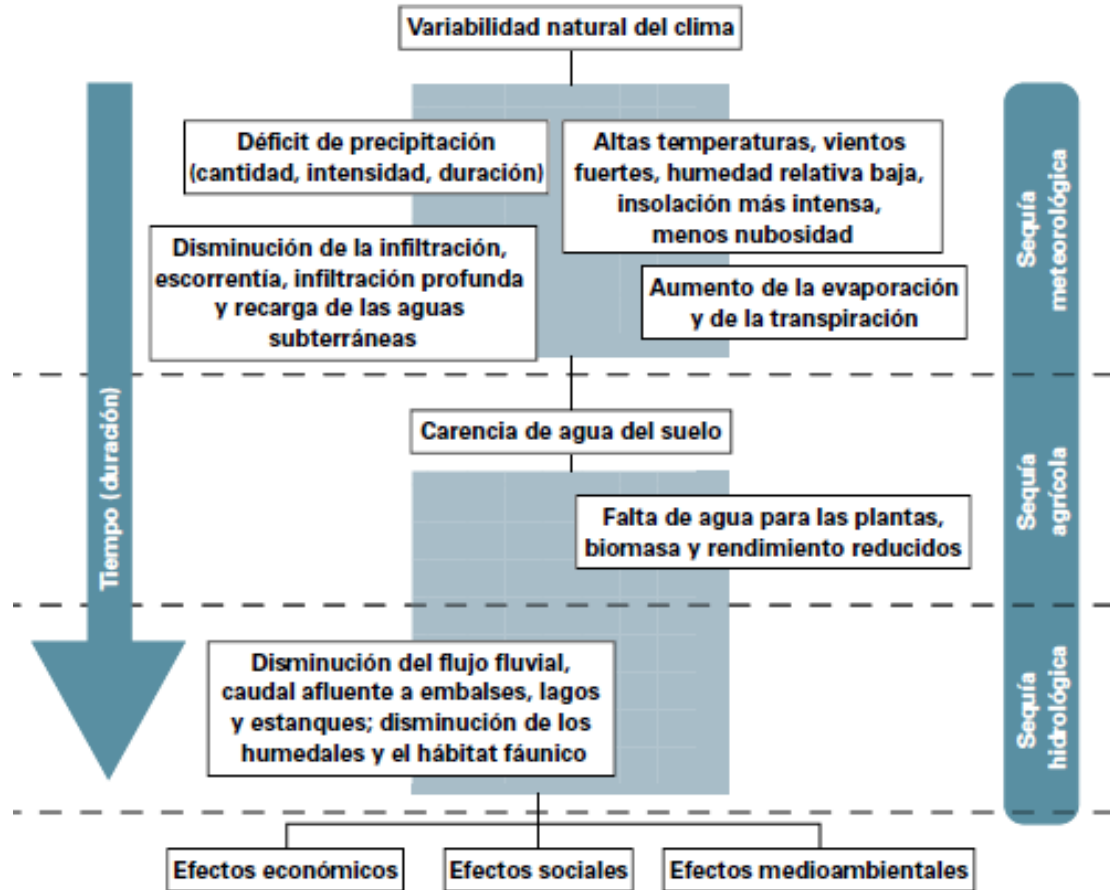


Gráfica 1. Interrelaciones entre las sequías meteorológica, agrícola, hidrológica y socioeconómica. (Fuente: Centro Nacional de Mitigación de Sequías, Universidad de Nebraska–Lincoln, Estados Unidos de América).

En el informe OMM-N° 1006, las sequías presentan tres rasgos distintivos:

- I. *La intensidad* reflejada en el déficit de precipitación y la gravedad de los efectos asociados a ese déficit. Su magnitud suele determinarse en términos de la desviación, respecto de las condiciones normales de los parámetros climáticos, tales como la precipitación o el nivel de los embalses, o de índices como, por ejemplo, el índice de precipitación normalizado.
- II. *La duración* de las sequías pueden sobrevenir rápidamente en algunas zonas climáticas pero, por lo general, tardan como mínimo dos o tres meses en hacerse patentes. Una vez comenzada, la sequía puede durar meses o años.
- III. *La magnitud* de los efectos de una sequía está estrechamente relacionada con el momento en que comienza la escasez de precipitación y con la intensidad y duración del fenómeno. Un invierno seco, por ejemplo, puede tener escasas repercusiones en muchos climas templados de latitudes medias, debido a la menor demanda de agua durante esos meses.

Todas las sequías son consecuencia de un déficit de precipitación o de sequía meteorológica, que a su vez puede provocar otros tipos de sequía y de efectos, eso se observa detalladamente en la gráfica 9.



Gráfica 2. Secuencia de sucesos de sequía y de sus efectos para tipos de sequías comúnmente aceptados. (Fuente: Centro Nacional de Mitigación de Sequías, Universidad de Nebraska-Lincoln, Estados Unidos de América).

2.3. Hidrología Probabilística

En el libro *Hidrología Aplicada*, Chow *et al* (1996), al iniciar el estudio de la estadística hidrológica, indica respecto al enfoque estocástico de los procesos hidrológicos, que “Este tipo de tratamiento es apropiado para observaciones de eventos hidrológicos extremos, como crecientes o sequías, y para información hidrológica promediada a lo largo de intervalos de tiempo grandes, como la precipitación anual”.

Teniendo en cuenta que nos interesa saber con qué frecuencia o cada cuántos años es posible esperar una sequía meteorológica tan severa, como aquella que consiste

en la precipitación anual menor o igual al 40% de la esperada en un año considerado normal. Esto requiere desde una aproximación hidrológico-probabilística, estimar el periodo de retorno de dicho evento especificado, basándose en la serie histórica de registros de la precipitación anual en el área de interés, los cuales consistirán en datos obtenidos en una estación meteorológica.

Definimos que “Q” es la magnitud del evento que ocurre en el tiempo y en el sitio específico, suponiendo que las observaciones que disponemos, se realizan en un intervalo de tiempo frecuente en el punto de interés. Indicamos que Q es una cantidad aleatoria (una variable aleatoria, que para nuestro caso, concerniría a la magnitud de la precipitación anual, que puede tomar un valor potencial entre 0 y el infinito). En el análisis de frecuencia estadística convencional, la distribución de frecuencia, es el conjunto fundamental, que especifica cuán frecuentemente ocurren los posibles valores de Q. Denotemos a continuación por $F(x)$, la probabilidad de que el valor actual de Q sea menor o igual a un determinado valor x:

$$F(x) = Prob(Q < x) \quad (1)$$

$F(x)$ corresponde a la función de distribución de probabilidad acumulada de la distribución de frecuencia. Por otro lado, la función inversa de $F(x)$ corresponde a $x(F)$, que es conocida como la función cuantílica de la distribución de frecuencia y expresa la magnitud de un determinado evento (la precipitación acumulada en un año cualquiera, en nuestro caso) en términos de su probabilidad de no excedencia F.

Especifiquemos el periodo de retorno (T), que tiene como definición el valor esperado del intervalo de recurrencia promedio o tiempo entre ocurrencias promedio entre eventos que exceden o igualan una magnitud especificada de la variable aleatoria X. Entonces, un cuantil con periodo de retorno T, QT , es un evento de magnitud tan extrema, que tiene una probabilidad $1/T$ de ser excedido por cualquier evento específico. Para un evento extremo alto, es decir, ubicado en la cola superior de la distribución de frecuencias, QT está dado por:

$$QT = x \left(1 - \frac{1}{T}\right) \quad (2)$$

$$F(QT) = \left(1 - \frac{1}{T}\right) \quad (3)$$

Mientras que para un evento extremo bajo, ubicado en la cola inferior de la distribución de frecuencias (caso de las sequías), las relaciones correspondientes son:

$$QT = x \left(\frac{1}{T}\right) \quad (4)$$

$$F(QT) = \left(\frac{1}{T}\right) \quad (5)$$

Una “regla del pulgar” (rule of thumb), aplicada al esquema regional, recomienda que uno debiese tener, al menos, 5 veces la cantidad de años que el periodo de retorno objetivo que quiere determinar (Reed, 1999).

Se admite, habitualmente que un cuantil con periodo de retorno T puede ser estimado de manera confiable a partir de un registro de longitud n , solamente si $T < n$. En la gran mayoría de las situaciones prácticas de las toma de decisiones basadas en datos anuales; esta condición casi nunca se satisface, debido, por un lado, a que habitualmente $n < 50$, y *por otro* lado, a que $T=100$ o incluso mayor (las sequías que habitualmente producen los mayores impactos económicos se dan cada 100 años, por ejemplo).

Establecer en primer lugar qué tipo de modelos permite caracterizar adecuadamente la función de distribución de probabilidad y por otro lado, cuál es el procedimiento más adecuado para determinar las características del modelo que mejor representa esa distribución de probabilidad a partir de la serie de datos disponibles.

En cuanto a los métodos para la estimación de los parámetros de la distribución y su mejor ajuste, algunos de los más utilizados son:

- ✓ Método gráfico
- ✓ Método de los momentos

- ✓ Método de máxima verosimilitud
- ✓ Procedimiento de mínimos cuadrados
- ✓ Método de los momentos ponderados por probabilidad
- ✓ Método de los L-momentos

2.4. Introducción a los L-momentos

Los momentos lineales surgen de combinaciones lineales de los *momentos ponderados probabilísticamente*, introducidos por Greenwood *et al.* (1979), y constituyen un sistema alternativo al método tradicional de los momentos convencionales para describir las funciones de distribución. Los momentos ponderados probabilísticamente, de la variable aleatoria X con una función de distribución $F(X)$, quedan definidos por la expresión:

$$M_{p,r,s} = E[X^p \{F(X)\}^r \{1 - F(X)\}^s] \quad (6)$$

Donde E indica el valor de la Esperanza, siendo p , k y s números reales.

Son particularmente útiles los casos en que $p = 1$ con $k = 0$, y $p = 1$ con $s = 0$, denotados como $\alpha_r = M_{1,0,r}$ y $\beta_r = M_{1,r,0}$ respectivamente.

$$\alpha_r = \int_0^1 x(u)(1-u)^r du \quad (7)$$

$$\beta_r = \int_0^1 x(u) u^r du \quad (8)$$

Finalmente, el L-momento de orden r , se estima con base a una pequeña modificación de las fórmulas anteriores:

$$\lambda_r = \int_0^1 x(u) P_{r-1}^*(u) du \quad (9)$$

Así, los primeros 4 L-momentos, en función de alfa y beta, son:

$$\lambda_1 = \alpha_1 = \beta_1 \quad (10)$$

$$\lambda_2 = \alpha_0 - 2\alpha_1 = \beta_1 - 2\beta_0 \quad (11)$$

$$\lambda_3 = \alpha_0 - 6\alpha_1 + 6\alpha_2 = 6\beta_2 - 6\beta_1 + \beta_0 \quad (12)$$

$$\lambda_4 = \alpha_0 - 12\alpha_1 + 30\alpha_2 - 20\alpha_3 \quad (13)$$

No obstante, para poder comparar diferentes estaciones, resulta necesario definir los momentos lineales de forma adimensional. Para ello, se calculan los ratios o cocientes de momentos lineales, L-momentos_ratios, dividiendo aquellos de mayor orden por la medida de escala λ_2 (Hosking y Wallis, 1997).

$$\tau_r = \lambda_r / \lambda_2, \quad r = 3, 4, \dots \quad (14)$$

Obtienen especial importancia los *L*-coeficientes de variación, asimetría y kurtosis.

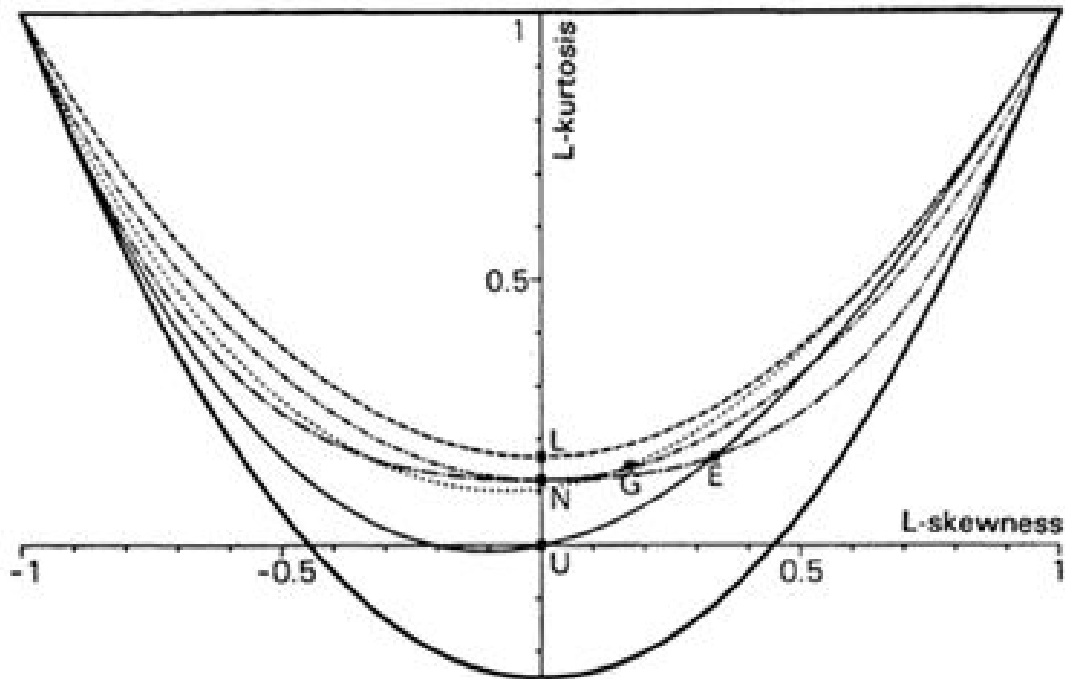
$$L - CV = \tau = \frac{\lambda_1}{\lambda_2} \quad (15)$$

$$L - asimetria = \tau_3 = \frac{\lambda_3}{\lambda_2} \quad (16)$$

$$L - kurtosis = \tau_4 = \frac{\lambda_4}{\lambda_2} \quad (17)$$

Sin embargo, Hosking y Wallis (1997) verifican que son varias las ventajas de los L-momentos desde el punto de vista teórico: capacidad para caracterizar un mayor rango de funciones de distribución, mayor robustez en presencia de valores atípicos y menor sensibilidad a la asimetría, entre otras.

Tal vez una de las formas más adecuadas de visualizar cómo los L-momentos contribuyen a definir un tipo de distribución es a través del denominado Diagrama de L-momento-ratio.



—	Pareto generalizada	-----	Valor extremo	-----	Logística generalizada
-----	Logarítmica normal	---	Gamma	—	Otras distribuciones

Gráfica 3. Diagrama L-momento-ratio de algunas distribuciones comunes. Las Distribuciones: U= Uniforme, E=Exponencial, G=Gumbel, N=Normal. (Gentileza M.G.S. Eng. Cons. WA, USA)

En el diagrama se observa que: las distribuciones de 2 parámetros (normal, logística, exponencial, Gumbel, etc.) se representan por un punto. Las distribuciones de 3 parámetros se representan por una línea. Las distribuciones de 4 parámetros (Kappa) se representan por un área.

2.5. Análisis Regional de Frecuencia

El Análisis Regional de Frecuencias es un procedimiento consistente en agregar la información proveniente de varias muestras, en una sola gran muestra, asumiendo previamente que todas éstas provienen de un mismo modelo de probabilidad, excepto por una diferencia entre ellas debidas a un factor de escala.

En nuestro caso, el procedimiento de “regionalización” es el conocido como “Índice de Avenida” o “Índice de Crecientes” (Index Flood).

En términos estadísticos, podríamos escribir esto como:

$$Q_i(F) = uiq(F) \quad (18)$$

Donde, Q_i es la función de cuantiles (montos de lluvia, caudal, temperatura, etc., para una determinada probabilidad de no excedencia); ui es el “factor de escala” (Índice de Avenida) y $q(F)$ es la función de cuantiles adimensional, proveniente de un mismo modelo de probabilidad que es similar para todas las estaciones que se encuentran dentro de una región denominada “homogénea”

¿Qué ventajas tiene, respecto al procedimiento convencional “at-site” o sitio por sitio, de forma separada?:

- Permite aumentar significativamente el tamaño total de la muestra.
- Permite tener en cuenta la variabilidad de los L-momentos de cada estación debido al tamaño muestral.
- Permite, por tanto, mejorar significativamente el ajuste a la distribución “parental”, de donde provienen estas muestras (registros de cada estación)
- En esencia, permite compensar la carencia de información en el tiempo, por su abundancia en el espacio

El término “Índice de Avenida”, aplicado por vez primera en el estudio de las avenidas por Dalrymple (1960) y de ahí su nombre, se ha utilizado desde entonces en el Análisis de Frecuencia de otras variables diferentes a la original, entre ellas las precipitaciones máximas (p.e. Saénz de Ormijana *et al.*, 1991; Guttman, 1993; Ferrer y Ardiles, 1994). Conocidos los cuantiles a escala regional y los valores medios locales, se pueden extrapolar extremos locales con cierta fiabilidad.

Conviene señalar que el método del Índice de Avenida parte de asumir una serie de supuestos:

- I. Las observaciones de una estación están idénticamente distribuidas.
- II. Las observaciones en una estación son independientes.
- III. Observaciones en diferentes estaciones son independientes.
- IV. Las distribuciones de frecuencia en distintos sitios de una misma región son idénticas excepto por un factor de escala.
- V. La forma de la curva regional de frecuencias se especifica correctamente desde el punto de vista matemático.

Las dos primeras suposiciones se cumplen para la mayor parte de los casos. Sin embargo, la tercera suposición rara vez se cumple estrictamente. En estudios habituales de precipitaciones y caudales, por ejemplo, cabe esperar la existencia de una cierta correlación entre las observaciones de estaciones cercanas entre sí. Más aún: es obvio que las dos últimas suposiciones nunca se van a cumplir en la práctica. Como mucho podrán ser alcanzadas aproximadamente, mediante selección cuidadosa de las estaciones de cada región y elección de distribuciones de frecuencia consistentes con las observaciones.

Así pues, el método del Índice de Avenida será apropiado siempre y cuando constituya un modelo robusto que aproxime a estos cinco supuestos.

2.6. Los Cuantiles

Los cuantiles constituyen una generalización del concepto de mediana. Así como la mediana divide a la serie estudiada en dos partes con el mismo número de elementos cada una, si la división se hace en cuatro partes, o en diez partes, o en cien partes, llegamos al concepto de cuantil.

Hay, principalmente, tres cuantiles importantes: cuartiles, deciles y percentiles.

- ✓ Cuartiles son tres valores con las siguientes características:

Q_1 : Primer cuartil, que es el valor de la variable por debajo del cual queda 1/4 de los elementos de la serie estudiada.

Q_3 : Tercer cuartil, que es el valor de la variable por debajo del cual quedan los 3/4 de los elementos que constituyen la serie.

Evidentemente el segundo cuartil coincide con la mediana. Como puede comprobarse, no tendría ninguna utilidad definir el cuarto cuartil. El cálculo de los cuartiles se realiza por el mismo procedimiento que el cálculo de la mediana, pues hay únicamente una diferencia cuantitativa entre ambas medidas, pero tienen significados paralelos.

Así, el primer cuartil se hallará aplicando la siguiente fórmula:

$$Q_1 = l + \frac{l}{f} \left(\frac{N}{4} - f_i \right) \quad (19)$$

Y el tercer cuartil:

$$Q_3 = l + \frac{l}{f} \left(\frac{3N}{4} - f_i \right) \quad (20)$$

Donde:

l : límite inferior de la clase a la que pertenece el cuartil, que es la clase que deja por debajo de ella el 25% de las observaciones (o el 75% en el caso de Q_3)

i : amplitud del intervalo.

f : frecuencia de la clase cuartílica.

N : total de elementos de la muestra.

f_i : frecuencia acumulada de todos los valores inferiores a la clase que contiene el cuartil.

- ✓ Deciles es la segunda clase de cuantiles. Si se divide toda la serie en diez partes iguales tendremos los deciles.

D_1 , el decil 1, deja el 10% de los valores de la serie por debajo de él.

Análogamente ocurre con los deciles D_2, D_3, \dots, D_9 . El decil 8, por ejemplo, deja el 80% de la masa de datos investigada por debajo de él.

Las fórmulas para calcularlos son también análogas a las de la mediana.

Por ejemplo:

$$D_1 = l + \frac{l}{f} \left(\frac{N}{10} - f_i \right) \quad (21)$$

$$D_1 = l + \frac{l}{f} \left(\frac{9N}{10} - f_i \right) \quad (22)$$

- ✓ Percentiles hay 99 percentiles que se denotan: $P_1, P_2, P_3, \dots, P_{98}, P_{99}$. Así P_{90} , por ejemplo, deja por debajo de él el 90% de los elementos.

La fórmula para realizar el cálculo del percentil 45, por ejemplo sería:

$$P_{45} = l + \frac{l}{f} \left(\frac{45N}{100} - f_i \right) \quad (23)$$

2.7. Región Homogénea

La región homogénea es un tipo de región que presenta mayor semejanza entre las unidades que la componen que con las unidades que pertenecen a otras regiones. Criterios muy diferentes y desigualmente complejos pueden utilizarse para localizar regiones homogéneas.

La homogeneidad se define siempre en relación con un cierto nivel de resolución o de generalización de las unidades geográficas, y sólo tiene en cuenta las variaciones observadas a una cierta escala, y según un número limitado de criterios.

3. Control de calidad de los datos de precipitación

La calidad del dato es un requisito emergente, que se desprende de las necesidades de información confiable y certera. La intención es proporcionar una calidad que sea satisfactoria, adecuada, confiable y económica.

Durante el proceso de observación y recolección de las series de datos, se producen inevitablemente errores de distinta naturaleza.

El control de calidad pretende detectar principalmente errores de carácter no sistemático, distintos de los que pretendemos detectar. Los errores que intentamos detectar y ajustar mediante los procesos de control de calidad son siempre de carácter sistemático y responden a fenómenos como el cambio de emplazamiento, entorno o instrumentación de los observatorios meteorológicos.

Los controles de calidad a la base de datos se realizan en dos etapas. En la primera etapa, los datos climáticos se someten a una serie de tests estadísticos de distintos tipos. Estos tests identifican registros que contienen variables meteorológicas con valores sospechosos. Una segunda etapa, involucra la verificación manual de todos los valores sospechosos identificados en la etapa anterior. Se verifican los datos sospechosos utilizando los registros originales, registrados en papel o en formato digital pero con mayor resolución temporal.

3.1. Grupos de tests de control de calidad

Los controles de calidad se organizaron en seis familias que agrupan tests de características similares.

Tests generales	Tests de rango variable	Tests de consistencia entre variables
Tests de rango fijo	Tests de continuidad temporal	Tests de consistencia espacial

Tabla 1. Familia de tests de control de calidad de datos. (Fuente: Reporte Técnico CRC-SAS-2014-001.)

A continuación describimos cada control de calidad mencionados en la tabla 1. Se pueden llamar tests también a los controles ya que los datos pasan por diferentes procesos para analizar su calidad.

- ✓ Controles generales. Estos controles verifican la integridad general de los datos. Por ejemplo, se controla que no haya fechas duplicadas o fuera de secuencia en las observaciones diarias. Otra verificación que se realiza es la frecuencia con la cual se registran los valores decimales para cada variable.
- ✓ Controles de rango fijo. Estos controles aseguran que no existan valores físicamente imposibles o nunca antes observados en el registro histórico. Los límites propuestos son fijos para cada variable durante todo el periodo de datos y todas las estaciones meteorológicas.
- ✓ Controles de rango variable. En esta familia, los rangos o umbrales usados para “marcar” valores sospechosos varían con el tiempo, tomando valores específicos para cada día o mes del año, por lo que los controles son más finos o sensibles que los tests de rango fijo.
- ✓ Controles de continuidad temporal. Estos controles estudian las secuencias de valores de cada variable en días consecutivos. Algunos de los controles en esta familia detectan la presencia de saltos o picos inusuales en las series de datos. Otros tests en esta familia identifican secuencias largas con valores idénticos.

- ✓ Controles de consistencia entre variables. Una serie de controles en esta familia o grupo evalúan la consistencia entre valores de pares o grupos de variables que deben guardar cierta consistencia.
- ✓ Controles de consistencia espacial. Todos los controles descritos anteriormente se realizan sobre los datos de una única estación meteorológica (aunque en algunos controles se use más de una variable). En esta familia de tests, sin embargo, los valores de una variable para una estación determinada (que generalmente se denomina la “estación central”) se comparan con los valores de esa variable registrados en estaciones geográficamente cercanas (o “estaciones vecinas”).

Dentro de cada familia de controles puede haber varios tests basados en distintos cálculos. Además, no todos los controles se aplican a todas las variables. Hay controles, por ejemplo, que se utilizan solamente para datos de precipitación.

3.1.1. Implementación de los tests de control de calidad

Todos los tests de control de calidad se implementaron en el lenguaje R (R Core Team, 2013), un entorno de programación diseñado para realizar análisis estadísticos y visualizar datos. El R es software abierto y sin costo, y está disponible para varias plataformas (Windows, Mac OS, Linux) bajo los términos de la Licencia Pública General GNU (GNU-GPL, por sus siglas en inglés; ver <http://www.r-project.org/Licenses/GPL-3>).

Una ventaja del lenguaje R es la existencia de una gran variedad de paquetes o librerías contribuidos por la comunidad de usuarios a nivel mundial que expanden la funcionalidad del lenguaje (ver, por ejemplo <http://cran.r-project.org/> o <http://dirk.eddelbuettel.com/cranberries/>). Estos paquetes ahorran tener que programar todos los cálculos deseados. Por ejemplo, existen al menos dos paquetes que permiten calcular el largo teórico del día (o sea, el número máximo de horas de sol) en función de la latitud y día del año.

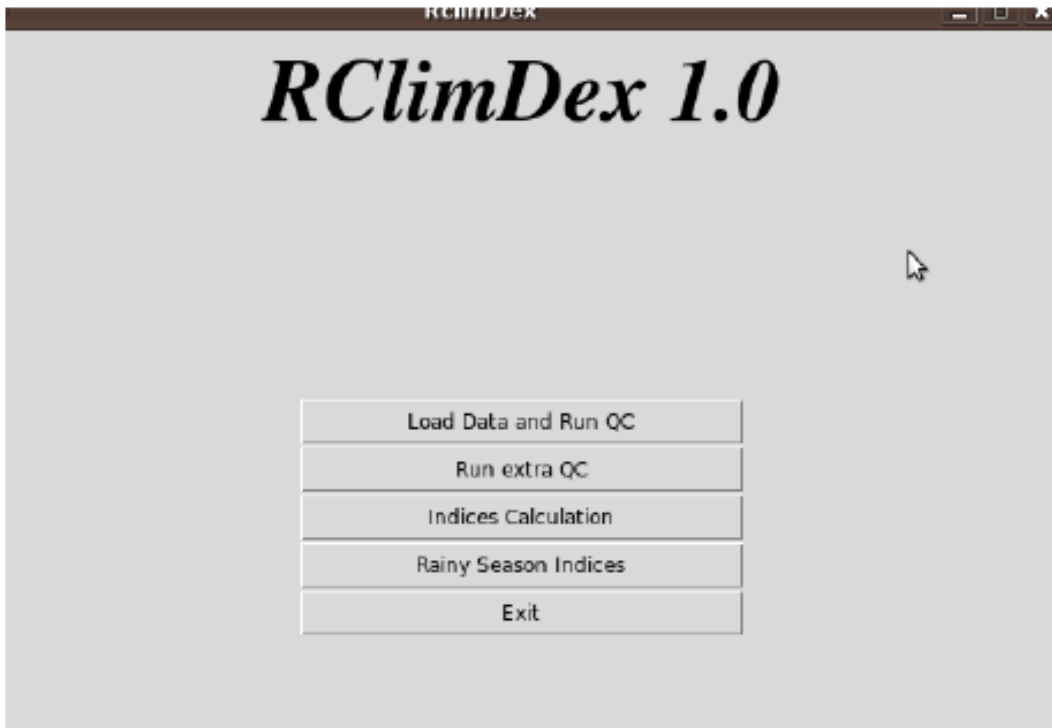
Para facilitar la organización y mantenimiento del código, cada familia de tests se implementó en un script separado. Los scripts que ejecutan cada familia de tests son llamados desde un script “maestro” que además realiza tareas generales como conectarse con la base de datos, etc. Otro script contiene funciones programadas en R que realizan tareas que se utilizan más de una vez. El encapsular una tarea repetida en una función permite evitar la duplicación de código y, por lo tanto, disminuir la chance de errores y facilitar el mantenimiento de los scripts.

3.2. RCLIMDEX_EXTRAQC.R.

- El primer control implementado fue el `rclimdex_extraQC.r`.

Es una rutina que corre sobre la plataforma R. Se origina en el software de control de calidad y cálculo de índices de cambio climático desarrollada por Environment Canadá para el Expert Team on Climate Change Detection and Índices y representa una modificación de la versión “oficial” de éste. Incluye un conjunto rutinas adicionales (Extra_QC) para un control de calidad más preciso e intensivo preparadas por Enric Aguilar (Centre for Climate Change, C3) y Marc Prohom (Servei Meteorològic de Catalunya).

Carga del software y utilización de las rutinas Extra_QC. Tras este preliminar, la carga de RCEQC se realizará mediante la escritura en el prompt de R de la orden `source('rclimdex_extraQC.r')`. Tras la carga de RCEQC aparecerá en nuestras pantallas un nuevo GUI, basado en TCL/TK, con el aspecto que aparece en la imagen.



Gráfica 4. Interfaz para ejecutar RClmDex 1.0.

La ejecución de las funcionalidades del paquete Extra_QC es extremadamente sencilla. Tan solo es necesario presionar con el puntero del ratón el botón “Run extra QC” (segundo por arriba) y utilizar la utilidad de navegación por el árbol de directorios para seleccionar el fichero que contiene nuestros datos originales (“raw data”).

Una vez cargado el fichero de datos, si RCEQC reconoce el formato adecuado, nos mostrará casi inmediatamente un mensaje de bienvenida. Tras presionar “OK”, y con un retraso inversamente proporcional a las prestaciones de vuestro ordenador, un segundo indicando que el proceso ha finalizado (ver abajo). Tras un nuevo 'OK', pasaremos a inspeccionar el nuevo directorio “extraqc” aparecido en nuestra área de trabajo. En este punto, es necesario hacer una advertencia. RCEQC, durante su ejecución, intentará crear el directorio mencionado.

3.3.ANDESQC

- El segundo control implementado fue el ANDESQC

ANDESQC se ha programado sobre R, a partir de RClimdex-extraqc y ha sido realizado bajo el contrato CCI-009-ATN/OC-12439-RG-2012 para CIIFEN. Se distribuye bajo licencia GNU Lesser GPL 3.0

TESTS	FECHA	Precipitación	Temperatura máxima	Temperatura mínima	Amplitud térmica
Outliers por meses		X	X	X	X
Outliers pentales			X	X	
Outliers gamma		X			
Outliers espaciales		X	X	X	
Tmax< Tmin			X	X	
Valores fuera de rango		X	X	X	
Valores duplicados	X	!	!	!	!
Diferencias interdiarias			X	X	
Valores consecutivos		X	X	X	
Precip. Acumulada		X			
Cadenas de Markov		X	X	X	
Cadena Precip.		X			
Redondeo		X	X	X	

Tabla 2. Los tests incluidos en ANDESQC, donde nos indica que tests se aplican para cada variable analizada en este control de calidad.

Como se muestra en dicha tabla un test puede ser aplicado a diferentes variables. En el test valores duplicados solo se usa para fechas, ya que debemos controlar que este cronológicamente. El signo de exclamación significa que no es válido para aquellas variables, como por ejemplo la precipitación, donde en ocasiones tenemos datos duplicados en varias fechas consecutivas.

Se definirán solamente los test que hacen controles a los datos de precipitación de la serie temporal analizada para el proyecto.

- Outliers por meses.

Identifican como outliers aquellos valores que:

- Se encuentren por encima del percentil $75 + n$ rangos intercuartílicos (RIC)
- Se encuentren por debajo del percentil $25 - n$ RIC
 - Outliers Gamma.

Estratifica lo valores de precipitación por meses naturales. Ajusta a cada mes una distribución de probabilidad gamma. Etiqueta como outliers aquellos valores que superan el cuantil especificado en la parametrización.

- Valores Fuera Rango.

Identifica valores que se consideran imposibles, como la precipitación negativa o temperaturas de 100°C . Los valores etiquetados por este tests no son considerados outliers o valores sospechosos sino **ERRORES**.

- Valores Duplicados.

Identifica fechas repetidas. Los valores etiquetados por este tests no son considerados outliers o valores sospechosos sino **ERRORES**.

- Valores Consecutivos.

Identifica rachas consideradas demasiado largas de valores consecutivos iguales para TX, TN y precipitación distinta de 0.

- Precipitación Acumulada.

Identifica valores de precipitación elevados precedidos de una racha de valores 0 o en ausencia de valores previos.

- Cadenas de Markov.

Se trata de un tests de notable complejidad estadística. Se basa en cadenas de Markov de dos estados (**Seco = 0; húmedo = distinto de cero**) y rango 1.

Las series de precipitación se estratifican por ventanas móviles de 11 días. Se evalúan las probabilidades transicionales de una observación a la siguiente:

- P0 = probabilidad de día lluvioso; P1 = probabilidad de día seco
- P00 = probabilidad día seco tras día seco; P01 probabilidad día lluvioso tras día seco

Con las probabilidades obtenidas, se utiliza la distribución geométrica para extraer de la misma la racha que representa un determinado cuantil (configurable por el usuario). De encontrarse una racha que exceda la determinada anteriormente, se etiquetarán como sospechosos todos los días que la componen más el anterior y el posterior.

- Cadenas de Precipitación.

Se trata de una versión simplificada del anterior tests, que simplemente analiza la existencia de rachas superiores a n valores idénticos (presumiblemente, 0), siendo n configurable. La parametrización de esta función debe tener en cuenta la existencia de una estación seca.

- Redondeo.

Se basa en determinar el número máximo de observaciones que se permiten con el mismo valor tras el punto decimal (.0 a .9). Se aplica a TX, TN y precipitación distinta de 0.0.

3.4. RClimTool

- El tercer control implementado fue RClimTool.

RClimTool ha sido diseñada con el objetivo de facilitar a los usuarios el análisis estadístico, control de calidad, llenado de datos faltantes, análisis de homogeneidad y

cálculo de indicadores para las series climatológicas diarias de temperatura máxima, temperatura mínima y precipitación.

Para ejecutar la interfaz de la aplicación debemos cargar el código fuente. Una vez se ha cargado el código exitosamente aparecerá la siguiente interfaz gráfica:

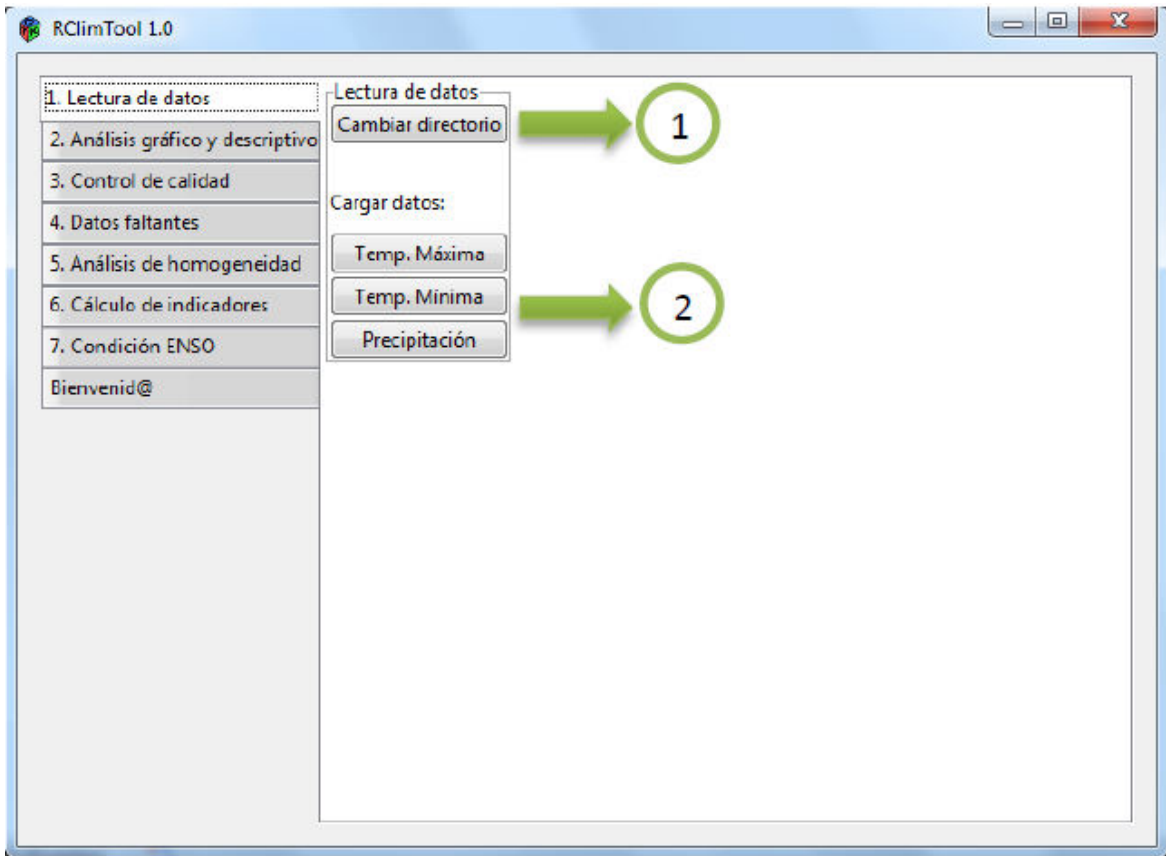


Gráfica 5. Interfaz para ejecutar RclimTool.

En el módulo de lectura de datos encontraremos diferentes botones que nos permitirán leer y cargar las bases de datos que contienen la información de las variables de interés.

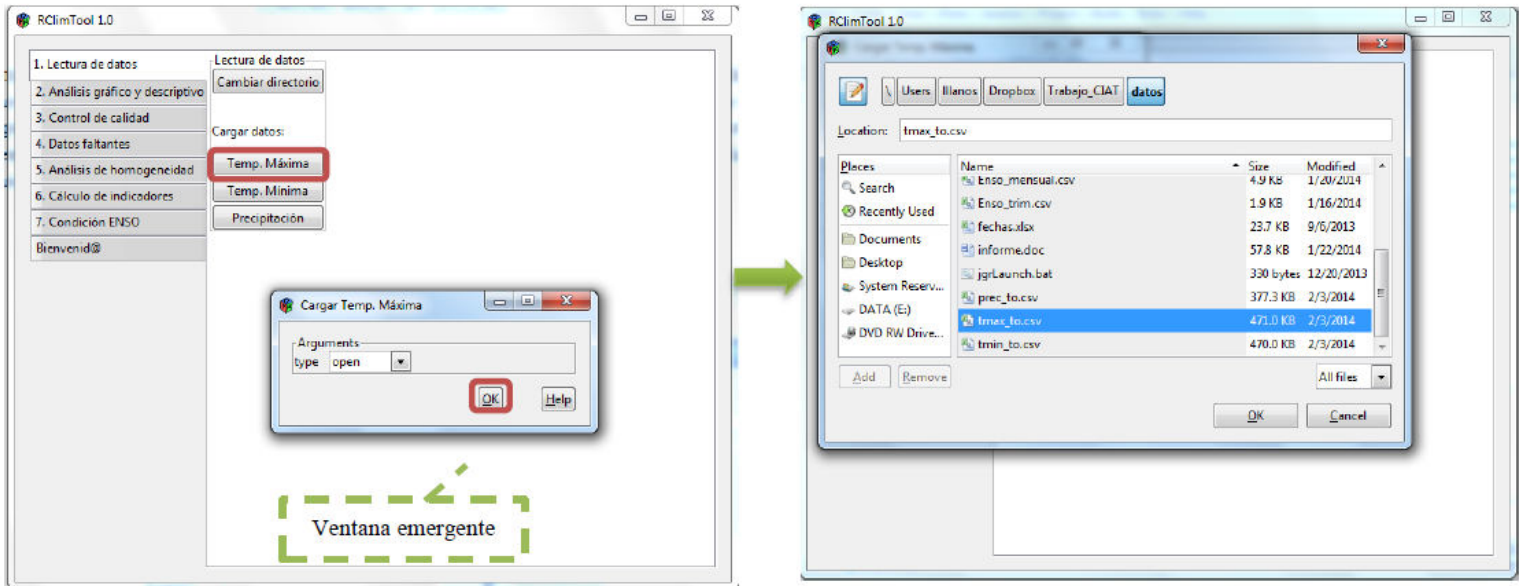
El botón *cambiar directorio* en la parte (1) de la gráfica 6, ofrece la opción de seleccionar el directorio donde se encuentran los archivos que se van a cargar, también será la ubicación para guardar todas las salidas de la aplicación.

En la parte (2) de la gráfica 6, se encuentran los botones que permite cargar la información cada una de las variables. Por ejemplo, al dar clic en el botón *Temp. Máxima* aparecerá una ventana emergente en la cual se debe ubicar el archivo que contiene las temperaturas máximas diarias de las diferentes estaciones. Este procedimiento se realiza para las demás variables a analizar.



Gráfica 6. Ventana para lectura de datos.

En esta ventana se selecciona la ubicación y el archivo que deseamos cargar. Seleccionamos el archivo y damos clic en *ok* como se ve en la gráfica 7. Recuerde cerrar la ventana emergente cada vez que se cargue una variable diferente.



Gráfica 7. Ejemplo de selección de archivo.

Los siguientes puntos de la interfaz fueron detallados en el Anexo II con ejemplos de los resultados.

4. Metodología

4.1. Población y muestra

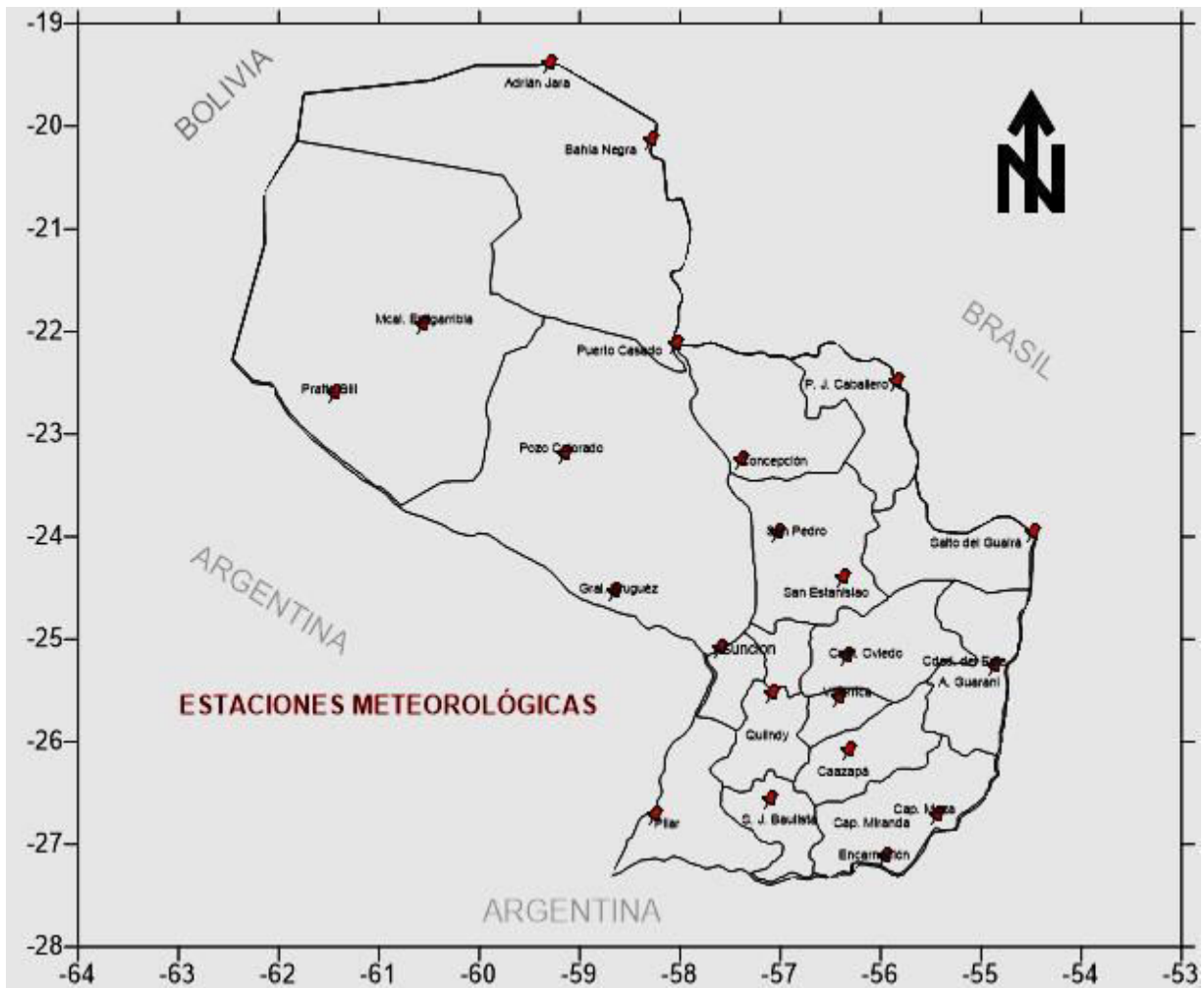
Los datos que se utilizarán provienen de la base de datos de la Dirección de Meteorología e Hidrología, con rigurosos controles de calidad, creando una base de datos controlados.

Al ser un estudio de caracterización es adecuado que se utilicen todos los datos posibles y al ser de tipo climatológico se necesitan al menos de 30 años o más. Para el cálculo de las frecuencias de probabilidad y los periodos de retorno en el evento de sequía se utilizó el periodo desde el 01/01/1961 hasta el 31/12/2013 para las 23 estaciones.

Las estaciones a ser utilizadas se especifican a continuación:

Código de la Estación	Nombre de la estación	Latitud	Longitud
86011	Adrián Jara	-19,53	-59,37
86033	Bahía Negra	-20,23	-58,17
86065	Pratts Gill	-22,56	-61,56
86068	Mariscal Estigarribia	-22,03	-60,62
86086	Puerto Casado	-22,28	-57,94
86097	Pedro Juan Caballero	-22,64	-55,83
86128	Pozo Colorado	-23,50	-58,79
86134	Concepción	-23,44	-57,43
86170	General, Bruguez	-24,74	-58,84
86185	San Pedro	-24,09	-57,09
86192	San Estanislao	-24,67	-56,46
86210	Salto Guairá	-24,03	-54,35
86218	Asunción (Aeropuerto)	-25,24	-57,52
86221	Paraguarí	-25,63	-57,15
86233	Villarrica	-25,75	-56,44
86234	Coronel, Oviedo	-25,45	-56,38
86246	Aeropuerto Guaraní	-25,46	-54,84
86251	Quyquyó	-26,22	-56,98
86255	Pilar	-26,88	-58,32
86260	San Juan Bautista Misiones	-26,67	-57,13
86268	Caazapá	-26,18	-56,36
86285	Capitán Meza	-26,83	-55,33
86297	Encarnación	-27,30	-55,09

Tabla 3. Lista de estaciones, donde se detallan: código de identificación de dicha estación; nombre de la localidad donde se encuentra ubicada; latitud y longitud para saber su ubicación geográfica.



Gráfica 8. Mapa de las estaciones analizadas.

4.2. Ajuste de los datos de precipitación

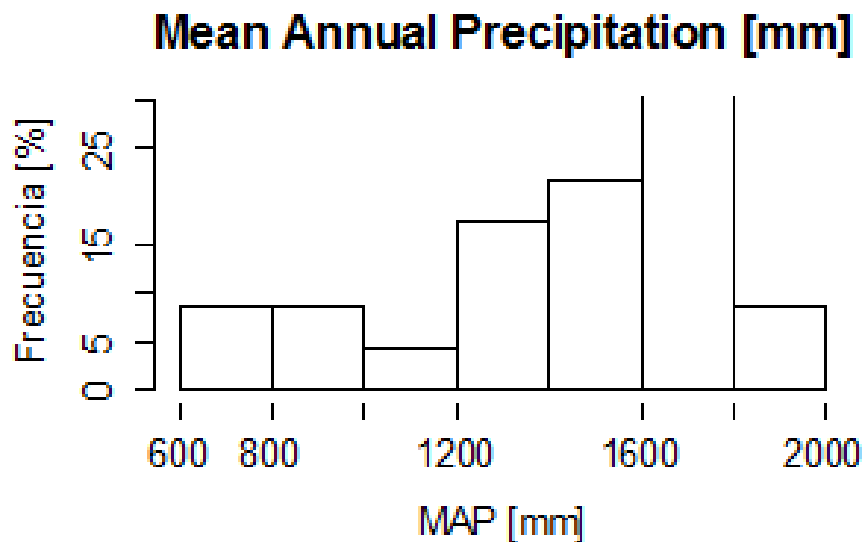
La disponibilidad de datos confiables es una exigencia indefectible para cualquier tipo de análisis hidrológico, en particular, aquel relacionado con la determinación de probabilidad de ocurrencia de eventos extremos, en la que el error, presencia o ausencia de un dato en particular, puede tener importantes consecuencias sobre las estimaciones de probabilidad con altos periodos de retorno.

Los principales objetivos de esta etapa son:

1. Maximizar nuestro entendimiento de los datos.

2. Detectar las variables importantes.
3. Detectar valores atípicos y anomalías.
4. Comprobar los supuestos implícitos en los análisis.
5. Determinar los ajustes óptimos de los factores.
6. Proporcionar al analista información relevante en cuanto al modelo que mejor ajusta los datos.

Se puede utilizar varios análisis para la base de datos que se examina, el más común es por ejemplo el Análisis Exploratorio que mediante técnicas y herramientas como el uso de los histogramas de frecuencias, los box-plot, las curvas q-q, etc., nos dan un mejor panorama del comportamiento de los registros.

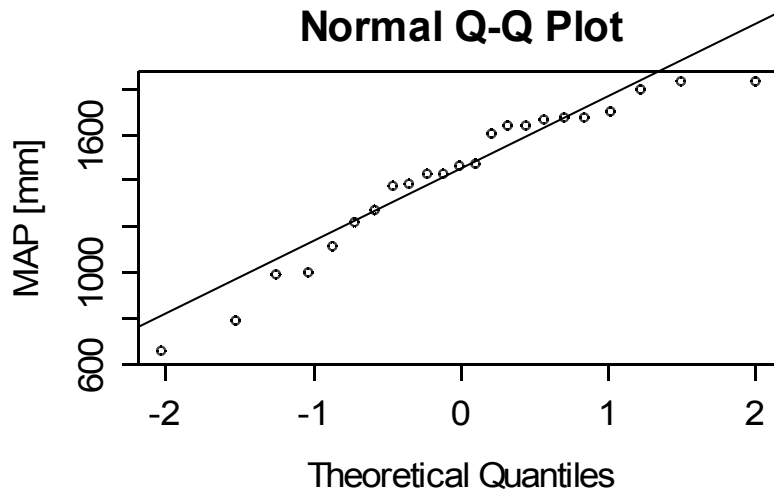


Gráfica 9. Histograma de frecuencia (%) de la precipitación media anual.

En la tabla 4 se muestra los valores calculados por el método para la precipitación media anual para cada estación estudiada.

Identificación de cada estación	Nombre de cada estación	Precipitación Media Anual (mm) (PMA)
st-py-011	Adrián Jara	9830
st-py-033	Bahía Negra	993
st-py-065	Pratts Gill	646
st-py-068	Mariscal Estigarribia	777
st-py-086	Puerto Casado	1270
st-py-097	Pedro Juan Caballero	1660
st-py-128	Pozo Colorado	1100
st-py-134	Concepción	1370
st-py-170	General Bruguez	1210
st-py-185	San Pedro	1470
st-py-192	San Estanislao	1590
st-py-210	Salto Guairá	1670
st-py-218	Asunción (Aeropuerto)	1380
st-py-221	Paraguarí	1450
st-py-233	Villarrica	1690
st-py-234	Coronel Oviedo	1820
st-py-246	Aeropuerto Guaraní	1830
st-py-255	Pilar	1420
st-py-260	San Juan Bautista Misiones	1670
st-py-268	Caazapá	1630
st-py-285	Capitán Meza	1630
st-py-297	Encarnación	1790

Tabla 4. Precipitación Media Anual calculada para cada estación.



Gráfica 10. Gráfico entre la precipitación media anual vs los cuantiles teóricos.

En el gráfico normal q-q si las dos distribuciones coinciden, veremos un gráfico muy parecido a una recta.

4.3. Identificación regiones homogéneas

Para la identificación de las regiones homogéneas los métodos más utilizados son el Análisis Clúster, Métodos Multivariados, Métodos basados en atributos geográficos y climáticos del área de estudio, región de influencia, lógica difusa, mapas auto organizados e Índice de Estacionalidad (Gaál *et al*, 2007; Gaál y Kyselý, 2009; Fowler y Kilsby, 2003; Chavochi y Soleiman, 2009; Burn y Goel, 2000; Kohnova *et al*, 2009; Lin y Chen, 2004).

En el Análisis Clúster se utiliza la información de una serie de variables para cada sujeto u objeto y, conforme a estas variables se mide la similitud entre ellos. Una vez medida la similitud se agrupan en: grupos homogéneos internamente y diferentes entre sí. La "nueva dimensión" lograda con el clúster se aprovecha después para facilitar la aproximación "segmentada" de un determinado análisis.

El análisis de datos multivariantes tiene por objeto el estudio estadístico de varias variables medidas en elementos de una población.

Usaremos dos métodos que son el Índice de Estacionalidad y el Día Juliano Medio. El primero se relaciona con el grado de concentración de la precipitación mientras que el segundo se refiere a la época en que ésta se concentra.

4.4. Determinación del modelo de probabilidad de mejor ajuste

Cuando una región homogénea ha satisfecho la condición de homogeneidad es posible, entonces, determinar, basados tanto en el uso del diagrama de los L-momentos regionales como en el uso de la prueba de bondad de ajuste Z^{DIST} , la distribución de probabilidad regional, así como estimar los parámetros de esta distribución (Wallis *et al*, 2007).

La utilización de un diagrama que relacione los L-momentos regionales respecto a los L-momentos teóricos, para diversos tipos de modelos de distribución de probabilidad, es una etapa fundamental en el ARF basado en L-momentos. Vogel y Fennessey (1993) y Peel *et al*, (2001) indican que los diagramas de L-momento-ratios son una herramienta útil para la selección de la distribución de probabilidad aunque no suficientes, por lo que recomiendan el uso de medidas de bondad de ajustes basados en los L-momentos regionales.

Hosking y Wallis (1996) desarrollan una medida de bondad de ajuste basado en las estadísticas medidas de los L-momentos regionales. Los autores indican que el procedimiento para aceptar una o varias distribuciones de mejor ajuste se basan en el valor Z^{DIST} . Si $Z^{DIST} < |1.64|$, entonces se acepta la hipótesis de buen ajuste de la distribución, caso contrario, se rechaza.

$$Z^{DIST} = \frac{\tau_4^{DIST} - \bar{\tau}_4 + \beta_4}{\sigma_4} \quad (24)$$

4.5. Determinación de cuantiles

Tal como se había indicado respecto al fundamento del Índice de Avenida, la distribución de frecuencias, que pueden ser expresadas en términos de la curva de cuantiles, se define por:

$$Q_i(F) = u_i q(F) \quad (25)$$

Donde, Q_i es la función de cuantiles (montos de lluvia, etc., para una determinada probabilidad de no excedencia); u_i es el “factor de escala” (Índice de Avenida) y $q(F)$ es la función de cuantiles adimensional.

De este modo, lo que interesa es determinar la “curva de crecimiento regional” que es la función de cuantiles de la región, similar a la de cada sitio dentro de la región, excepto por el factor de escala u_i .

Para determinar la curva de cuantiles, se debe:

- Seleccionar un modelo de distribución.
- Utilizar los L-momentos regionales estimados para calcular los parámetros de la distribución seleccionadas.

4.6. Mapeo

Como lo explica en el curso Internacional Análisis Regional de Frecuencia de Sequías basado en L-momentos Jorge Núñez C. “La base del mapeo espacial de la distribución de probabilidad de un evento, es determinar, para cada celda o píxel en el mapa de salida, los parámetros de la función de distribución de probabilidad, que especifican la curva de cuantiles en cada celda. Para ello, se debe, a partir de los L-momentos en cada sub-región, determinar cómo éstos varían en el espacio”.

Esto involucra conocer y relacionar la variación espacial de los L-momentos (desconocida), con la variación espacial de una variable auxiliar que sí se conoce y tiene una buena capacidad explicativa sobre los L-momentos.

La base para el mapeo de cuantiles y probabilidad, es disponer de un buen mapa de Precipitación Media Anual.

La precipitación media anual afecta las estimaciones de los cuantiles a través de dos formas:

- En la estimación de los L-momentos
- En la estimación del factor de escala (Índice de Avenida)

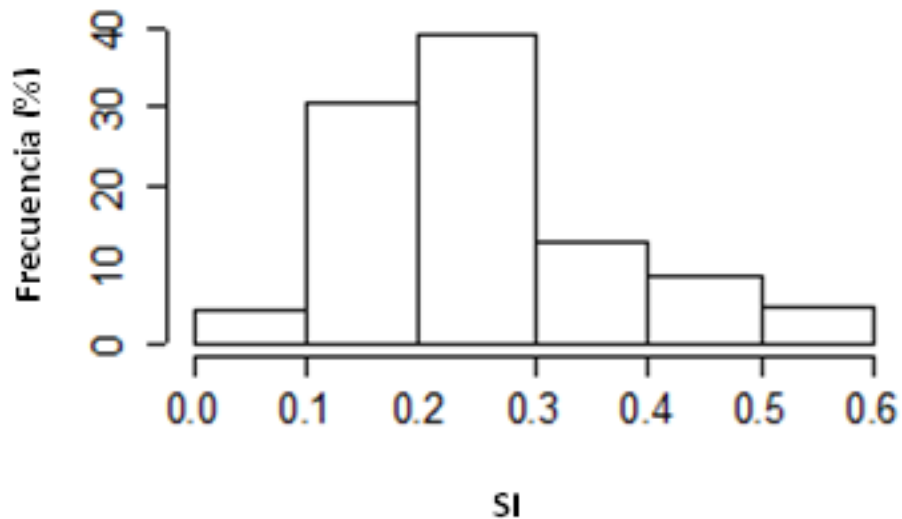
Para mapear los periodos de retorno mayores a 100 años se aplica una técnica llamada máscara, ésta consiste en poner en blanco aquellos valores que en la práctica carecen de valor.

5. Resultados

Para la elaboración de los gráficos y mapas se utilizó la herramienta R y su visualizador Rstudio.

5.1. Procedimientos de regionalización

La gráfica 11 se obtuvo el Índice de Estacionalidad (SI sus siglas en inglés Seasonality Index) de los datos anuales de precipitación. El SI toma valores entre 0 y 1. Valores cercanos a 0 indican una gran variación en los días de ocurrencia, mientras que valores cercanos a 1 indican una baja variación en los días de ocurrencia y, por tanto, alta concentración de los datos.



Gráfica 11. Índice de Estacionalidad. (SI).

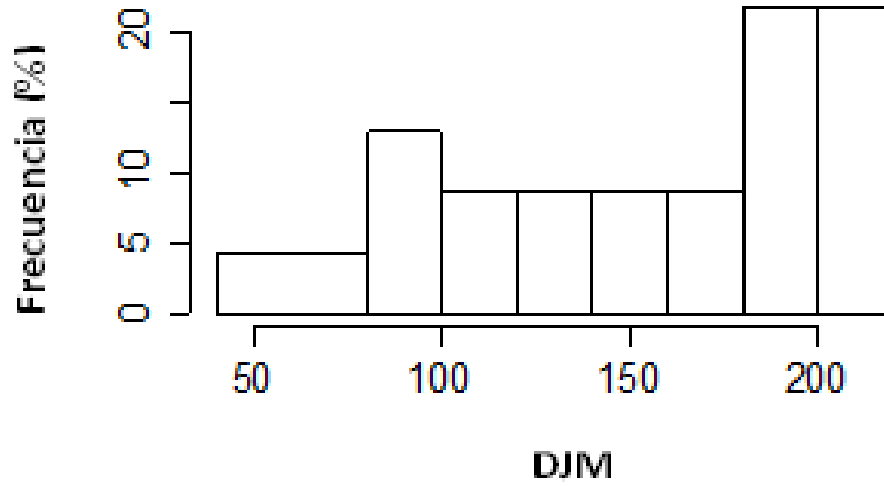
En la tabla 5 se muestra los valores calculados por el método para Índice de Estacionalidad para cada estación estudiada.

Identificación de cada estación	Nombre de cada estación	Índice de estacionalidad (SI)
st-py-011	Adrián Jara	0,44459
st-py-033	Bahía Negra	0,39981
st-py-065	Pratts Gill	0,51384
st-py-068	Mariscal Estigarribia	0,47062
st-py-086	Puerto Casado	0,31038
st-py-097	Pedro Juan Caballero	0,23802
st-py-128	Pozo Colorado	0,35434
st-py-134	Concepción	0,28204
st-py-170	General Bruguez	0,29708
st-py-185	San Pedro	0,21320
st-py-192	San Estanislao	0,19750
st-py-210	Salto Guairá	0,23296
st-py-218	Asunción (Aeropuerto)	0,23274
st-py-221	Paraguarí	0,22998
st-py-233	Villarrica	0,17863
st-py-234	Coronel Oviedo	0,15483
st-py-246	Aeropuerto Guaraní	0,17802
st-py-255	Pilar	0,27863
st-py-260	San Juan Bautista Misiones	0,20551
st-py-268	Caazapá	0,17069
st-py-285	Capitán Meza	0,18590
st-py-297	Encarnación	0,17910

Tabla 5. SI para cada estación.

En la gráfica 12 se muestra los datos agrupados en función de su Día Juliano Medio (MJD sus siglas en inglés Julian Mean Day).

Se usa este método debido a que puede existir estaciones con SI similar, pero cuya concentración ocurre en distintas épocas (zonas con precipitaciones concentradas en verano y otras en invierno).



Gráfica 12. Día Juliano Medio. (MJD)

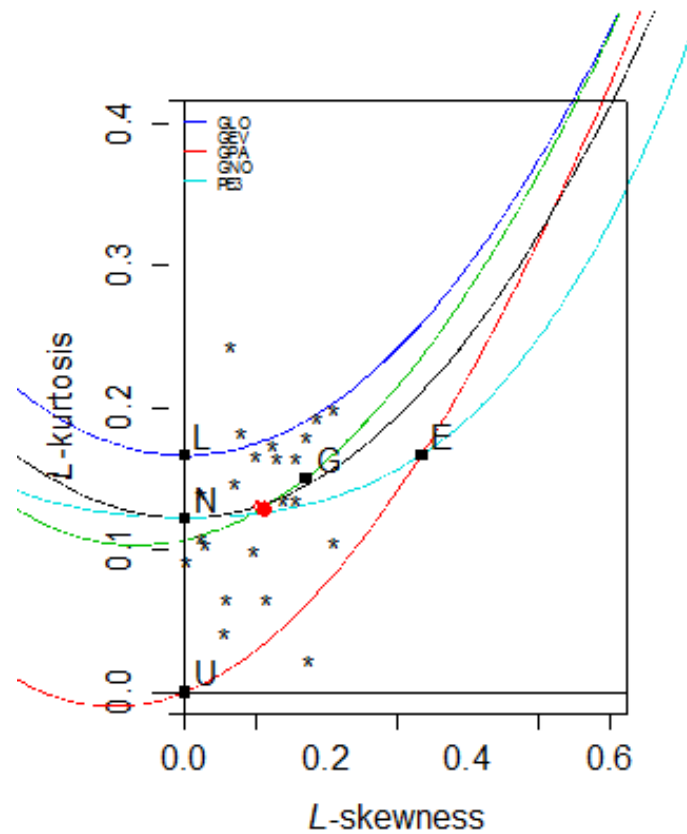
En la tabla 6 se muestra los valores calculados por el método para el día juliano medio para cada estación estudiada.

Identificación de cada estación	Nombre de cada estación	Día Juliano Medio (DJM)
st-py-011	Adrián Jara	0,01032
st-py-033	Bahía Negra	0,00477
st-py-065	Pratts Gill	0,01862
st-py-068	Mariscal Estigarribia	0,01355
st-py-086	Puerto Casado	0,00618
st-py-097	Pedro Juan Caballero	0,00615
st-py-128	Pozo Colorado	0,00754
st-py-134	Concepción	0,00692
st-py-170	General Bruguez	0,01183
st-py-185	San Pedro	0,00735
st-py-192	San Estanislao	0,00523
st-py-210	Salto Guairá	0,00507
st-py-218	Asunción (Aeropuerto)	0,00834
st-py-221	Paraguarí	0,00876
st-py-233	Villarrica	0,00553
st-py-234	Coronel Oviedo	0,00480
st-py-246	Aeropuerto Guaraní	0,00495
st-py-255	Pilar	0,00652
st-py-260	San Juan Bautista Misiones	0,00549
st-py-268	Caazapá	0,00519
st-py-285	Capitán Meza	0,00480
st-py-297	Encarnación	0,00489

Tabla 6. JMD para cada estación.

5.2. Determinación del modelo de probabilidad de mejor ajuste

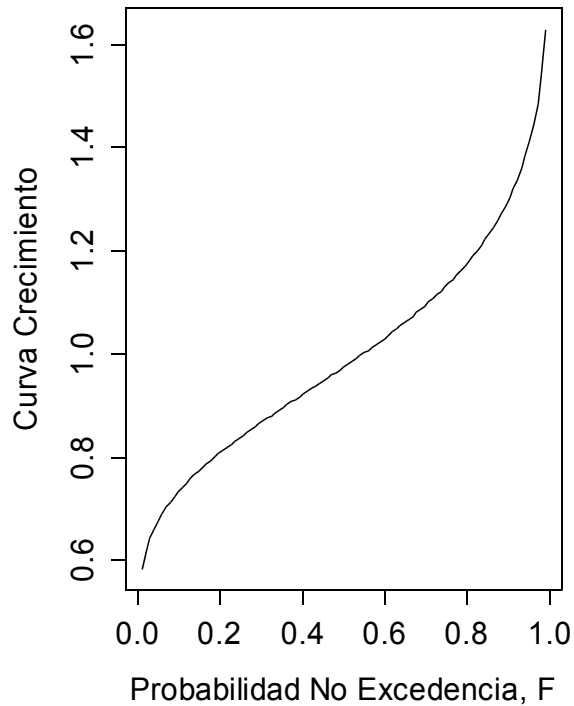
En la gráfica 13, se obtuvo la relación de L-kurtosis y L-asimetría. Se presenta el análisis de bondad de ajuste para la región, como ejemplo, obtenido mediante el uso de los diagramas L-momento-ratios.



Gráfica 13. Gráfico entre la relación de L-kurtosis y L-asimetría.

5.3. Cuantiles

En la gráfica 14 se presentan los cuantiles regionales de la región homogénea. Una de las dudas a resolver en este estado del análisis, se refiere a seleccionar, dentro de las posibles alternativas de distribuciones aceptadas, cual será utilizada en la determinación de los cuantiles y el posterior mapeo de la distribución.



Gráfica 14. Curva de Crecimiento Regional.

En la tabla 4 se se presentan los parámetros de la distribución regional seleccionada para cada región. En este caso solo se agrupan en la región 1. Clúster Regional Adaptado muestra la región según cada distribución.

Identificación de cada estación	Nombre de cada estación	Clúster Regional Adaptado	Distribuciones
st-py-011	Adrián Jara	1	Distribución de Pearson 3 (pe3)
st-py-033	Bahía Negra	1	Generalizada Normal (gno)
st-py-065	Pratts Gill	1	Generaliza de valor extremo (gev)
st-py-068	Mariscal Estigarribia	1	Pareto Generalizada (gpa)
st-py-086	Puerto Casado	1	Generaliza de valor extremo (gev)
st-py-097	Pedro Juan Caballero	1	Distribución de Pearson 3 (pe3)
st-py-128	Pozo Colorado	1	Generalizada Logística (glo)
st-py-134	Concepción	1	Generaliza valor extremo (gev)
st-py-170	General Bruguez	1	Generalizada Logística (glo)
st-py-185	San Pedro	1	Generaliza valor extremo (gev)
st-py-192	San Estanislao	1	Generalizada Logística (glo)
st-py-210	Salto Guairá	1	Generalizada Normal (gno)
st-py-218	Asunción (Aeropuerto)	1	Generalizada Logística (glo)
st-py-221	Paraguarí	1	Generalizada Logística (glo)
st-py-233	Villarrica	1	Generaliza valor extremo (gev)
st-py-234	Coronel Oviedo	1	Generaliza valor extremo (gev)
st-py-246	Aeropuerto Guaraní	1	Pareto Generalizada (gpa)
st-py-255	Pilar	1	Generalizada Logística (glo)
st-py-260	San Juan Bautista Misiones	1	Distribución de Pearson3 (pe3)
st-py-268	Caazapá	1	Generaliza valor extremo (gev)
st-py-285	Capitán Meza	1	Pareto Generalizada (gpa)
st-py-297	Encarnación	1	Pareto Generalizada (gpa)

Tabla 7. Parámetros de la distribución según la región homogénea.

5.4. Mapeo

Para obtención de los siguientes mapas se utilizaron dos herramientas denominadas ZAGA GISy Q-GIS

SAGA (acrónimo inglés de System for Automated Geoscientific Analyses o Sistema para Análisis Automatizados Geocientíficos en español) es un software híbrido de información geográfica.

- a. El primer objetivo de SAGA es dar una plataforma eficaz y fácil para la puesta en práctica de métodos geocientíficos mediante su interfaz de programación (API).
- b. El segundo objetivo es hacer estos métodos accesibles de una manera fácil.

Esto se consigue principalmente mediante su interfaz gráfica de usuario (GUI). Juntos, API Y GUI son el verdadero potencial de SAGA: un sistema cada vez mayor y rápido de métodos geocientíficos.

QGIS is la mejor herramienta SIG en la comunidad de software libre y abierto (FOSS).

QGIS es un Sistema de Información Geográfica (SIG) de Código Abierto licenciado bajo GNU - General Public License. QGIS es un proyecto oficial de Open Source Geospatial Foundation (OSGeo). Corre sobre Linux, Unix, Mac OSX, Windows y Android y soporta numerosos formatos y funcionalidades de datos vector, datos ráster y bases de datos.

QGIS es un Sistema de Información Geográfica (SIG) de Código Abierto licenciado bajo GNU - General Public License . QGIS es un proyecto oficial de Open Source Geospatial Foundation (OSGeo). Corre sobre Linux, Unix, Mac OSX, Windows y Android y soporta numerosos formatos y funcionalidades de datos vector, datos ráster y bases de datos.

5.4.1. Período de retorno de eventos de sequía

Para el cálculo de los periodos de retorno es posible determinar las propiedades de la función de densidad de probabilidad, esto es, sus parámetros (mismos que son estimados a partir de los L-momentos) a partir de alguna variable explicativa o predictiva (la principal variable predictiva de los L-momentos es la magnitud de la precipitación media anual), con una mejor distribución espacial de información que aquella basada solamente en la de los sitios de observación disponibles.

En la tabla 8, se tiene los valores calculados de los diferentes periodos de retorno para cada estación meteorológica analizada.

Identificación de cada estación	Nombre de cada estación	Periodo de Retorno del 10%	Periodo de Retorno del 60%	Periodo de Retorno del 70%	Periodo de Retorno del 80%	Periodo de Retorno del 90%	Periodo de Retorno del 120%
st-py-011	Adrián Jara	2,7	1	1,41	4,92	2,64	5,44
st-py-033	Bahía Negra	1,61	2,9	1,11	5,26	2,97	5,34
st-py-065	Pratts Gill	5,02	3,11	1,15	5,37	3,02	5,31
st-py-068	Mariscal Estigarribia	2,7	1,95	7,03	3,64	2,49	3,92
st-py-086	Puerto Casado	2,7	2,63	2,39	6,05	2,76	6,2
st-py-097	Pedro Juan Caballero	2,7	2,63	7,2	7,82	2,94	6,98
st-py-128	Pozo Colorado	2,87	4,04	1,5	6,3	3,13	6,46
st-py-134	Concepción	4,27	8,47	1,79	6,12	2,99	6,08
st-py-170	General Bruguez	3,36	6,09	1,91	7,09	3,23	7,11
st-py-185	San Pedro	4,27	8,47	1,79	6,12	2,99	6,08
st-py-192	San Estanislao	2,7	6	1,45	5,24	2,64	6,11
st-py-210	Salto Guairá	1,68	4,57	1,31	5,37	2,88	5,58
st-py-218	Asunción (Aeropuerto)	1,34	5,95	1,72	6,31	2,98	6,6
st-py-221	Paraguarí	1,34	5,95	1,72	6,31	2,98	6,6
st-py-233	Villarrica	7,73	7,69	1,51	5,3	2,74	5,64
st-py-234	Coronel Oviedo	7,73	7,69	1,51	5,3	2,74	5,64
st-py-246	Aeropuerto Guaraní	2,7	2,63	9,36	3,76	2,43	4,27
st-py-255	Pilar	2,7	1,21	2,21	6,54	2,87	6,87
st-py-260	San Juan Bautista Misiones	2,7	1,09	1,8	5,88	2,89	6,01
st-py-268	Caazapá	7,73	7,69	1,51	5,3	2,74	5,64
st-py-285	Capitán Meza	2,7	2,63	9,01	3,74	2,43	4,24
st-py-297	Encarnación	2,7	2,63	9,01	3,74	2,43	4,24

Tabla 8. Periodos de retorno.

En las gráficas del 15 al 20 se detallan los mapas de los diferentes periodos de retorno obtenidos.

Se obtuvo un mapa de periodo de retorno del 10%, esto implica un 90% de déficit de PMA. **Gráfica 15.**

Se obtuvo un mapa de periodo de retorno del 60%, esto implica un 40% de déficit de PMA. Se observa en el gráfico, las zonas blancas con periodos de retornos superiores a los 100 años, que para cualquier actividad que se pueda desarrollar, es un periodo excesivamente largo y por ende no es útil para un estudio a corto o mediano plazo. **Gráfica 16.**

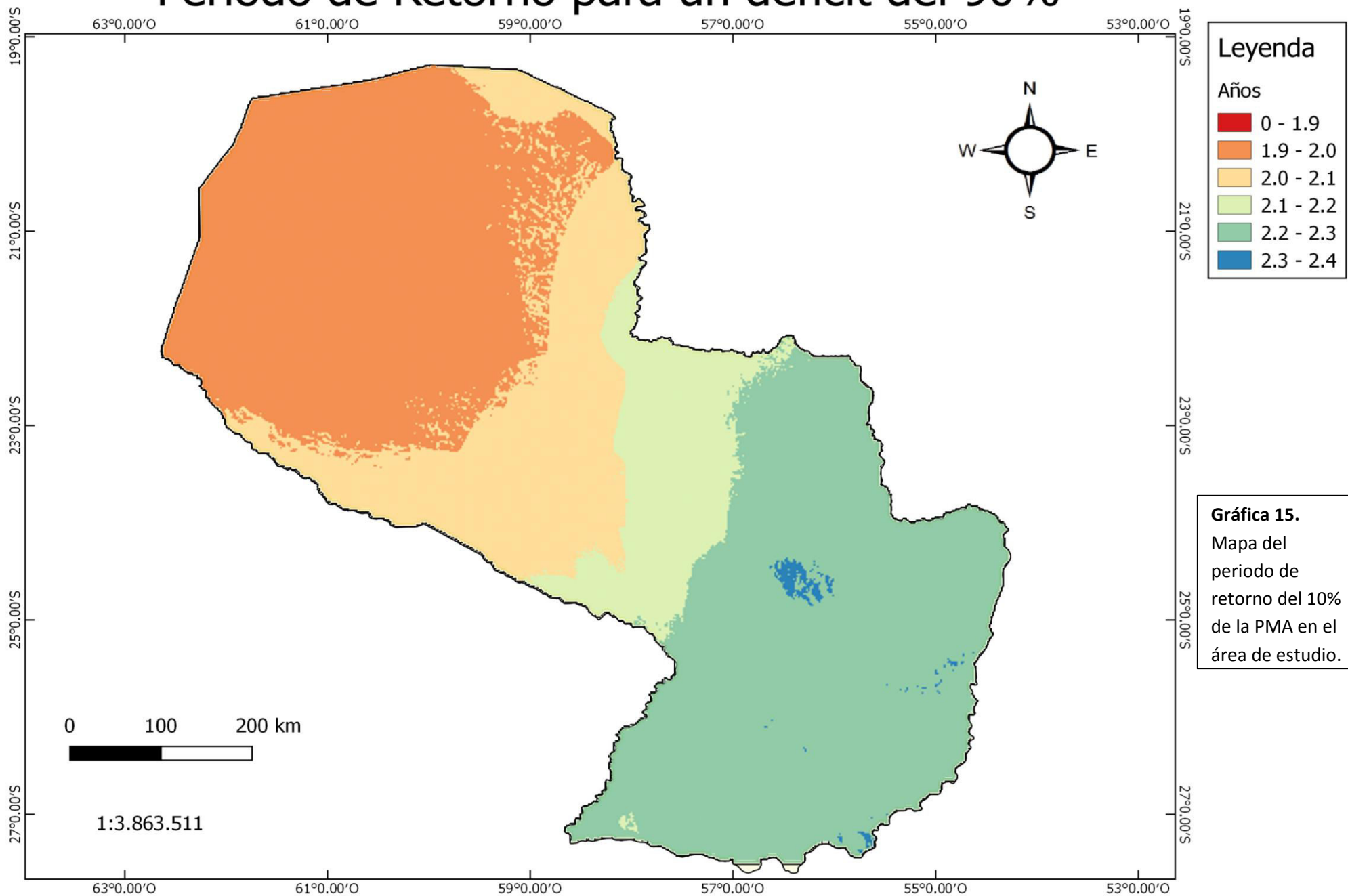
Se obtuvo un mapa de periodo de retorno del 70%, esto implica un 30% de déficit de PMA. **Gráfico 17.**

Se obtuvo un mapa de periodo de retorno del 80%, esto implica un 20% de déficit de PMA. **Gráfica 18.**

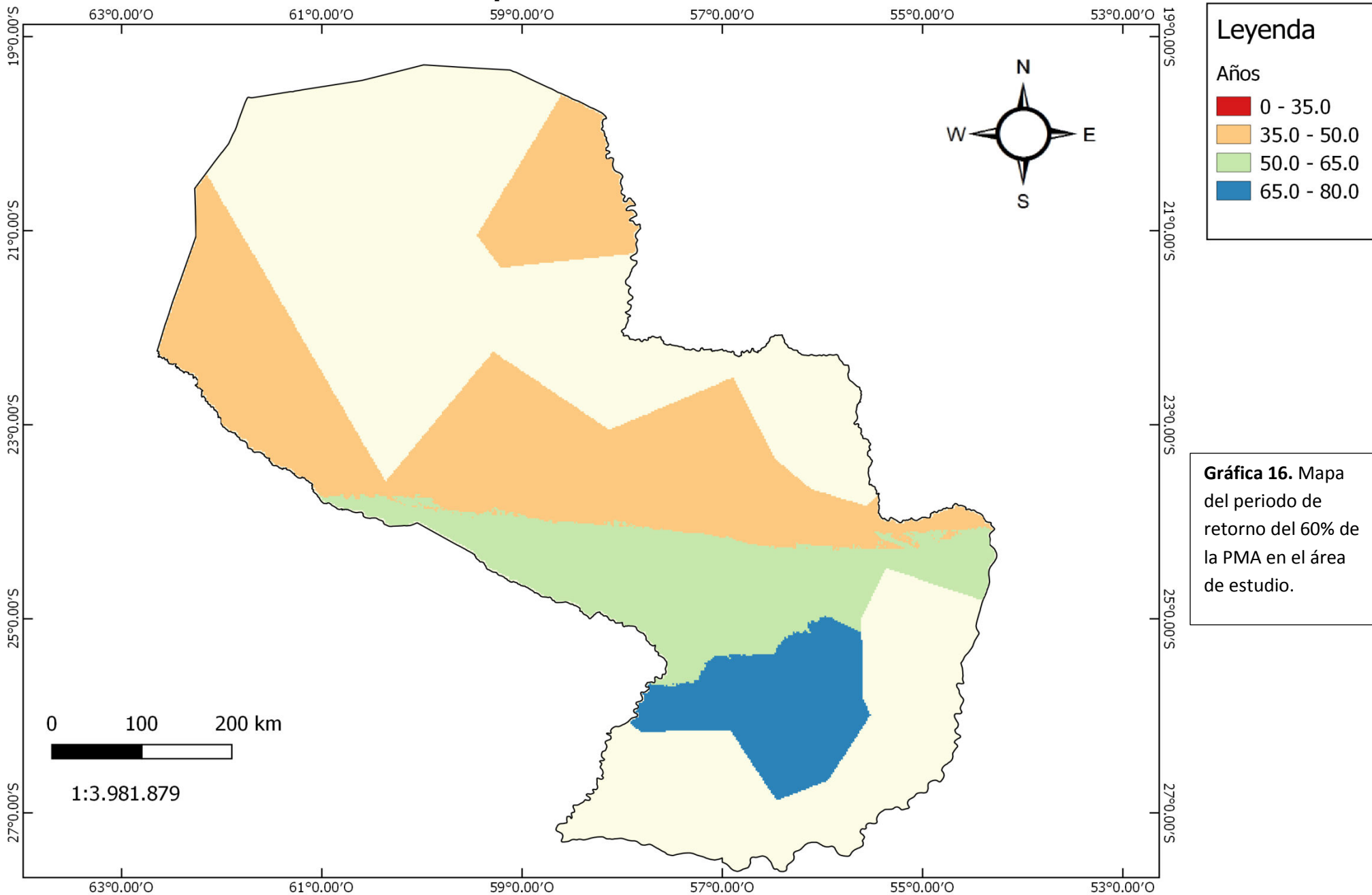
Se obtuvo un mapa de periodo de retorno del 90%, esto implica un 10% de déficit de PMA. **Gráfica 19.**

Se obtuvo un mapa de periodo de retorno del 120%, esto implica un 20% de exceso de PMA. **Gráfica 20.**

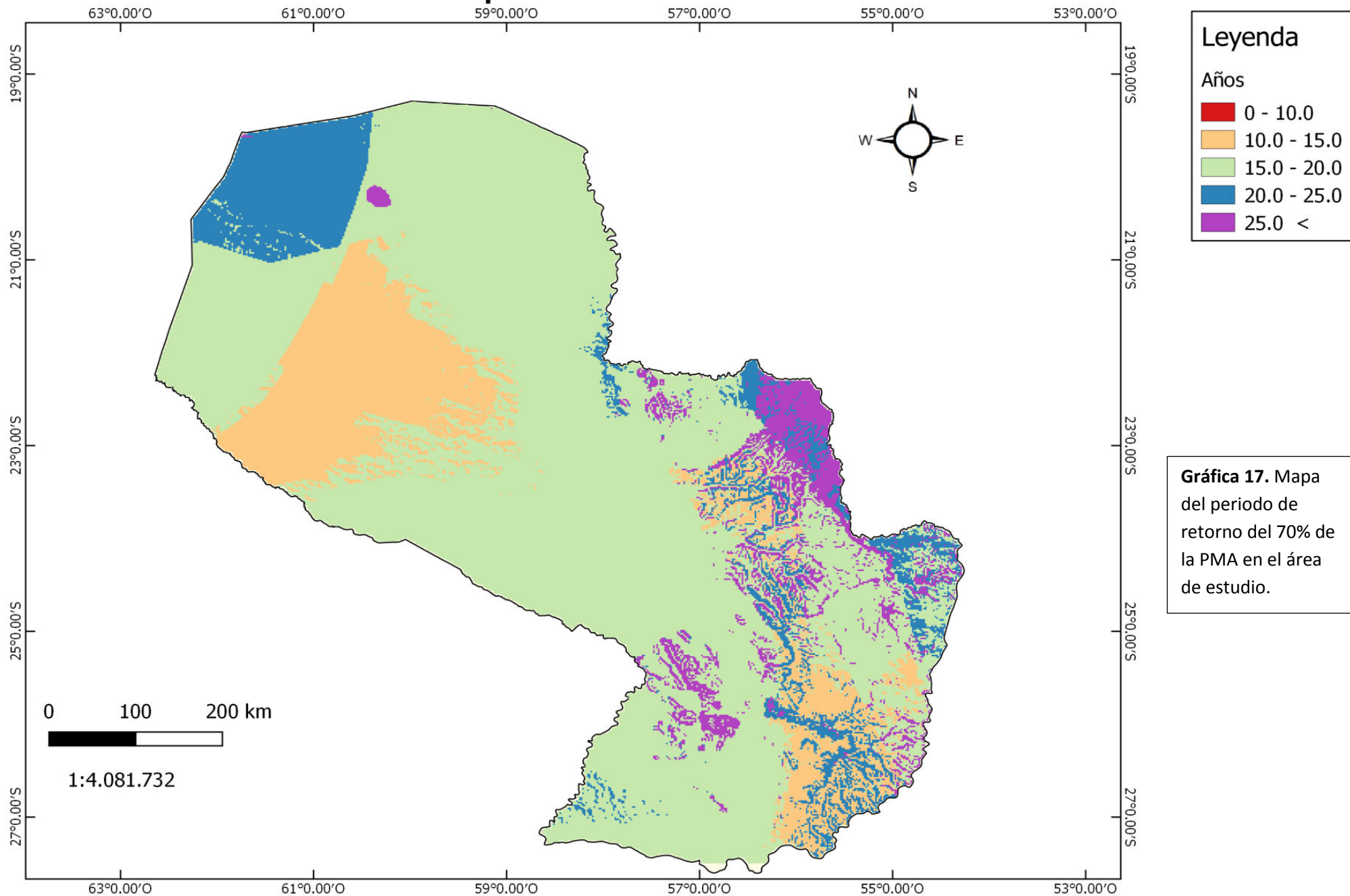
Periodo de Retorno para un deficit del 90%



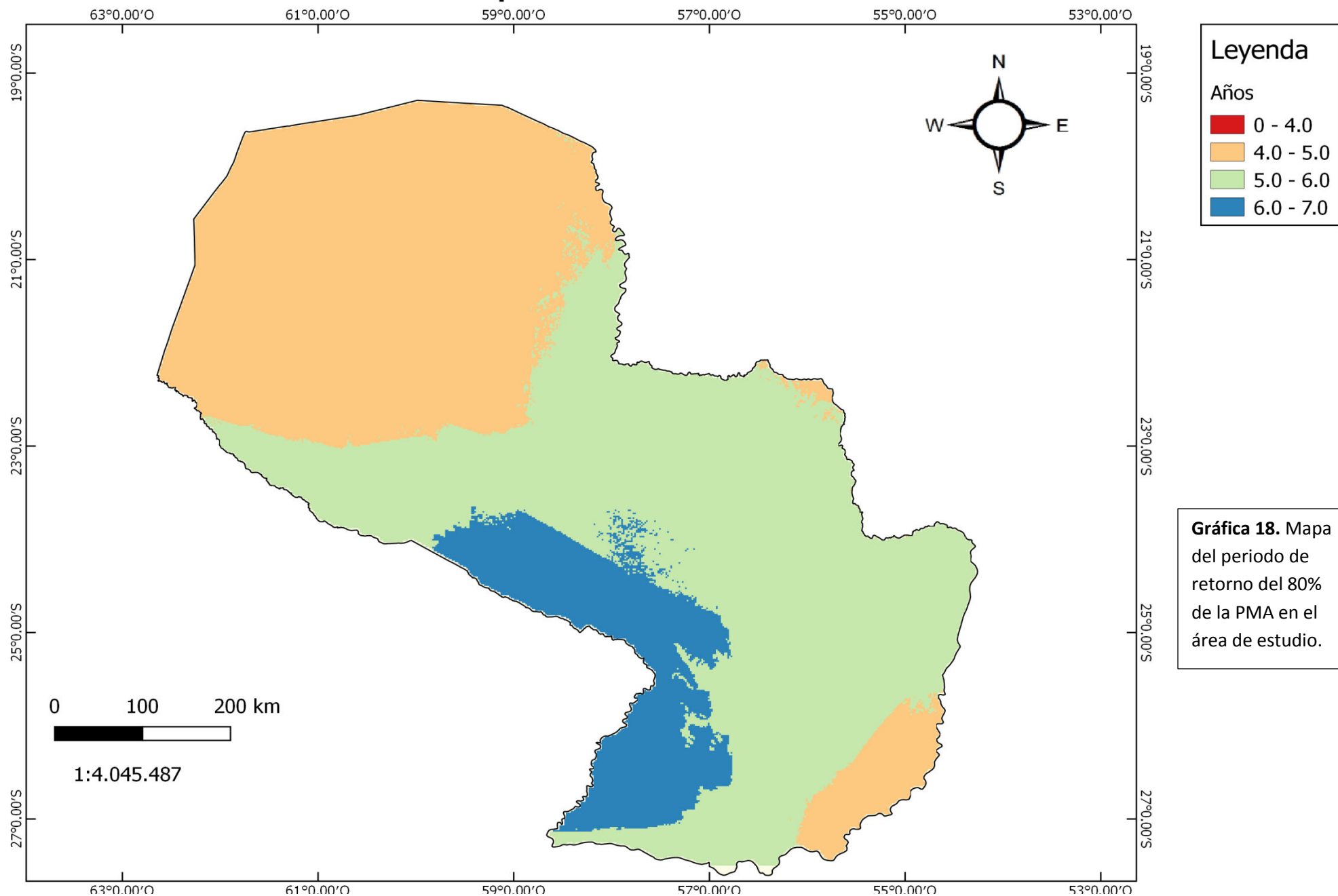
Periodo de Retorno para un déficit del 40%



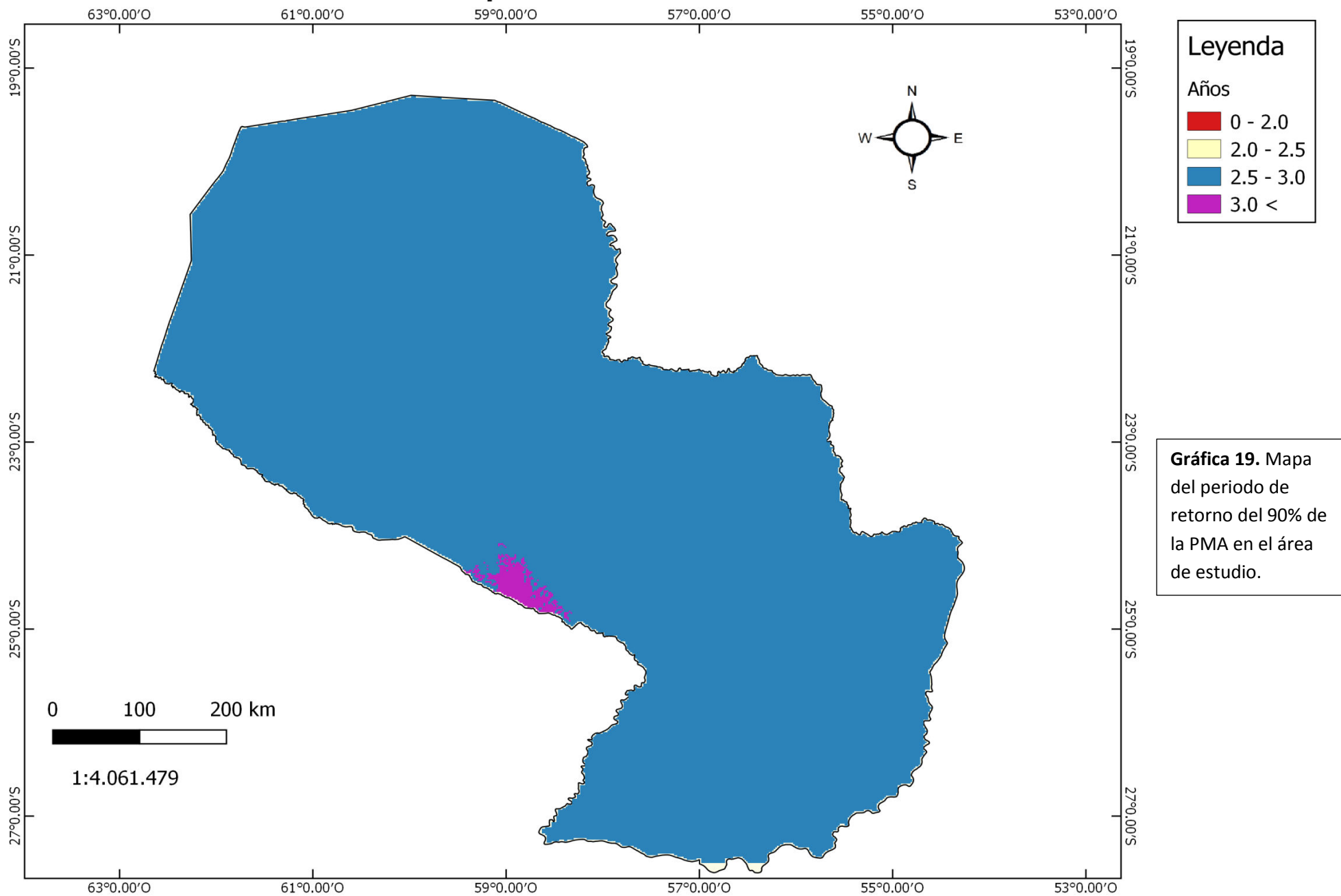
Periodo de Retorno para un déficit del 30%



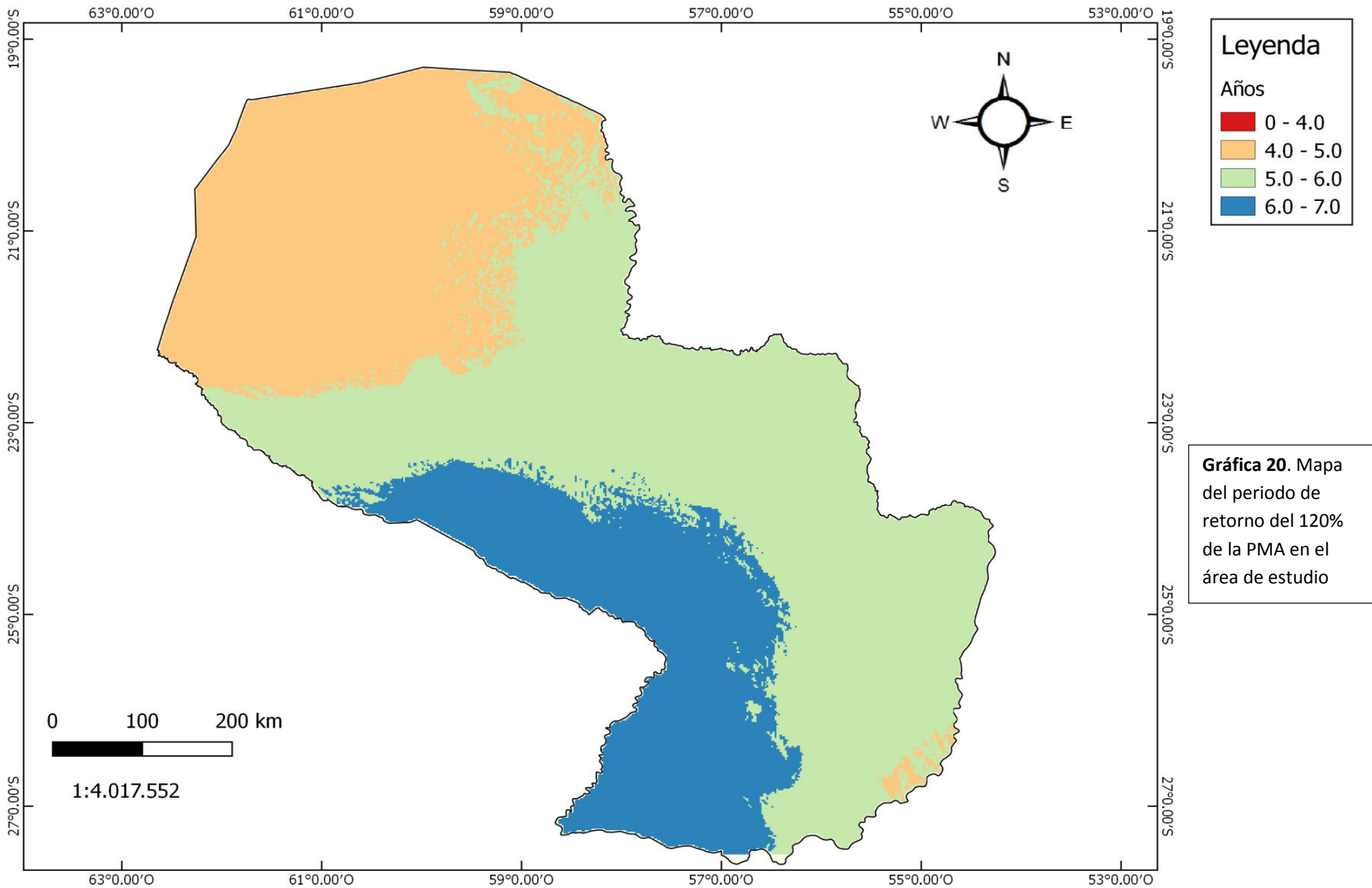
Periodo del Retorno para un déficit del 20%



Periodo de Retorno para un déficit del 10%



Periodo de Retorno para un exceso del 20%



6. Conclusiones

En las gráficas 11 y 12 se aprecia los valores obtenidos para el Índice de Estacionalidad (SI) y el Día Juliano Medio (MJD), con relación a la Precipitación Media Anual (PMA) (gráfica 36). De esto se puede concluir dos aspectos importantes. El primero está basado en los valores de SI, el criterio que se usa para agrupar estaciones en sub-regiones homogéneas es: grupo 1, estaciones con SI entre 0 y 0,2; grupo 2, con SI entre 0,2 y 0,6 y grupo 3, con SI mayor a 0,6. El segundo está basado en función de su MJD (Día Juliano Medio), en el caso de que las estaciones analizadas se encuentren todas dentro de un mismo rango de SI, debido a que puede existir estaciones con SI similar, pero cuya concentración ocurre en distintas épocas (zonas con precipitaciones concentradas en verano y otras en invierno).

Teniendo en cuenta lo citado arriba, podemos agrupar las estaciones en una región homogénea, esto se puede deber a la poca cantidad de estaciones que se poseen en la región Occidental que puede influenciar a la hora del análisis de homogeneidad. Ya que la Región Oriental y Occidental tienen diferencias marcadas en su climatología. Por eso se hace la siguiente aclaración: para este trabajo la homogeneidad de la región no es geográfica, sino homogeneidad desde el punto de vista del modelo de probabilidad parental. De este modo, cobra relevancia la forma en que las estaciones se agrupen para poder generar estas "regiones homogéneas".

En las gráficas 13 y 14 se aprecia que existe un mayor rango de variación en la curva de crecimiento. Esto implica una mayor variabilidad en los cuantiles regionales y en los de cada sitio. Esto tiene relación con los valores de precipitación media anual, que se incrementan en el mismo sentido. Es decir, la variabilidad, medida en términos de L-Cv, L-Skewness y L-Kurtosis, disminuye desde las zonas más secas a las zonas más húmedas.

Cuando se tiene un periodo de retorno del 10% de la PMA (gráfica 37), esto implica un 90% de déficit de PMA. Se observa en el gráfico 15, que la zona más afectada sería la Región Occidental, en especial los departamentos de Boquerón, Alto Paraguay, y el centro-norte de Presidente Hayes. Con un promedio de 1 a 2 años aproximadamente. En la Región Oriental tardaría en producirse dicho déficit alrededor de 2 a 3 años aproximadamente.

Cuando se tiene un periodo de retorno del 60% de la PMA (gráfica 37), esto implica un 40% de déficit de PMA. Se observa en el gráfico 16, las zonas blancas con periodos de retornos superiores a los 100 años, que para cualquier actividad que se pueda desarrollar, es un periodo excesivamente largo y por ende no es útil para un estudio a corto o mediano plazo. En ese gráfico se aprecia las zonas más afectadas por dicho déficit, están ubicados al noreste del departamento de Alto Paraguay y el noroeste del departamento de Boquerón, ambas en la Región Occidental. Se tiene un promedio de alrededor de 35 a 50 años. Los departamentos del centro y sur de la región Oriental tienen un promedio de alrededor de 65 y 80 años. Los que están ubicados en el este y norte tienen un promedio de alrededor de 40 y 50 años.

Para un periodo de retorno del 70% de la PMA (gráfica 37), esto implica un 30% de déficit de dicha precipitación. En el gráfico 17, se resaltan las zonas más afectadas por dicho déficit, ubicadas en el noreste del departamento de Boquerón con un promedio de alrededor de 20 a 55 años y también parte del departamento de Presidente Hayes, con un promedio de alrededor de 10 a 15 años, ambos en la región Occidental. En la Región Oriental no resulta tan notorio las zonas que serían afectadas por este déficit, en el sur donde se ubican los departamentos de Caazapá y Encarnación, tienen un promedio de alrededor de 10 a 15 años. El departamento de Amambay se ve afectado en promedio alrededor de 20 a 26 años.

Para un periodo de retorno del 80% de la PMA (gráfica 37), esto implica un 20% de déficit de dicha precipitación. Se observa en el gráfico 18 que gran parte de ambas regiones en el Paraguay, tendría este déficit en promedio de 5 a 6 años aproximadamente.

En la zona del el centro-norte de Presidente Hayes y el centro - sur de la Región Oriental tendríamos este exceso alrededor de 6 a 7 años. En la región Occidental los departamentos más afectados serían Boquerón y Alto Paraguay.

Para un periodo de retorno del 90% de la PMA (gráfica 37), esto implica un 10% de déficit de dicha precipitación. En la gráfica 19, con dicho déficit es muy ínfima la diferencia que se puede encontrar en ambas regiones del país para decir que zona sería más perjudicada. Y esto ocurría en promedio de 2 a 3 años aproximadamente.

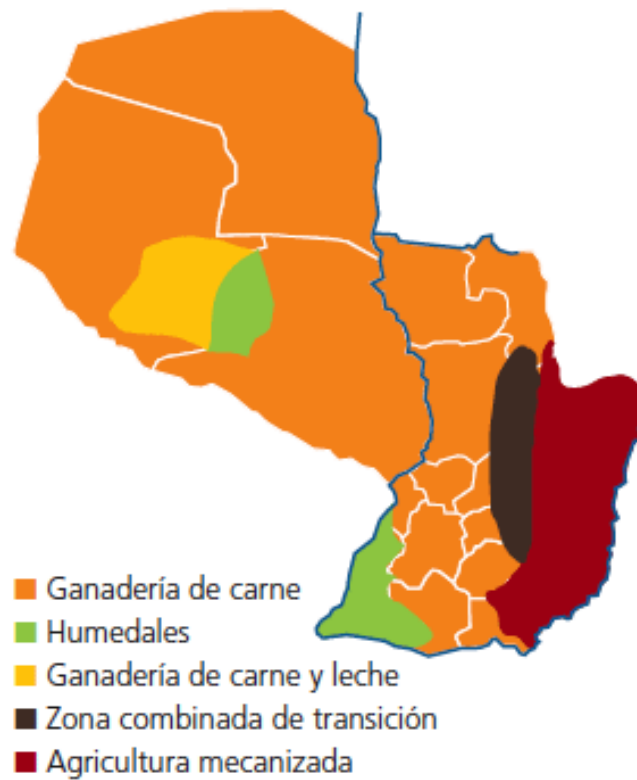
Para un periodo de retorno del 120% de la PMA (gráfica 37), esto implica un 20% de exceso de dicha precipitación. En la gráfica 20, que la zona más afectada sería la Región Occidental, en especial los departamentos de Boquerón, Alto Paraguay. Con un promedio de 4 a 5 años aproximadamente. En la zona del el centro-norte de Presidente Hayes y el centro - sur de la Región Oriental tendríamos este exceso alrededor de 6 a 7 años.

En este trabajo se analizaron los datos de precipitación de las 22 estaciones meteorológicas ubicadas en ambas regiones del Paraguay, se obtuvieron los valores calculados de los L-momentos para el territorio paraguayo, que se relacionan directamente con la obtención de los periodos de retorno de los diferentes umbrales de déficit de la precipitación media anual.

Con la metodología analizada el principal obstáculo encontrado fue la insuficiente cantidad de datos históricos de precipitación para la obtención de resultados que reflejen de manera más correcta los diferentes valores de los periodos de retorno. Teniendo en cuenta que los valores de los L-momentos no presentaron una variación muy significativa con los datos analizados, apenas de 0 a 0,20; y sus umbrales van de 0 a 1. Pero con lo mencionado anteriormente se puede concluir que la metodología fue aplicable a Paraguay obteniendo resultados interesantes.

Se puede clasificar las sequías más severas cuando las lluvias tienen un porcentaje muy bajo con relación a la PMA; tenemos que el Chaco paraguayo y el sur de la Región Oriental son los más afectados por eventos de menor tiempo de retorno.

En el gráfico 21 se puede ver cómo está distribuida la utilización de la tierra y esto da una idea que zonas son más propensas a sufrir los efectos de los eventos de sequía.



Gráfica 21. Distribución del uso de la tierra en Paraguay. (Fuente: [http://tierra y conocimiento.inbio.org.py/la gente que siembra el futuro.html](http://tierra_y_conocimiento.inbio.org.py/la_gente_que_siembra_el_futuro.html))

6.1. Recomendaciones para trabajos futuros

Para optimizar los resultados será necesario ensanchar el periodo de análisis de datos. Ya que para un estudio de la precipitación, es transcendental contar con el elemento esencial que son los datos, por ende es preciso tener una mayor cantidad de estaciones meteorológicas.

Para profundizar en dicho método, es necesario modificar o agregar pasos en el script utilizado en R, con el propósito de obtener resultados más detallados con un tamaño muestral más pequeño.

7. ANEXO

7.1. ANEXO I: Formato de Entrada de los Datos

7.1.1. Control de calidad

1) rclimindex_extraQC.r.

El mismo debe estar en el formato internacionalmente conocido como “formato Rclimindex” que responde a las siguientes especificaciones:

- a) Un registro (fila) por día
- b) Seis campos (columnas) por registro: año (YYYY), mes (MM), día (DD), precipitación acumulada diaria (en mm, 0.1 para IP), temperatura máxima diaria (°C, con un valor decimal), temperatura mínima diaria (°C, con un valor decimal).
- c) Los campos deben ir delimitados por tabulador (otros delimitadores como espacio o incluso comas son tolerables, pero problemáticos).
- d) Los valores ausentes deben ser consignados como 99.9 y si para un registro (día) no existe ninguno de los tres valores (precipitación, temperatura máxima, temperatura mínima) no es necesario incluirlo.

	A	B	C	D	E	F
1	1959	1	1	-99.9	-99.9	-99.9
2	1959	1	2	-99.9	-99.9	-99.9
3	1959	1	3	-99.9	-99.9	-99.9
4	1959	1	4	-99.9	-99.9	-99.9
5	1959	1	5	-99.9	-99.9	-99.9
6	1959	1	6	-99.9	-99.9	-99.9
7	1959	1	7	-99.9	-99.9	-99.9
8	1959	1	8	-99.9	-99.9	-99.9
9	1959	1	9	-99.9	-99.9	-99.9
10	1959	1	10	-99.9	-99.9	-99.9
11	1959	1	11	-99.9	-99.9	-99.9
12	1959	1	12	-99.9	-99.9	-99.9
13	1959	1	13	-99.9	-99.9	-99.9
14	1959	1	14	-99.9	-99.9	-99.9
15	1959	1	15	-99.9	-99.9	-99.9
16	1959	1	16	-99.9	-99.9	-99.9
17	1959	1	17	-99.9	-99.9	-99.9
18	1959	1	18	-99.9	-99.9	-99.9
19	1959	1	19	-99.9	-99.9	-99.9
20	1959	1	20	-99.9	-99.9	-99.9
21	1959	1	21	-99.9	-99.9	-99.9
22	1959	1	22	-99.9	-99.9	-99.9
23	1959	1	23	-99.9	-99.9	-99.9
24	1959	1	24	-99.9	-99.9	-99.9
25	1959	1	25	-99.9	-99.9	-99.9

Gráfica 22. Ejemplo de formato de datos de entrada para rclimdex_extraQC.r.

2) ANDESQC

El fichero **masterqc** responde a las siguientes especificaciones:

- a) Columnas en las siguientes secuencias: nombre fichero, año, mes, día, precip, tx, tn.
- b) Los campos deben ir delimitados por tabulador (otros delimitadores como espacio o incluso comas son tolerables, pero problemáticos).
- c) Los valores ausentes deben ser consignados como -99.9. Los registros de datos deben estar en orden cronológico. No se permite fechas faltantes.

Ejemplo de formato de datos de entrada para ANDESQC

	A	B	C	D	E	F	
1	1959	1	1	-99.9	-99.9	-99.9	Año
2	1959	1	2	-99.9	-99.9	-99.9	
3	1959	1	3	-99.9	-99.9	-99.9	
4	1959	1	4	-99.9	-99.9	-99.9	Mes
5	1959	1	5	-99.9	-99.9	-99.9	
6	1959	1	6	-99.9	-99.9	-99.9	
7	1959	1	7	-99.9	-99.9	-99.9	Día
8	1959	1	8	-99.9	-99.9	-99.9	
9	1959	1	9	-99.9	-99.9	-99.9	
10	1959	1	10	-99.9	-99.9	-99.9	Precipitación
11	1959	1	11	-99.9	-99.9	-99.9	
12	1959	1	12	-99.9	-99.9	-99.9	
13	1959	1	13	-99.9	-99.9	-99.9	Temperatura mínima
14	1959	1	14	-99.9	-99.9	-99.9	
15	1959	1	15	-99.9	-99.9	-99.9	
16	1959	1	16	-99.9	-99.9	-99.9	Temperatura máxima
17	1959	1	17	-99.9	-99.9	-99.9	
18	1959	1	18	-99.9	-99.9	-99.9	
19	1959	1	19	-99.9	-99.9	-99.9	
20	1959	1	20	-99.9	-99.9	-99.9	
21	1959	1	21	-99.9	-99.9	-99.9	
22	1959	1	22	-99.9	-99.9	-99.9	
23	1959	1	23	-99.9	-99.9	-99.9	
24	1959	1	24	-99.9	-99.9	-99.9	
25	1959	1	25	-99.9	-99.9	-99.9	

Gráfica 23. Ejemplo de formato de datos de entrada para ANDESQC

3) RCLimTool

Los archivos que vayan a ser utilizados deben estar en formato CSV (delimitado por comas). Se deben emplear bases diferentes para cada una de las variables, con las estaciones que se deseen analizar. Estas bases deben cumplir con los siguientes aspectos:

- a) Columnas en las siguientes secuencias: day, month, year seguido de los nombres de las estaciones. *NOTA: unidades de Precipitación= milímetros y unidades de Temperatura= grados Celsius.*
- b) Para los casos en el que se presenten datos faltantes, se deben codificar como NA; los registros de datos deben estar en orden cronológico. No se permite fechas faltantes.

14-INV-338

1	day	month	year	aja	bne	pgi
2	1	1	1959		0	0 NA
3	2	1	1959		0	0 NA
4	3	1	1959	9.7	9.7	NA
5	4	1	1959	15.4	15.4	NA
6	5	1	1959	0.5	0.5	NA
7	6	1	1959	16.5	16.5	NA
8	7	1	1959		0	0 NA
9	8	1	1959		0	0 NA
10	9	1	1959		0	0 NA
11	10	1	1959	35.5	35.5	NA
12	11	1	1959		0	0 NA
13	12	1	1959		0	0 NA
14	13	1	1959		0	0 NA
15	14	1	1959		0	0 NA
16	15	1	1959		0	0 NA
17	16	1	1959	0.3	0.3	NA
18	17	1	1959	0.8	0.8	NA
19	18	1	1959		0	0 NA
20	19	1	1959		0	0 NA
21	20	1	1959		0	0 NA
22	21	1	1959		0	0 NA
23	22	1	1959		0	0 NA
24	23	1	1959		0	0 NA
25	24	1	1959		0	0 NA

Gráfica 24. Ejemplo de formato de datos de entrada para RClimTool.

7.1.2. Análisis Regional de Frecuencia basados en L-momentos

Para poder ejecutar el script para obtener los diferentes resultados en el Análisis Regional de Frecuencia basados en L-momentos se deben preparar dos planillas una con los datos de precipitación y otra con los datos de las estaciones a ser utilizadas.

Los archivos que vayan a ser utilizados deben estar en formato CSV (delimitado por comas). Estas bases deben cumplir con los siguientes aspectos:

- ✓ Base de datos de estaciones
- a) Columnas en las siguientes secuencias: id_estacion (código de la estación), nombres de las estaciones, país, latitud, longitud, elevación.

id_estacion	nombre_estacion	País	Latitud	Longitud	Ele
st-py-011	Adrian_Jara	Paraguay	-19.53	-59.37	143
st-py-033	Bahia_Negra	Paraguay	-20.23	-58.17	82
st-py-065	Pratts_Gill	Paraguay	-22.56	-61.56	206
st-py-068	Mcal_Estigarribia	Paraguay	-22.03	-60.62	167
st-py-086	Puerto_Casado	Paraguay	-22.28	-57.94	78
st-py-097	Pedro_Juan_Caballero	Paraguay	-22.64	-55.83	563
st-py-128	Pozo_Colorado	Paraguay	-23.5	-58.79	98
st-py-134	Concepción	Paraguay	-23.44	-57.43	75
st-py-170	Gral_Bruguez	Paraguay	-24.74	-58.84	89
st-py-185	San_Pedro	Paraguay	-24.09	-57.09	81
st-py-192	San_Estanislao	Paraguay	-24.67	-56.46	183
st-py-210	Salto_Guaira	Paraguay	-24.03	-54.35	297
st-py-218	Asuncion_Aeropuerto	Paraguay	-25.24	-57.52	83
st-py-221	Paraguarí	Paraguay	-25.63	-57.15	116
st-py-233	Villarrica	Paraguay	-25.75	-56.44	163
st-py-234	Cnel_Oviedo	Paraguay	-25.25	-54.45	159
st-py-246	Aeropuerto_Guarani	Paraguay	-25.46	-54.84	247
st-py-251	Quyquyó	Paraguay	-26.22	-56.99	141
st-py-255	Pilar	Paraguay	-26.88	-58.32	58
st-py-260	San_Juan_Bautista_Misiones	Paraguay	-26.67	-57.13	131
st-py-268	Caazapá	Paraguay	-26.18	-56.36	142
st-py-285	Capitan_Meza	Paraguay	-26.83	-55.33	263
st-py-297	Encarnación	Paraguay	-27.3	-55.9	90

Tabla 9. Ejemplo de formato de datos de entrada de la Base de datos de estaciones.

- ✓ Base de datos de Registros
 - a) Un registro (fila) por año.
 - b) Columnas en las siguientes secuencias: id_estacion (código de la estación), año, mes. *NOTA: unidades de Precipitación= milímetros.*
 - c) Para los casos en el que se presenten datos faltantes, se deben codificar como NA; los registros de datos deben estar en orden cronológico. No se permite fechas faltantes.

En forma de ejemplo se puso solo la estación que se ubica en la localidad de Adrián Jara, se puede observar cómo es el formato que deben tener los datos a ser analizados.

st-py-011	1961	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
st-py-011	1962	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
st-py-011	1963	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
st-py-011	1964	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
st-py-011	1965	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
st-py-011	1966	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
st-py-011	1967	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
st-py-011	1968	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
st-py-011	1969	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
st-py-011	1970	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
st-py-011	1971	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
st-py-011	1972	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
st-py-011	1973	NA	NA	NA	NA	40.8	67.9	18.3	33.9	3	95.7	146.7	113.4
st-py-011	1974	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
st-py-011	1975	38.6	101.1	111.9	110.4	2.4	129.1	41.1	10	NA	24.4	50	88.8
st-py-011	1976	68.9	NA	NA	68.2	22.2	5.3	0	22.7	NA	201.1	NA	NA
st-py-011	1977	192.6	194.6	134.7	0	62.4	48.8	3.8	34.2	78	89	126.2	0
st-py-011	1978	155	198.3	108.6	64.3	7.2	NA	0	NA	55	208	100.2	108.1
st-py-011	1979	349.8	50.1	215.8	81	10.4	2.2	11.5	45.7	27.6	79.2	103.9	119.9
st-py-011	1980	NA	NA	NA	NA	54	101.2	0	NA	NA	NA	NA	113.4

Tabla 10. Ejemplo de formato de datos de entrada de la Base de datos de Registros.

7.2. ANEXO II: GRÁFICAS De Los Resultados

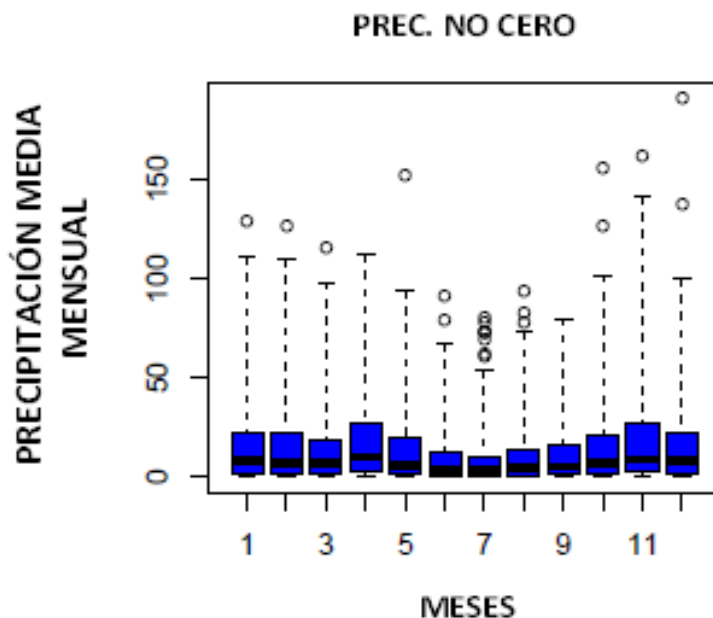
7.2.1. Control de calidad

Se detalla a continuación los resultados obtenidos de la variable precipitación que arrojaron para la estación 86218, que corresponde al Aeropuerto Internacional Silvio Pettirossi.

1) rclimindex_extraQC.r.

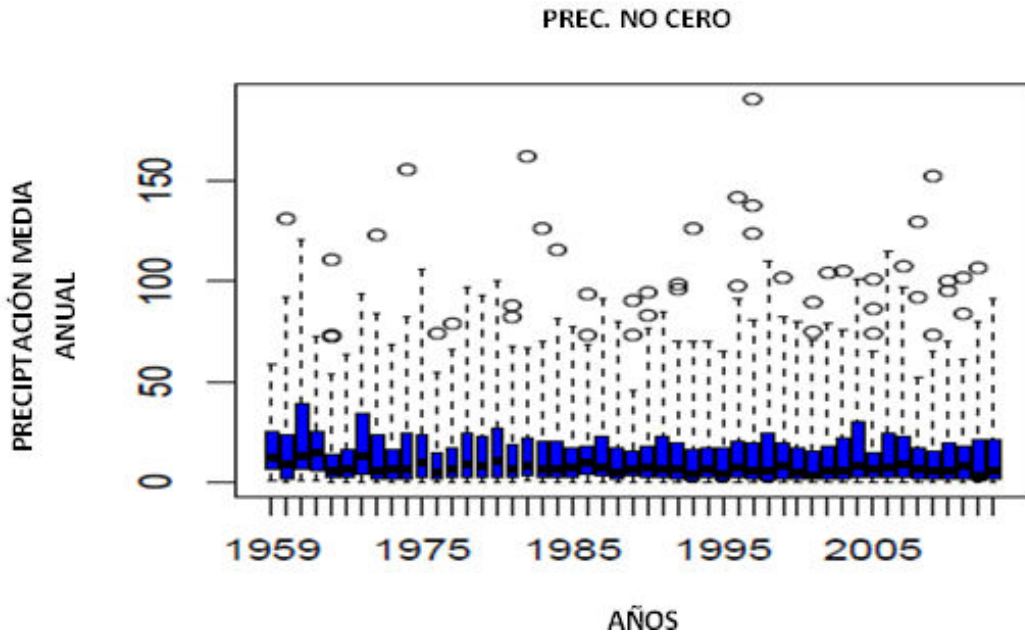
En el directorio aparecerán los siguientes ficheros:

- a) Tres ficheros con gráfico:
 - i. ra0086128_boxex.pdf: diagramas box-plot mensuales para el conjunto de la serie, para precipitación por encima de 0 mm. Aparecen identificados en forma de círculo aquellos valores que superan el tercer cuartil + 3 veces el rango intercuartílico (5 en precipitación) o los que no alcanzan el primer cuartil – 3 veces el rango intercuartílico (5 en precipitación). Los datos así resaltados representan outliers.



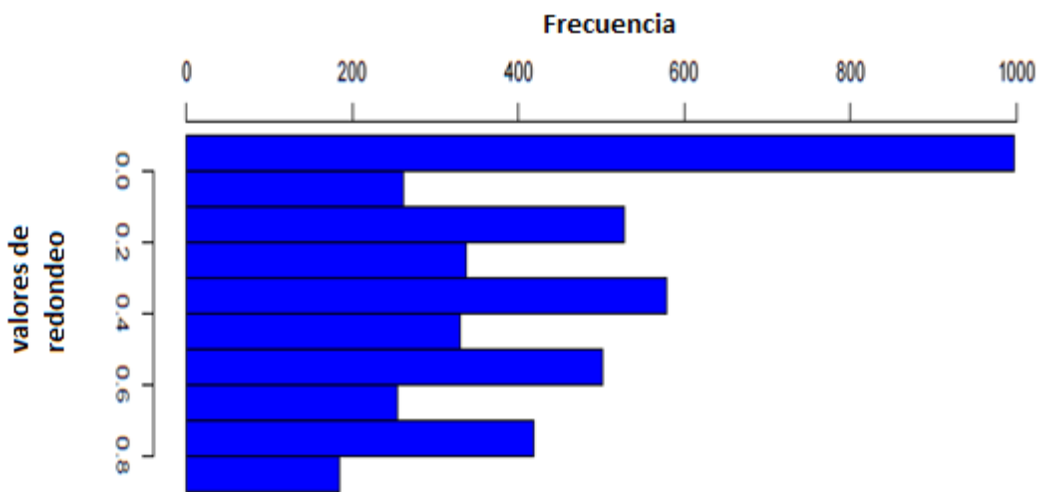
Gráfica 25. Serie para la precipitación mensual por encima de 0 mm.

- ii. ra0086128_boxseries.pdf: diagramas de caja para las mismas variables, pero graficados como series anuales.



Gráfica 26. Serie para la precipitación anual por encima de 0 mm.

- iii. ra0086128_rounding.pdf: diagramas de frecuencia de los 10 valores posibles del punto decimal. En precipitación se omiten los valores 0 y permite comprobar si las series están redondeadas.



Gráfica 27. Gráfica de los diferentes valores de redondeo.

Como se puede observar en el Gráfico se redondea a 0.0 mm casi la mayoría de los datos.

b) Otros ficheros de texto

- i. miestacion_duplicates.txt: volcado de fechas duplicadas. En caso de que una combinación de año/mes/día aparezca más de una vez, será mostrada en este fichero.

De la siguiente manera se presenta el fichero:

año	mes	día	precipitación

Tabla 11. Tabla de fechas duplicadas.

Con este control no se encontraron datos duplicados.

- ii. miestacion_toolarge.txt: muestra los valores considerados excesivamente elevados.

año	mes	día	precipitación
2008	1	11	129
1983	2	15	126,4
2006	3	24	115
2009	5	23	152,2
1977	6	19	78,8
2013	6	14	91,6
1967	7	14	73,3
1967	7	20	72,5
1978	7	1	80,4
1983	7	26	61,7
1985	7	30	77,7
2009	7	7	73,4
2010	7	9	69,7
2011	7	9	61,4
1974	8	3	78,4
1990	8	18	83
1990	8	25	94
1974	10	15	155,2
1993	10	17	126
1982	11	19	161,7
1997	12	13	137,7
1997	12	22	190,8

Tabla 12. Tabla de valores elevados de precipitación.

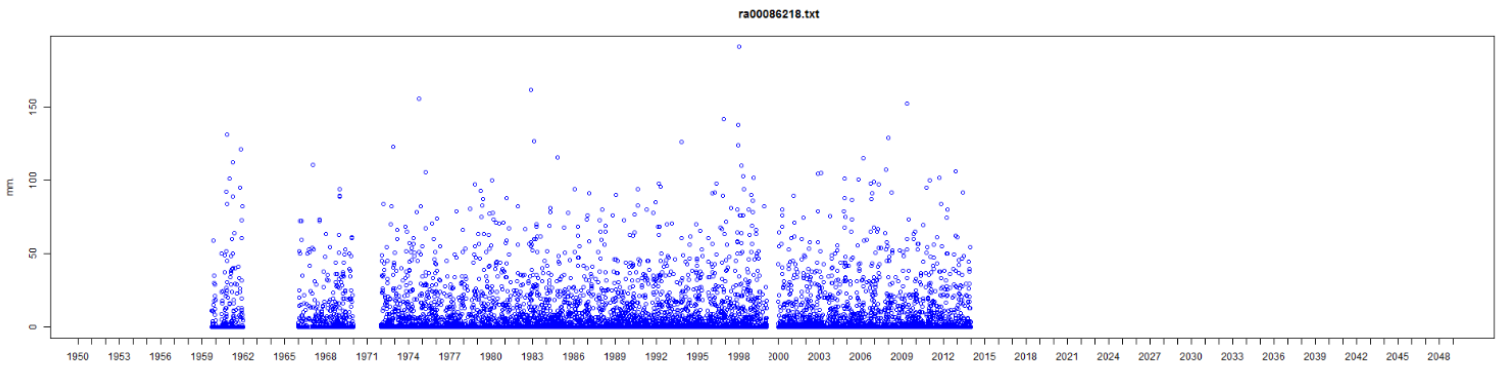
2) ANDESQC

En el directorio aparecerán los siguientes ficheros:

a) Tres ficheros con gráfico.

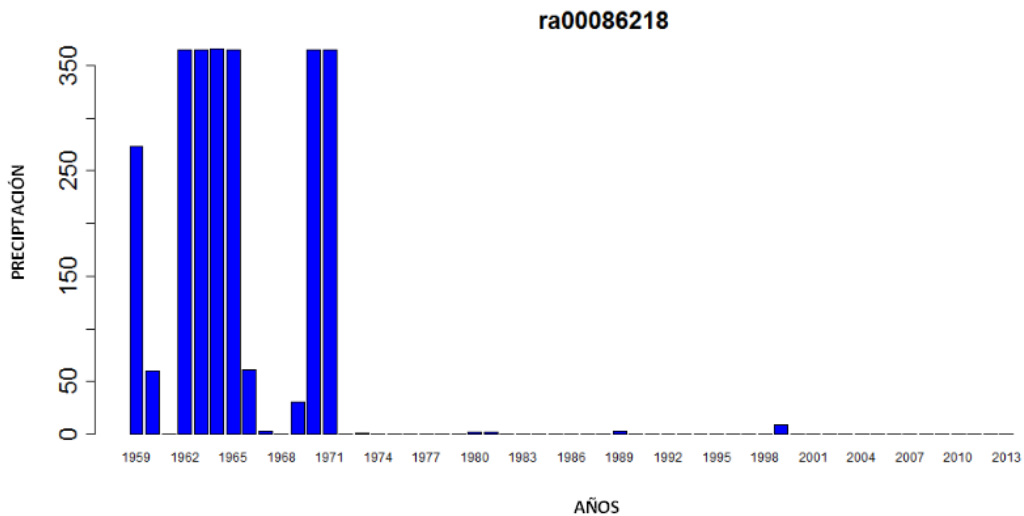
i. Visión general de la serie.

Te permite ver el comportamiento temporal de la serie de datos.



Gráfica 28. Serie completa de datos.

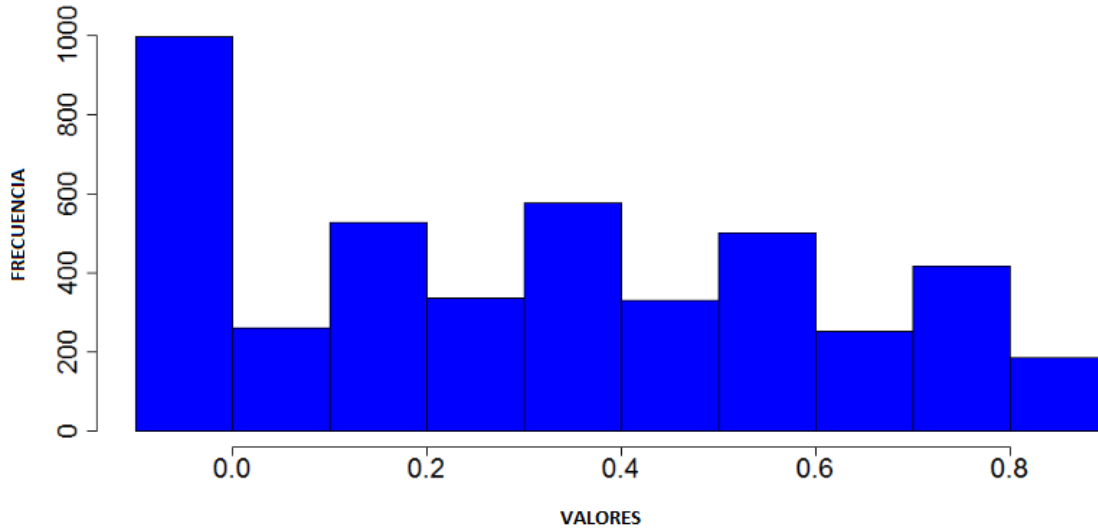
ii. Valores faltantes



Gráfica 29. Gráfica de Valores faltantes.

Muestra los años donde faltan los valores de precipitación.

iii. Control de redondeo



Gráfica 30. Gráfica de los diferentes valores de redondeo.

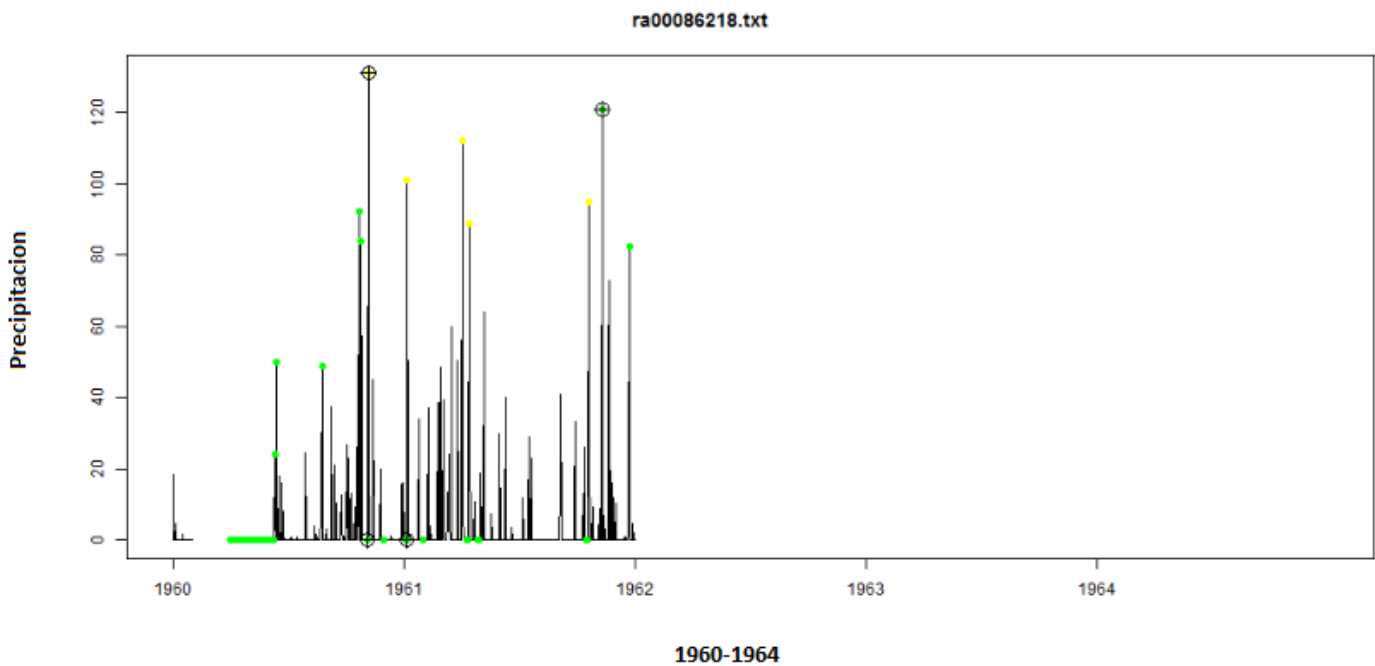
Como se puede observar en el gráfico 30 se tiende a redondear los valores de la precipitación medidos a 0.0 mm.

b) Contenido output (salida) por períodos de cinco años.

En este apartado se verifica que tests fallaron al analizar los valores de precipitación de cada estación meteorológica, esto son los llamados valores atípicos, que no siempre son errores.



Gráfica 31. Gráfica a) y b) las diferentes representaciones de los tests fallados.



Gráfica 32. Serie de cuatro años de lo serie completa, donde se ve los tests errados.

3) RClimTool

En el directorio aparecerán los siguientes ficheros

a) Análisis Gráfico – Descriptivo

Una vez tenemos los datos cargados para todas las variables a analizar, procedemos a realizar el análisis descriptivo para cada una de ellas, para ello debemos especificar el período de análisis, útil si se desea analizar solo una sección de la serie, si por el contrario se desea analizar la serie completa entonces estos campos deben estar vacíos.

i. Análisis descriptivo

Pre-informe. El siguiente informe analiza la variable precipitación para el período comprendido entre 1959-2014.

En la tabla 13 se observa el comportamiento de la serie temporal de los datos de precipitación. En la primera columna se tiene las estaciones con sus respectivos nombres; la siguiente cuantifica los datos analizados; en las siguientes se observa ya un análisis estadístico descriptivo de dicha serie de datos, con el fin de describir apropiadamente las diversas características de ese conjunto.

Estaciones	Valores Analizados	Media	Varianza	Desviación Estándar	Coficiente de variación	Datos Faltantes	Datos Faltantes en porcentaje
Adrián Jara	18628	2,815	96,117	9,804	348,275	5998	32,199
Bahía Negra	18628	2,815	96,117	9,804	348,275	5998	32,199
Pratts Gill	18628	1,789	63,873	7,992	446,733	7822	41,991
Mariscal Estigarribia	18628	2,112	78,073	8,836	418,366	427	2,292
Puerto Casado	18628	3,427	127,549	11,294	329,552	409	2,196
Pedro Juan Caballero	18628	4,437	144,089	12,004	270,537	1372	7,365
Pozo Colorado	18628	3,084	117,982	10,862	352,203	16103	86,445
Concepción	18628	3,715	154,034	12,411	334,079	42	0,225
General Bruguez	18628	3,311	133,98	11,575	349,591	11398	61,187
San Pedro	18628	3,997	182,887	13,524	338,343	12510	67,157
San Estanislao	18628	4,274	157,754	12,56	293,87	7360	39,51
Salto Guairá	18628	4,65	165,622	12,869	276,762	7074	37,975
Asunción (Aeropuerto)	18628	3,785	144,646	12,027	317,751	1575	8,455
Paraguarí	18628	4,004	161,033	12,69	316,93	12112	65,02
Villarrica	18628	4,559	179,828	13,41	294,143	33	0,177
Coronel Oviedo	18628	4,969	180,914	13,45	270,687	13362	71,731
Aeropuerto Guaraní	18628	4,935	176,492	13,285	269,2	3434	18,435
Pilar	18628	3,835	159,047	12,611	328,85	600	3,221
San Juan Bautista Misiones	18628	4,615	187,057	13,677	296,357	301	1,616
Caazapá	18628	4,466	165,426	12,862	287,994	5053	27,126
Capitán Meza	18628	4,945	199,156	14,112	285,385	1597	8,573
Encarnación	18628	4,837	196,934	14,033	290,124	951	5,105

Tabla 13. Tabla del Análisis descriptivo.

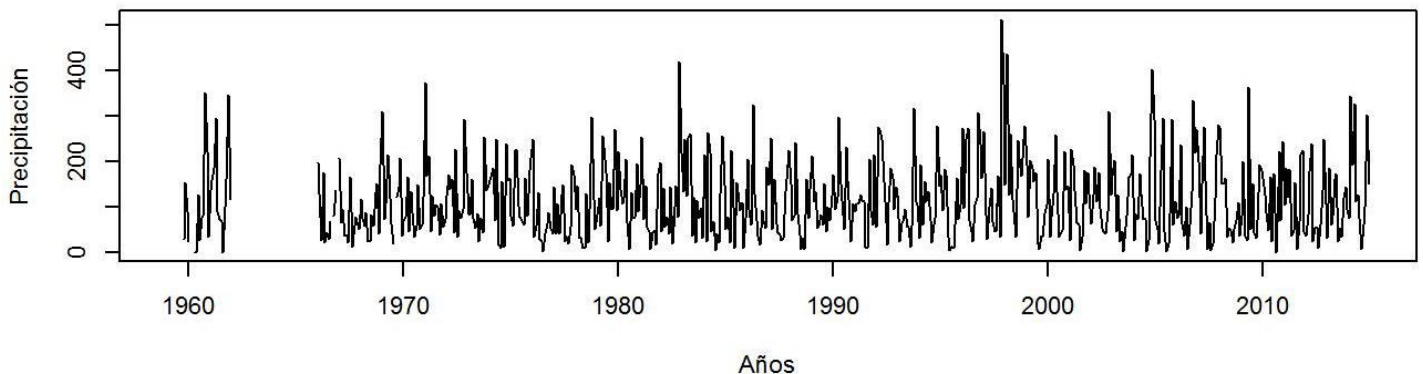
ii. Análisis Gráfico

Para el análisis gráfico se tiene la opción de generar diferentes tipos de *Gráficos automáticos*, los cuales se generan para todas las variables. Pueden generar gráficos diarios como mensuales.

Para la precipitación es mejor analizar los datos mensuales.

En este campo se genera estos tipos de gráficos.

✓ Gráficos Plot



Gráfica 33. Serie completa de los datos de precipitación.

Es una forma de visualizar los datos de la serie total.

En este tipo de gráfico podemos ver donde existen datos faltantes y que año es el más completo de la serie.

b) Control de Calidad

Ahora se realiza una validación para las diferentes estaciones en estudio, la cual consta del cálculo de varios criterios que ayudan a identificar datos atípicos y/o erróneos para su posterior corrección.

Para la identificación de datos atípicos se utilizaron 3 desviaciones estándar y se estableció un rango de valores permitidos $0 < \text{precip} < 300$, La última columna de la

siguiente tabla indica el % total de datos faltantes que serán llenados en la sección Datos faltantes. Se recomienda que este % no supere el 20% en cada una de las estaciones.

Estación	Porcentaje de datos atípicos	Porcentaje de datos fuera del rango	Porcentaje total de datos faltantes
Adrián Jara	1,69	0	0
Bahía Negra	1,69	0	0
Pratts Gill	1,14	0	0
Mariscal Estigarribia	1,97	0	0
Puerto Casado	2,43	0	0
Pedro Juan Caballero	2,57	0	0
Pozo Colorado	0,36	0	0
Concepción	2,29	0	0
General Bruguez	0,96	0	0
San Pedro	0,88	0	0
San Estanislao	1,48	0	0
Salto Guairá	1,57	0	0
Asunción (Aeropuerto)	2,29	0	0
Paraguarí	0,99	0	0
Villarrica	2,66	0	0
Coronel Oviedo	0,82	0	0
Aeropuerto Guaraní	2,1	0	0
Pilar	2,42	0	0
San Juan Bautista Misiones	2,44	0	0
Caazapá	1,75	0	0
Capitán Meza	2,26	0	0
Encarnación	2,42	0	0

Tabla 14. Tabla de validación.

7.2.2. Análisis Regional de Frecuencia basados en L-momento

- ✓ L-Momentos
- Se realizan los cálculos de los diferentes L-momentos

En la tabla se observa los diferentes valores obtenidos para los L-momentos en cada estación analizada.

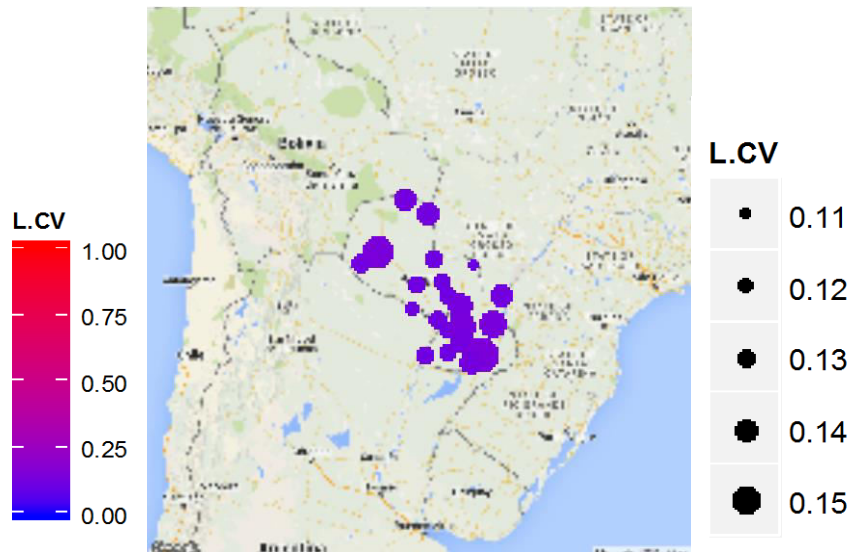
Identificación de cada estación	Nombre de cada estación	L-coeficientes de variación (L-CV)	L-Asimetría (L-Skewness)	L-Curtosis (L-Kurtosis)
st-py-011	Adrián Jara	0,13021	0,14002	0,13464
st-py-033	Bahía Negra	0,12791	0,02301	0,13945
st-py-065	Pratts Gill	0,12428	0,00351	0,09173
st-py-068	Mariscal Estigarribia	0,14544	0,05943	0,06451
st-py-086	Puerto Casado	0,11859	0,15952	0,16431
st-py-097	Pedro Juan Caballero	0,10461	0,15779	0,13438
st-py-128	Pozo Colorado	0,11933	0,06510	0,24321
st-py-134	Concepción	0,11452	0,02264	0,10730
st-py-170	General Bruguez	0,11135	0,07879	0,18218
st-py-185	San Pedro	0,11578	0,21027	0,10549
st-py-192	San Estanislao	0,13369	0,18737	0,19276
st-py-210	Salto Guairá	0,12519	0,07180	0,14701
st-py-218	Asunción (Aeropuerto)	0,11988	0,12482	0,17307
st-py-221	Paraguarí	0,11861	0,10205	0,16563
st-py-233	Villarrica	0,11795	0,13028	0,16441
st-py-234	Coronel Oviedo	0,13760	0,20996	0,19832
st-py-246	Aeropuerto Guaraní	0,14018	0,11424	0,06512
st-py-255	Pilar	0,11826	0,17210	0,17900
st-py-260	San Juan Bautista Misiones	0,11774	0,09875	0,09952
st-py-268	Caazapá	0,13039	0,05604	0,04138
st-py-285	Capitán Meza	0,15179	0,02980	0,10322
st-py-297	Encarnación	0,13228	0,17377	0,02201

Tabla 15. Valores de los diferentes L-momentos por estación.

En las gráficas del 34 al 36, se observa los diferentes valores obtenidos para los L-momentos en cada estación analizada que figuran en la tabla 15.

Los valores normales de L-CV (*L*-coeficientes de variación) varían de 0 a 1 como.

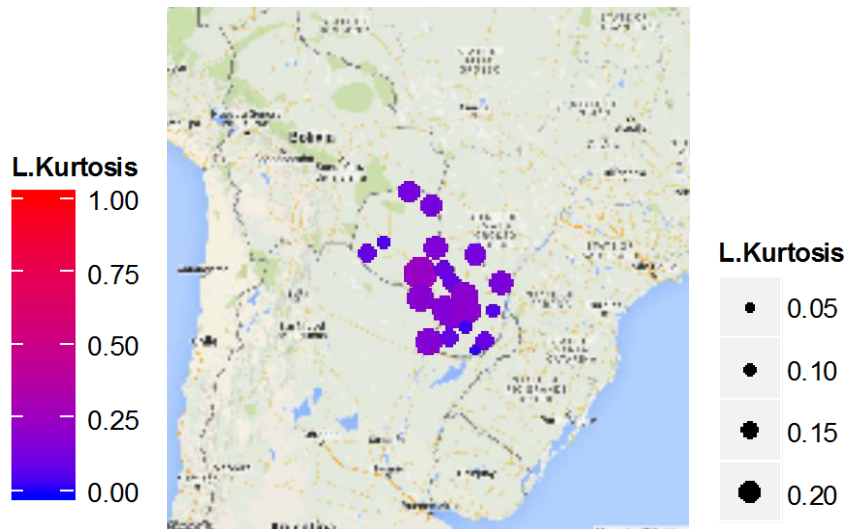
En la gráfica 34 se observa que la variación para las estaciones analizadas tiene un valor que va de 0.11 a 0.15, es decir poca dispersión con relación a los datos.



Gráfica 34. Los valores de los L-CV sobre el Paraguay.

Los valores normales de L-kurtosis varían de 0 a 1.

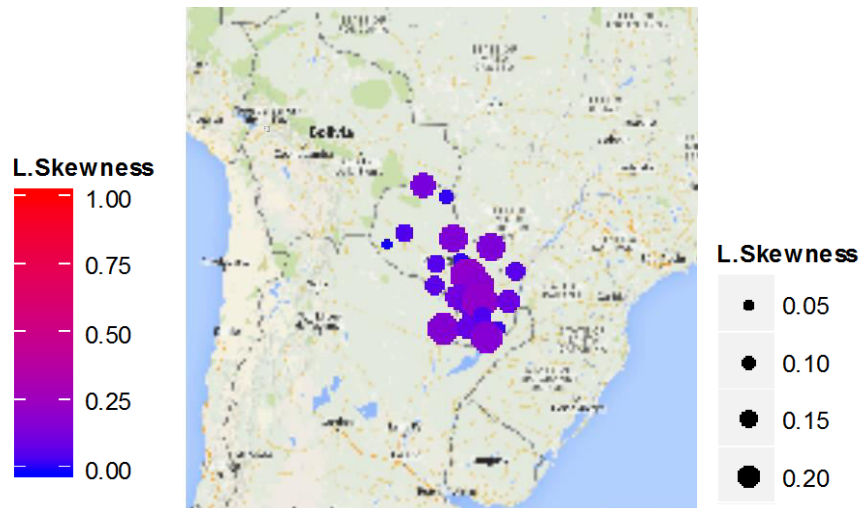
En la gráfica 35 se observa que la variación para las estaciones analizadas tiene un valor que va de 0.02 a 0.20, es decir la mayoría de los datos están concentrados en un lugar.



Gráfica 35. Los valores de los L-Kurtosis sobre el Paraguay.

Los valores normales de L-Skewness (*L*-Asimetría) varían de 0 a 1.

En la gráfica 36 se observa que la variación para las estaciones analizadas tiene un valor que va de 0.05 a 0.20, es decir la curva es asimétricamente positiva por lo que los valores se tienden a reunir más en la parte izquierda que en la derecha de la media.



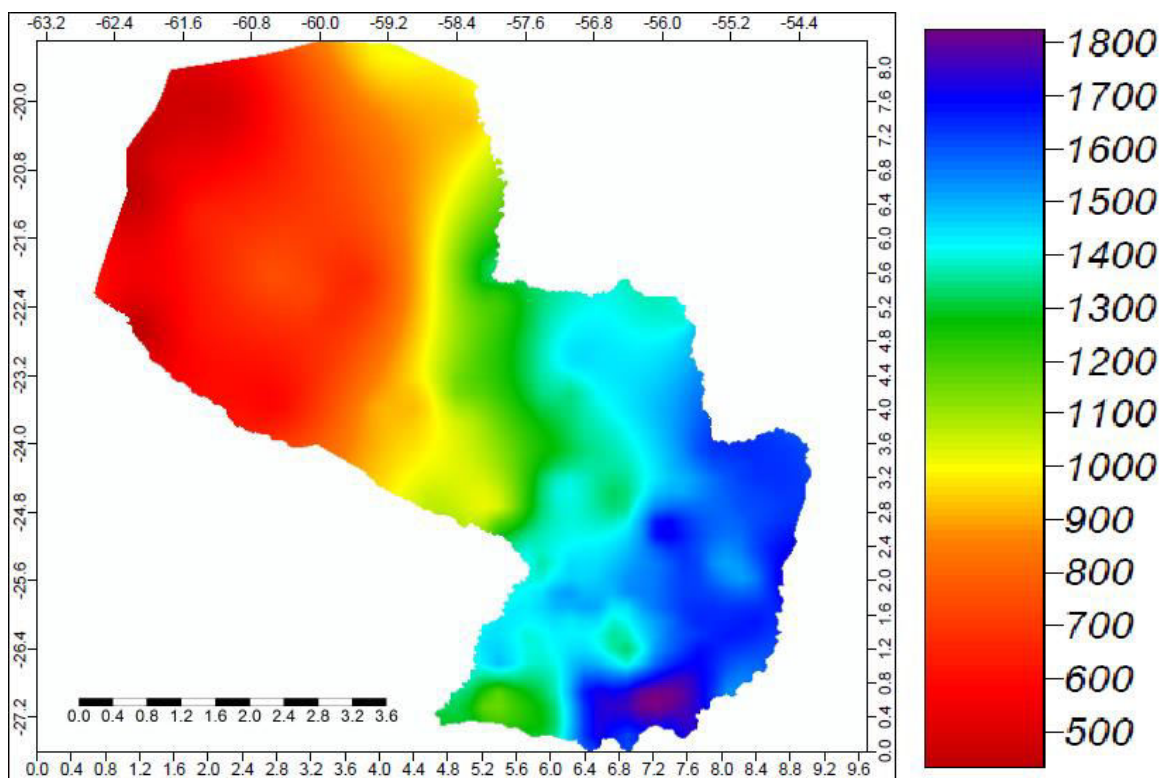
Gráfica 36. Los valores de los L-Skewness sobre el Paraguay.

A continuación se muestran las figuras entre los L-momentos y las variables auxiliares analizadas para ver su relación con cada una.

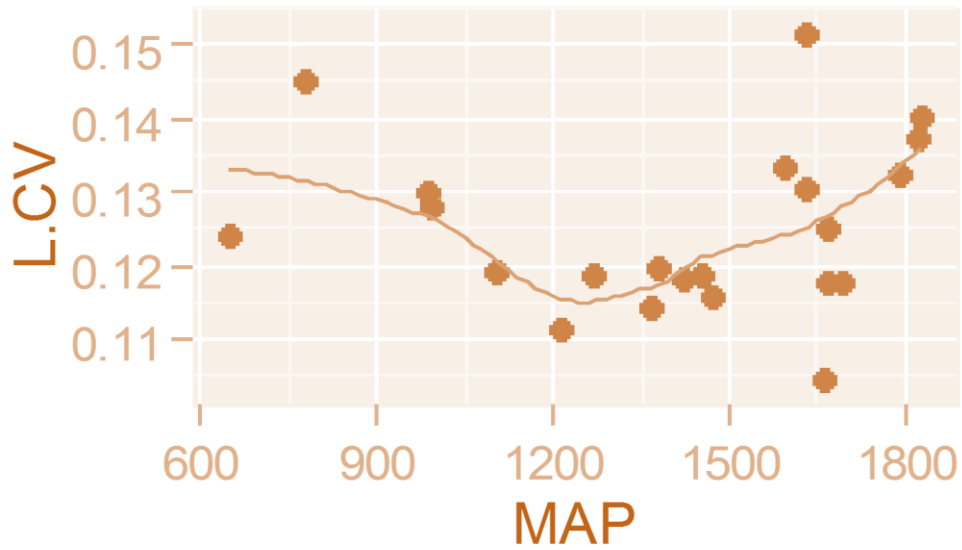
En las gráficas del 38 al 46, los valores de los diferentes L-momentos se observa en forma de puntos. Los valores de las variables auxiliares se verifican en forma de líneas.

Primeramente la PMA con los diferentes L-momentos.

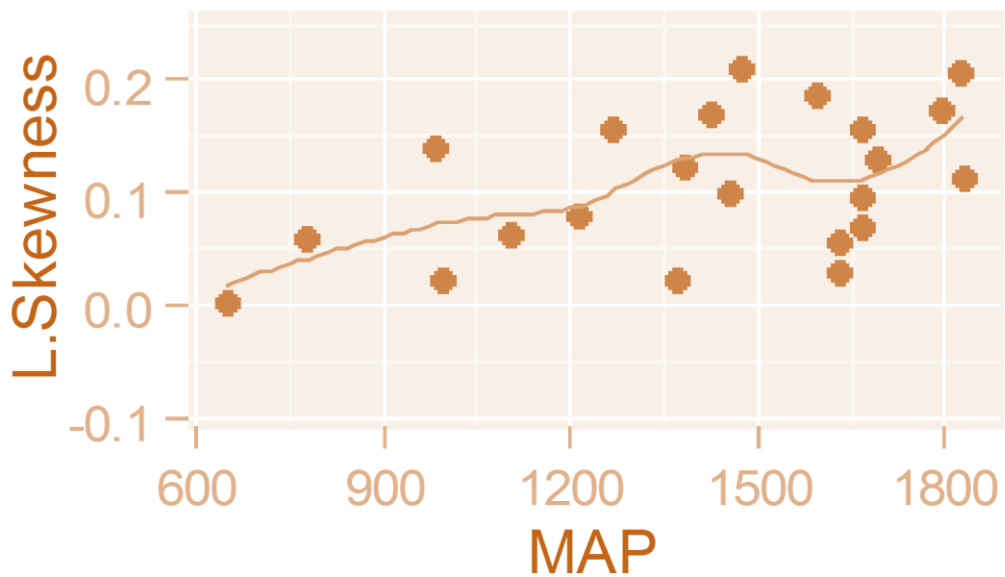
Se muestra a continuación la Precipitación Media Anual.



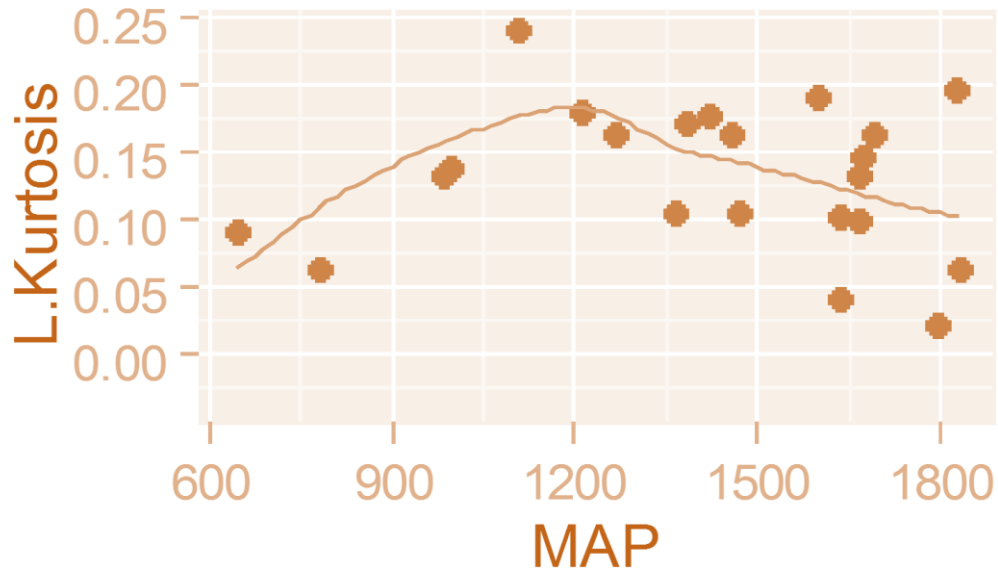
Gráfica 37. Los valores de la Precipitación media anual (PMA).



Gráfica 38. Curvas de mejor ajuste de la MAP vs L-CV.

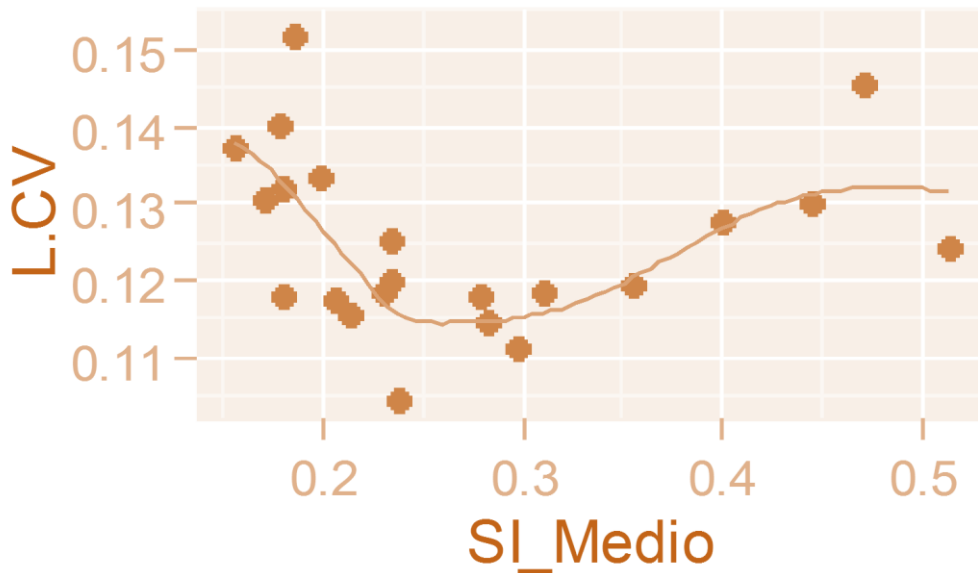


Gráfica 39. Curvas de mejor ajuste de la MAP vs L-Skewness.

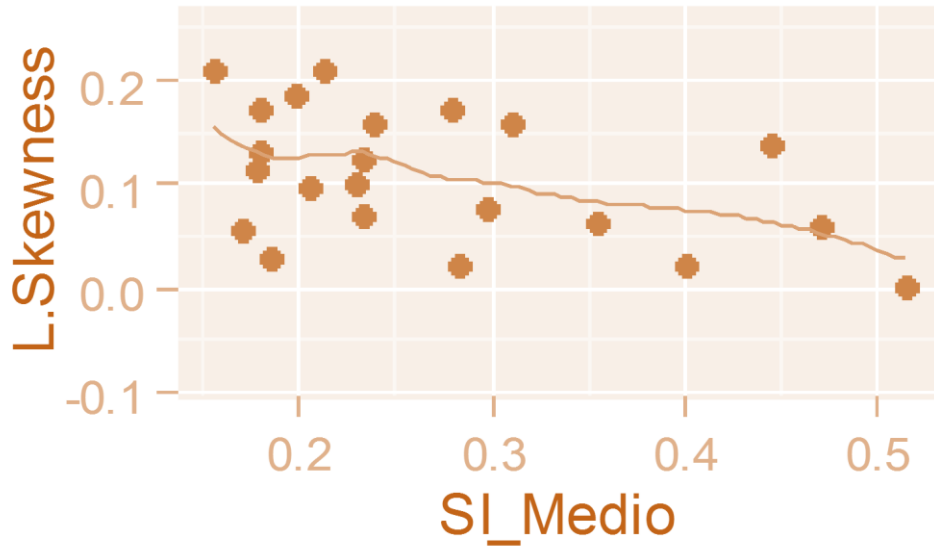


Gráfica 40. Curvas de mejor ajuste de la PMA vs L-Kurtosis.

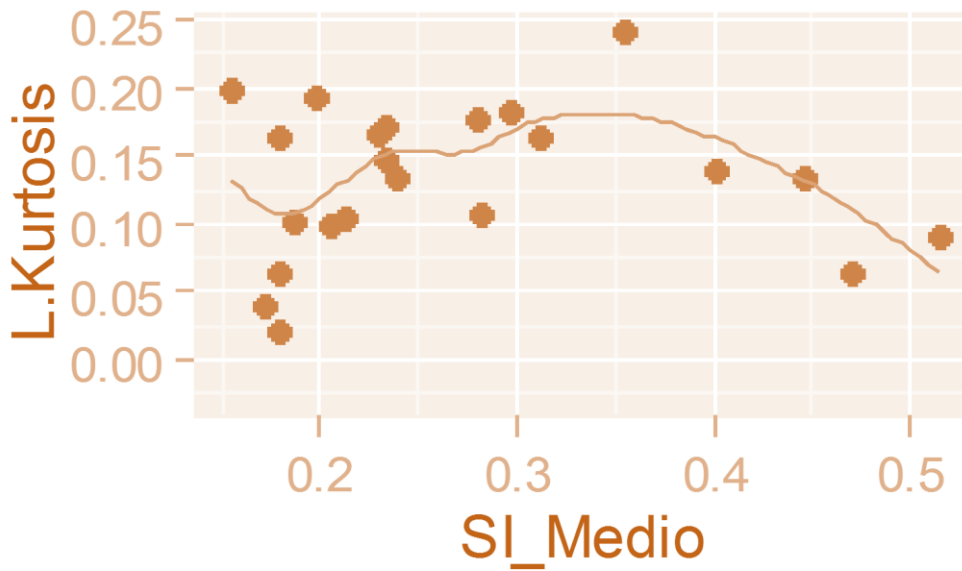
En segunda instancia el Índice de Estacionalidad Medio (SI_Medio) con los diferentes L-momentos.



Gráfica 41. Curvas de mejor ajuste de la SI_Medio vs L-CV.

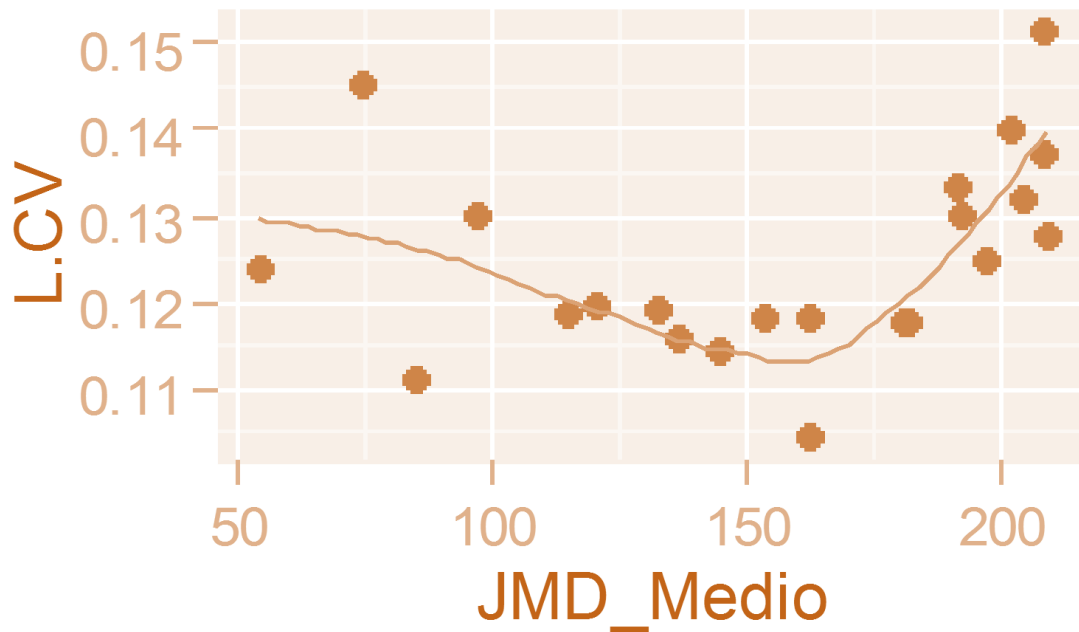


Gráfica 42. Curvas de mejor ajuste de la SI_Medio vs L-Skewness.

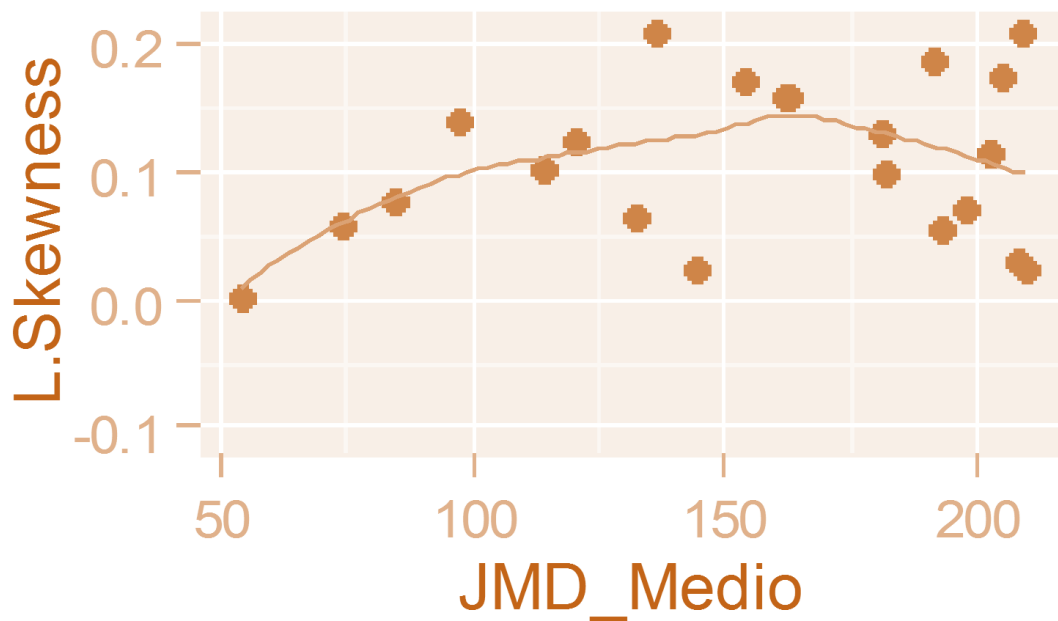


Gráfica 43. Curvas de mejor ajuste de la SI_Medio vs L-Kurtosis.

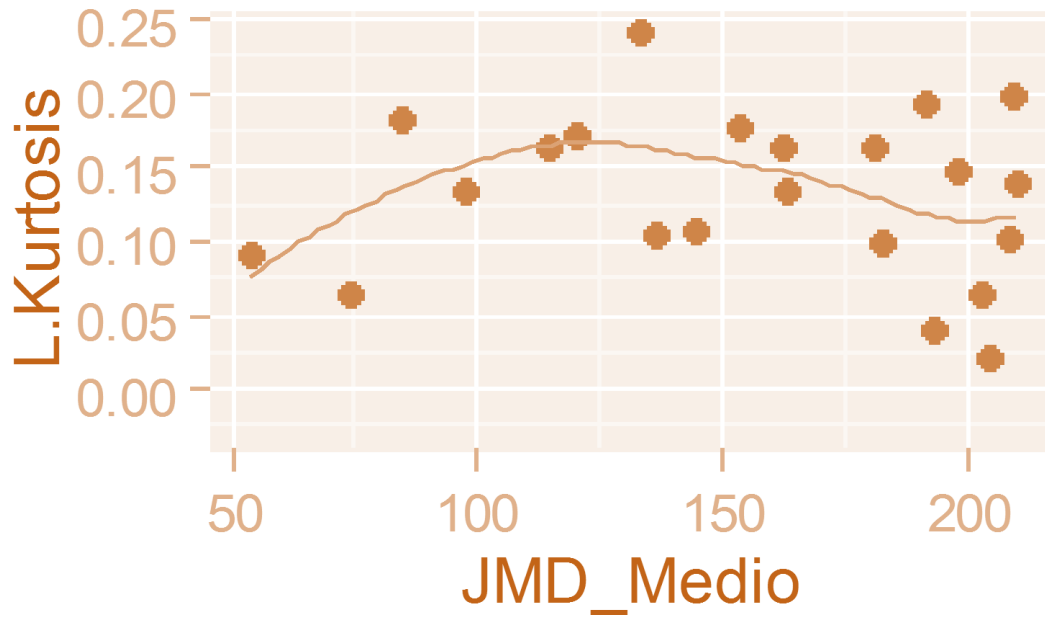
Y por último en tercera instancia Día Juliano Medio (JMD_Medio) con los diferentes L-momentos.



Gráfica 44. Curvas de mejor ajuste de la JMD_Medio vs L-CV.



Gráfica 45. Curvas de mejor ajuste de la JMD_Medio vs L-Skewness.



Gráfica 46. Curvas de mejor ajuste de la JMD_Medio vs L-Kurtosis.

En la tabla 16 se muestra los valores utilizados en la comparación para las diferentes gráficas presentadas anteriormente.

Identificación de cada estación	Nombre de cada estación	Precipitación Media Anual (mm) (PMA)	Índice de estacionalidad (SI)	Día Juliano Medio (DJM)	L-coeficientes de variación (L-CV)	L-Asimetría (L-Skewness)	L-Curtosis (L-Kurtosis)
st-py-011	Adrián Jara	983	0,44459	0,01032	0,13021	0,14002	0,13464
st-py-033	Bahía Negra	993	0,39981	0,00477	0,12791	0,02301	0,13945
st-py-065	Pratts Gill	646	0,51384	0,01862	0,12428	0,00351	0,09173
st-py-068	Mariscal Estigarribia	777	0,47062	0,01355	0,14544	0,05943	0,06451
st-py-086	Puerto Casado	1270	0,31038	0,00618	0,11859	0,15952	0,16431
st-py-097	Pedro Juan Caballero	1660	0,23802	0,00615	0,10461	0,15779	0,13438
st-py-128	Pozo Colorado	1100	0,35434	0,00754	0,11933	0,06510	0,24321
st-py-134	Concepción	1370	0,28204	0,00692	0,11452	0,02264	0,10730
st-py-170	General Bruguez	1210	0,29708	0,01183	0,11135	0,07879	0,18218
st-py-185	San Pedro	1470	0,21320	0,00735	0,11578	0,21027	0,10549
st-py-192	San Estanislao	1590	0,19750	0,00523	0,13369	0,18737	0,19276
st-py-210	Salto Guairá	1670	0,23296	0,00507	0,12519	0,07180	0,14701
st-py-218	Asunción (Aeropuerto)	1380	0,23274	0,00834	0,11988	0,12482	0,17307
st-py-221	Paraguarí	1450	0,22998	0,00876	0,11861	0,10205	0,16563
st-py-233	Villarrica	1690	0,17863	0,00553	0,11795	0,13028	0,16441
st-py-234	Coronel Oviedo	1820	0,15483	0,00480	0,13760	0,20996	0,19832
st-py-246	Aeropuerto Guaraní	1830	0,17802	0,00495	0,14018	0,11424	0,06512
st-py-255	Pilar	1420	0,27863	0,00652	0,11826	0,17210	0,17900
st-py-260	San Juan Bautista Misiones	1670	0,20551	0,00549	0,11774	0,09875	0,09952
st-py-268	Caazapá	1630	0,17069	0,00519	0,13039	0,05604	0,04138
st-py-285	Capitán Meza	1630	0,18590	0,00480	0,15179	0,02980	0,10322
st-py-297	Encarnación	1790	0,17910	0,00489	0,13228	0,17377	0,02201

Tabla 16. L-momentos y las variables auxiliares analizadas para cada estación.

8. Referencias

Vigilancia y alerta temprana de la sequía: conceptos, progresos y desafíos futuros. Información meteorológica y climática para el desarrollo agrícola sostenible, tipos de sequía. OMM-N° 1006 2006

Greenwood, J.A., Landwehr, J.M., Matalas, N.C., Wallis, J.R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research* 15 (5), 1049–1054.

F. Javier Sánchez San Román-Dpto. Geología-Univ. Salamanca (España)
<http://hidrologia.usal.es/temas/Precipitaciones.pdf>

Hosking, J.R.M., Wallis, J.R. (1987). Correlation and Dependence between Annual Maximum Flood Series. Research Report RC12822, IBM Research Division, Yorktown Heights, N.Y.

Dalrymple, T. (1960). Flood Frequency Analysis. Water Supply Paper 1543-A, USGS, Reston, VA, USA.

Saenz de Ormijana, F., Hidalgo, F.J., Santa Pérez, A. (1991). Estimación de las precipitaciones máximas mediante el método regional del índice de avenida. *Revista de Obras Públicas*, Feb., 9-22.

Guttman, N.B. (1993). The Use of L-Moments in the Determination of Regional Precipitation Climates. *Journal of Climate* 6, 2309-2325.

Ferrer, J., Ardiles, L., (1994). Análisis estadístico de las series anuales de máximas lluvias diarias en España. *Ingeniería Civil* 95, 87-100.

Erro, J., López, J., (2011). Análisis regional de las Precipitaciones diarias extremas En la cuenca del río Arga con R.

Análisis Espacial. Índice. Homogeneidad. <http://www.hypergeo.eu/spip.php?article135>

Enric Aguilar, Javier Sigró y Manola Brunet, C3. Manual_rclimdex_extraQC.r. Manual de Uso, versión 1.0.

Dr. Enric Aguilar, C3, Urv De Tarragona, España. Control De Calidad Con Andesqc. Tests Implementados.

Lizeth Llanos Herrera, estudiante de Estadística. RclimTool. MANUAL DEL USUARIO.

Prof. MSc. Grassi, Benjamín; Lic. Pasten, A, M;Téc. Sup. Met. Armoa, J. (2005). Un análisis del comportamiento de la precipitación en el Paraguay.

Universidad Autónoma De Baja California Sur, introducción a la Estadística Descriptiva, Método de Análisis Gráfico, Histograma De Frecuencia.
<http://www.uabcs.mx/maestros/descartados/mto07/histfrec.htm>

Dra. Diana M. Kelmansky. ANÁLISIS DE DATOS 2008.

9. Glosario

Densidad de probabilidad en la teoría de la probabilidad, la función de densidad de probabilidad, función de densidad, o, simplemente, densidad de una variable aleatoria continua describe la probabilidad relativa según la cual dicha variable aleatoria tomará determinado valor.

El periodo de retorno se define como el intervalo de recurrencia (T), al lapso promedio en años entre la ocurrencia de un evento igual o mayor a una magnitud dada. Este periodo se considera como el inverso de la probabilidad, del m-ésimo evento de los n registros.

$$T = \frac{n + 1}{m}$$

$$P = \frac{m}{n + 1}$$

Donde:

T= Periodo de retorno (años)

n= Número de años de registro

m= Numero de orden

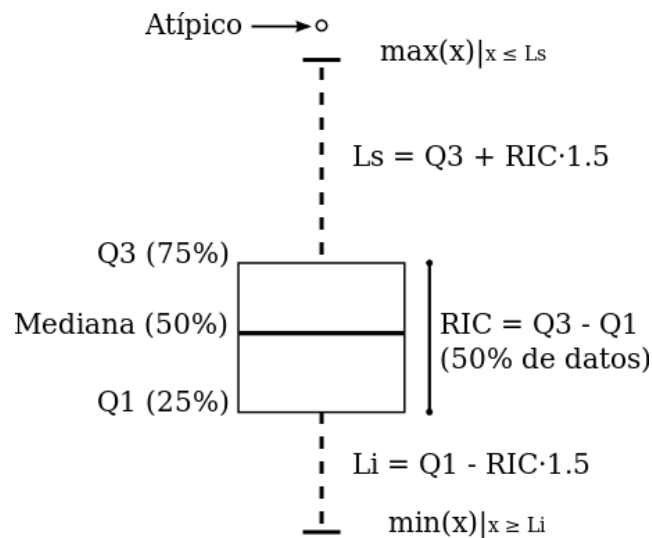
P= Probabilidad

Los outliers (valores atípicos) en estadística un valor atípico es una observación que es numéricamente distante del resto de los datos.

El rango intercuartílico es una medida de variabilidad adecuada cuando la medida de posición central empleada ha sido la mediana. Se define como la diferencia entre el tercer cuartil (Q_3) y el primer cuartil (Q_1), es decir: $RQ = Q_3 - Q_1$, Se usa para construir los diagramas de caja y box plots que sirven para visualizar la variabilidad de una variable y comparar distribuciones de la misma variable; además de ubicar valores extremos.

Un **Diagrama de caja (box-plot)** es un gráfico, basado en cuartiles, mediante el cual se visualiza un conjunto de datos, Está compuesto por un rectángulo, la “caja”, y dos brazos, los “bigotes”.

Es un gráfico que suministra información sobre los valores mínimo y máximo, los cuartiles Q1, Q2 o mediana y Q3, y sobre la existencia de valores atípicos y la simetría de la distribución. Primero es necesario encontrar la mediana para luego encontrar los 2 cuartiles restantes.

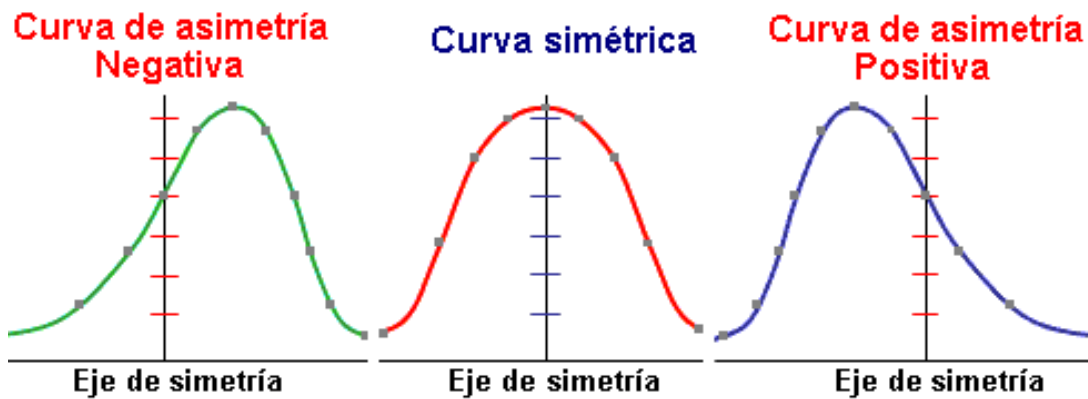


Cuartiles son tres valores con las siguientes características:

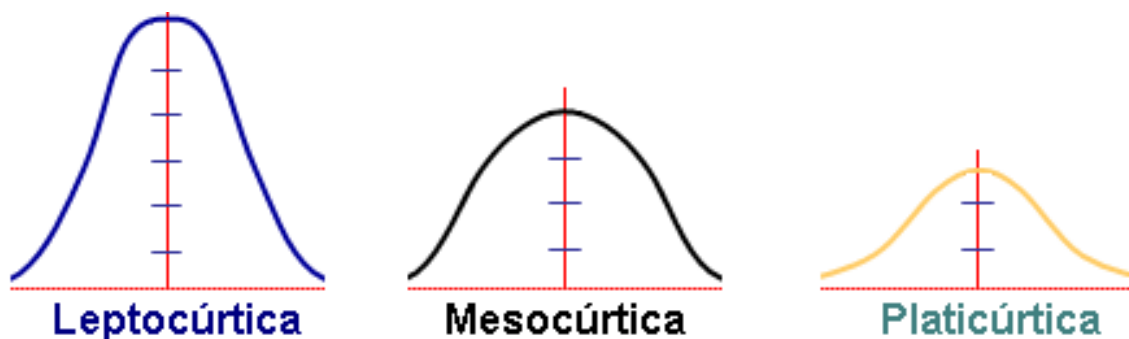
- ◆ Q_1 : Primer cuartil, que es el valor de la variable por debajo del cual queda 1/4 de los elementos de la serie estudiada.
- ◆ Q_3 : Tercer cuartil, que es el valor de la variable por debajo del cual quedan los 3/4 de los elementos que constituyen la serie.
- ◆ Evidentemente el segundo cuartil coincide con la mediana.

El **coeficiente de variación** es una medida de dispersión que describe la cantidad de variabilidad en relación con la media. Puesto que el coeficiente de variación no se basa en unidades, se puede utilizar en lugar de la desviación estándar para comparar la dispersión de los conjuntos de datos que tienen diferentes unidades o diferentes medias.

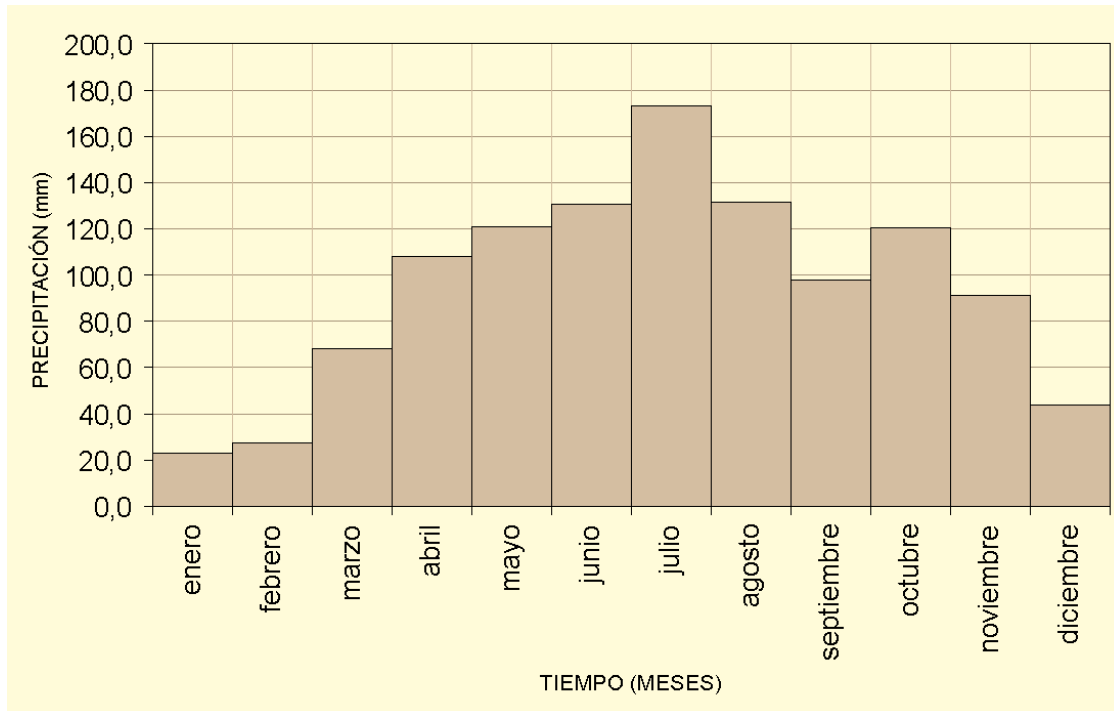
Asimetría nos permite identificar si los datos se distribuyen de forma uniforme alrededor del punto central (Media aritmética). La asimetría presenta tres estados diferentes, cada uno de los cuales define de forma concisa como están distribuidos los datos respecto al eje de asimetría. Se dice que la *asimetría es positiva* cuando la mayoría de los datos se encuentran por encima del valor de la media aritmética, la curva es *Simétrica* cuando se distribuyen aproximadamente la misma cantidad de valores en ambos lados de la media y se conoce como *asimetría negativa* cuando la mayor cantidad de datos se aglomeran en los valores menores que la media.



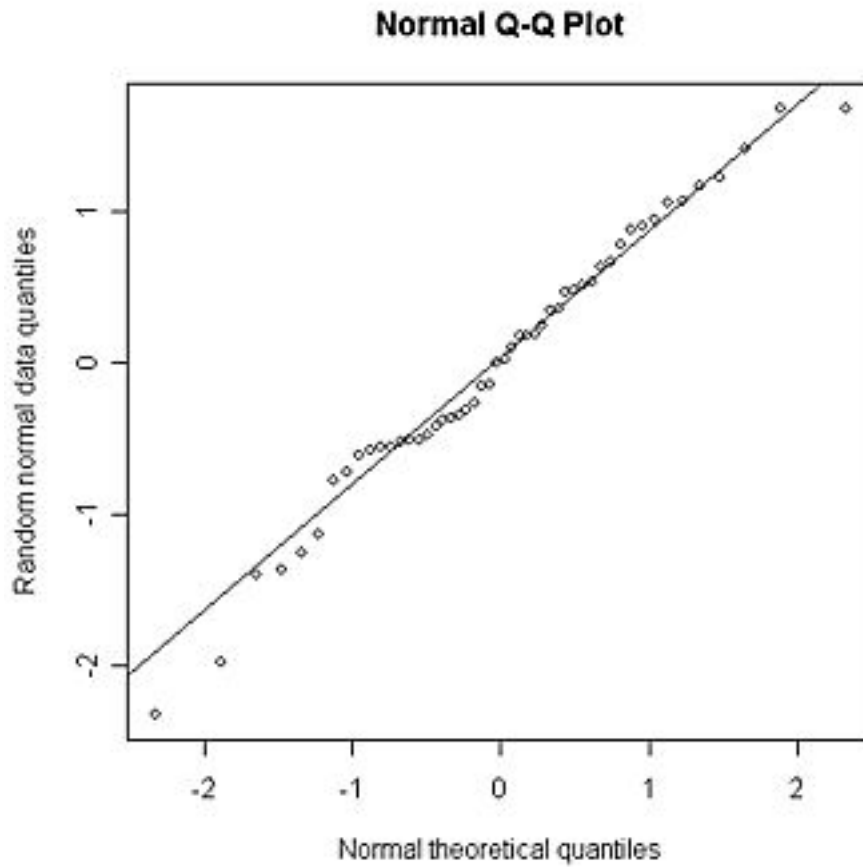
Curtosis esta medida determina el grado de concentración que presentan los valores en la región central de la distribución. Por medio del *Coefficiente de Curtosis*, podemos identificar si existe una gran concentración de valores (*Leptocúrtica*), una concentración normal (*Mesocúrtica*) o una baja concentración (*Platicúrtica*).



Los histogramas son diagramas de barras verticales en los que se construyen barras rectangulares en los límites de cada clase. La variable aleatoria o fenómeno de interés se despliega a lo largo del eje horizontal; el eje vertical representa el número, proporción o porcentaje de observaciones por intervalo de clase, dependiendo de si el histograma particular, es un histograma de frecuencia, un histograma de frecuencia relativa o histograma de porcentaje.



Gráficos Cuantil-Cuantil (Q-Q plots) Un gráfico Cuantil-Cuantil permite observar cuán cerca está la distribución de un conjunto de datos a alguna distribución ideal o comparar la distribución de dos conjuntos de datos. Comparación de la distribución de dos conjuntos de datos La función `qqplot(x, y, plot=T)` grafica las funciones cuantiles de una muestra vs la de la otra. Vemos que el Q-Q plot no cambia por una transformación lineal de los datos.





“Este Proyecto es financiado por el CONACYT a través del Programa PROCIENCIA con recursos del Fondo para la Excelencia de la Educación e Investigación – FEEL del FONACIDE”.