

CAPÍTULO I

INTRODUCCIÓN

Los resultados obtenidos en estudios de hidrología dependen en gran medida de los datos con los que se realice, es por esto, que seleccionar de buena manera las estaciones que se aplicarán para algún tipo de estudio es de suma importancia ya que si estas cuentan con registros buenos los resultados obtenidos serán mejores.

Previo a realizar cualquier estudio, se debe realizar un análisis de la información brindada por las estaciones para verificar la calidad de sus registros, es decir, identificar, evaluar y eliminar los posibles errores sistemáticos que hubieron podido ocurrir sea por causas naturales u ocasionadas por la intervención del hombre.

Los criterios con los que se seleccionaron estaciones en estudios anteriores son:

1. Balance Hídrico Superficial de Bolivia (1982). - La corrección y homogeneización de los datos anuales fue efectuada por el método de doble acumulada, tomando como vector de referencia la media de los valores del grupo de estaciones de cada región individualizada. Se han utilizado algunas estaciones como simple referencia a pesar de que sus datos no puedan homogeneizarse en razón de la corta duración del período de observación.
2. Balance Hídrico Superficial de Bolivia (2012). – Para detectar inconsistencias en los datos de precipitación de una estación, en este estudio se nos presenta el análisis de doble masa. Para la depuración de registros diarios de temperaturas nos presenta como herramienta el MatLab.
3. Guía Metodológica para la elaboración de balances hídricos superficiales (2016). – Este trabajo nos propone una visualización cualitativa de los datos que se aplica a través de histogramas, distribuciones acumuladas de frecuencias, diagramas de caja y diagramas de series cronológicas entre otros.

1.1.PLANTEAMIENTO DEL PROBLEMA

La disponibilidad de información climática de calidad y espacialmente distribuida es importante para el desarrollo de una investigación en distintas disciplinas como la Hidrología, Agronomía, Climatología y Ecología.

Por eso, una evaluación precisa del balance hídrico está estrechamente ligada con la cantidad y calidad de la información disponible y en particular, de la red de estaciones y número de años de sus registros.

El problema que se plantea es que la información climatológica con la cual se dispone para aplicar cualquier método indirecto que permita estimar las variables a considerar en la generación de un modelo climatológico, generalmente no es satisfactoria o existe falta de calidad en ella, Por lo tanto, se hace necesario el conocer cada uno de los factores relacionados con la cobertura espacial y temporal de los datos, así como falta de homogeneidad en los métodos para registrar la información.

Sin embargo, no existe un criterio definido para seleccionar las estaciones a emplear en un plan de aprovechamiento en las diferentes zonas; es por esto, que en el presente trabajo se pretende realizar un *Análisis de los Criterios de Selección de Estaciones* que se utilizarán en la generación de la grilla meteorológica GMET, para analizar si es que alguno de ellos se ajustará a la Cuenca del Río Guadalquivir en función a la cantidad de estaciones que se tiene en la zona, la calidad de sus registros y el periodo de años registrado de las mismas.

1.2.JUSTIFICACIÓN DEL TEMA

Se puede observar que en la zona que respecta a la Cuenca del Río Guadalquivir hay deficiencia en los proyectos relacionados al contexto del agua. Por un lado, se hace énfasis en que uno de los principales problemas es el fuerte racionamiento de agua durante la época de estiaje en el municipio de Tarija. Por otro, la carencia de este servicio básico en algunas zonas de otros municipios que comprenden el área en estudio como ser San Lorenzo, Uriondo y Padcaya hace necesario para estos últimos proyectos de riego, ya que es la principal fuente para sus actividades agropecuarias.

Esta es la razón por la que se considera importante realizar un análisis de los datos que brinden las estaciones, debido a que la información que proporcionan actualmente es deficiente, escasa y en el peor de los casos inexistente. Es por eso, que se pretende realizar el análisis de los criterios de selección de estaciones para analizar si alguno de estos se ajusta a la zona estudiada y así obtener información termopluviométrica corregida y procesada con la aplicación de la grilla meteorológica GMET que fue además la herramienta utilizada en el Balance Hídrico Superficial de Bolivia 2017, del cual se pretende realizar un ajuste a nivel regional en la Cuenca del Río Guadalquivir cuya importancia radica en que las fuentes de agua más importantes tanto para consumo humano como para la agricultura se encuentran ubicadas en esta cuenca.

1.3.OBJETIVOS

1.3.1. OBJETIVO GENERAL

Realizar el *Análisis de los Criterios de Selección de Estaciones* con datos de precipitación y temperaturas a nivel diario presentes en la Cuenca del Río Guadalquivir, con herramientas de control de calidad para su aplicabilidad en la grilla meteorológica GMET.

1.3.2. OBJETIVOS ESPECÍFICOS

- Identificar las estaciones presentes dentro de la Cuenca del Río Guadalquivir y establecer el número de años de registro de las mismas.
- Identificar las estaciones cercanas al área de estudio que nos brindarán información relevante para el trabajo que se pretende realizar.
- Analizar los datos de precipitaciones y temperaturas de las estaciones meteorológicas aplicando herramientas como RClindex y DECADE.
- Analizar los diferentes criterios de selección de estaciones para la generación de la Grilla Meteorológica (GMET), aumentando la cantidad de estaciones de entrada y definiendo un periodo diferente a 1980-2016.

1.4.HIPÓTESIS A COMPROBAR

Analizar los datos de las estaciones con herramientas de control de calidad como RClindex y DECADE, y aplicar test de homogeneidad con el método de las Alternancias y de Helmer

para seleccionar la información que brindan las estaciones presentes en la Cuenca del Río Guadalquivir proporcionará información termopluviométrica corregida y procesada para su uso en proyectos de desarrollo en el aprovechamiento hídrico en la zona de estudio.

1.5.LIMITACIONES DEL ESTUDIO

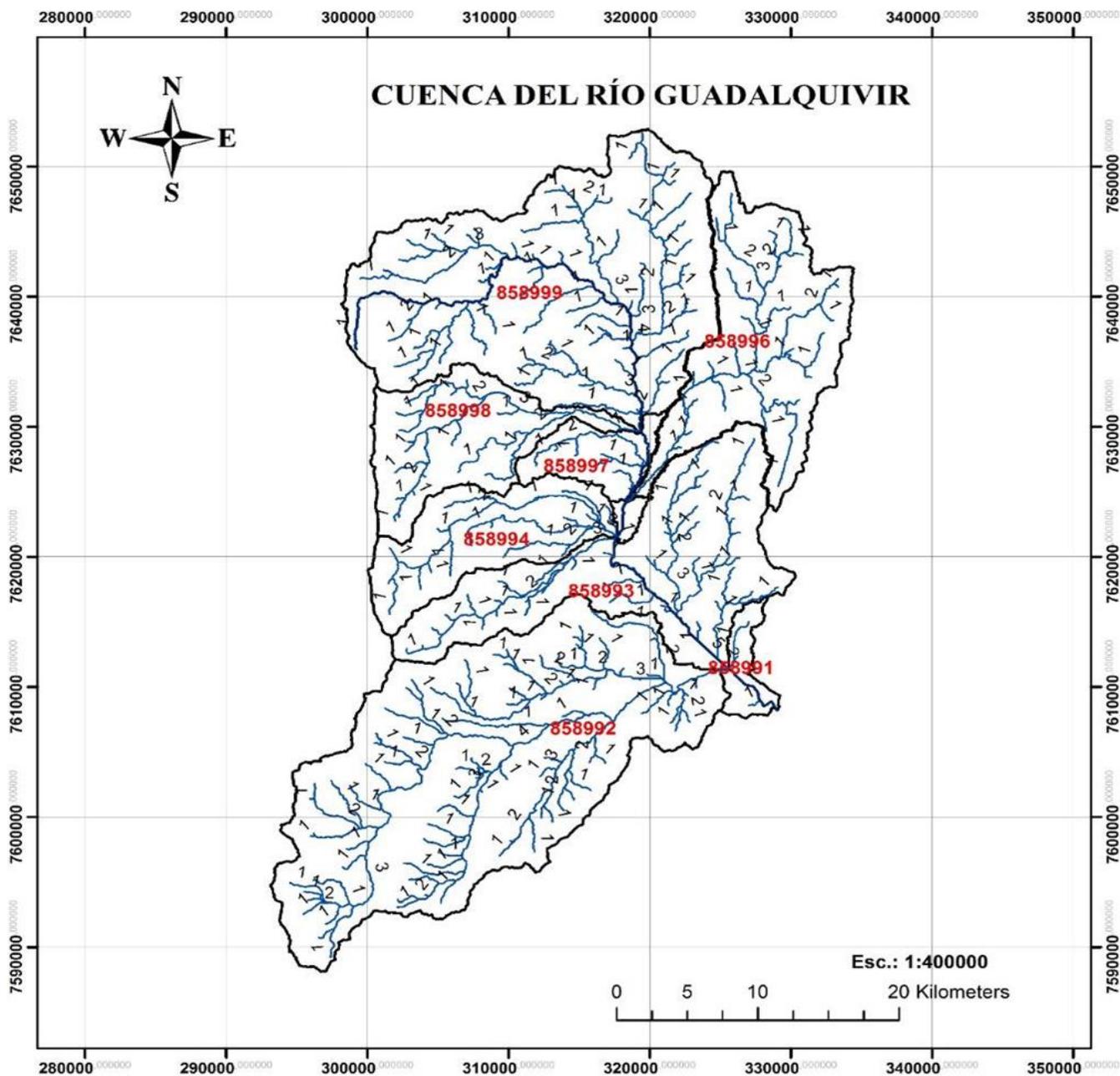
Para la realización del análisis de los criterios de selección de estaciones, se hará un estudio de los datos que nos brindan, aplicando herramientas para el control de calidad como el RClindex y DECADE herramientas aplicadas en el Balance Hídrico Superficial de Bolivia 2017. Este análisis se realizará con datos de precipitaciones y temperaturas a nivel diario. Por esto se presentan las siguientes limitaciones:

Tabla N° 1 Limitaciones del estudio

Temática	Análisis de los criterios de selección de estaciones para la generación de la grilla meteorológica GMET en la Cuenca del Río Guadalquivir.
Espacial	<p>La cuenca del Río Guadalquivir conocido como Valle Central de Tarija, comprende los municipios de San Lorenzo, Cercado, Uriondo y Padcaya.</p> <p>Geográficamente se encuentra delimitada por la serranía de Sama al oeste, la serranía de Gamoneda al este y la salida de la cuenca en el angosto de San Luis, abarcando aproximadamente 1536 km².</p>

Fuente: Elaboración Propia

Figura N°1 Cuenca del Río Guadalquivir



Fuente: Balance Hídrico Integral para la Cuenca del Río Guadalquivir

CAPÍTULO II

MARCO TEÓRICO

2.1. MARCO HISTÓRICO

2.1.1. CUENCA DEL RÍO GUADALQUIVIR

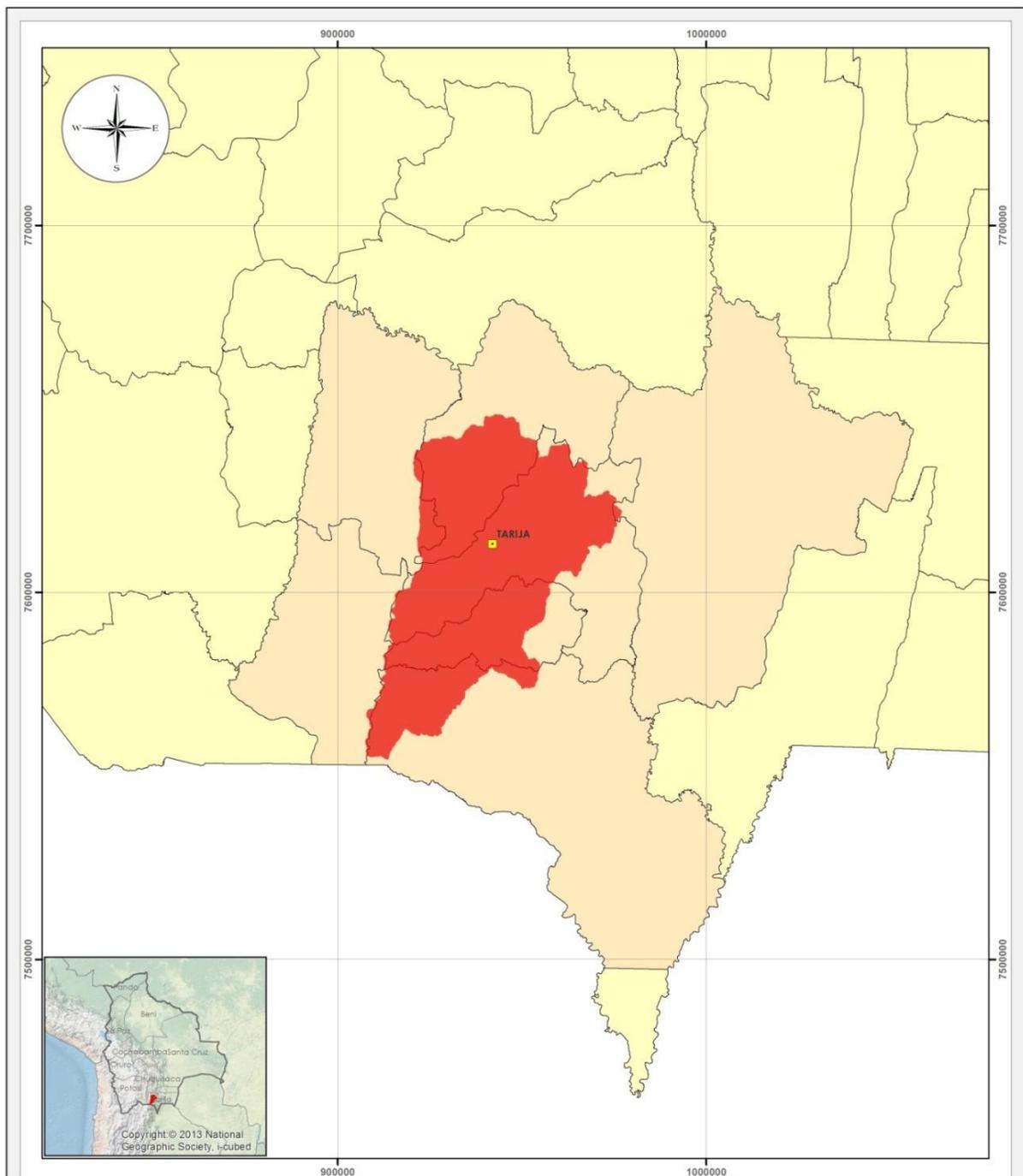
2.1.1.1 Ámbito geográfico

La Cuenca del Río Guadalquivir conocida como Valle Central de Tarija, situada al sur de Bolivia en el departamento del mismo nombre, comprende los municipios de San Lorenzo, Cercado, Uriondo y Padcaya.

La altura varía desde los 4674 msnm que corresponde al pico del cerro Negro del Chiquiro en la serranía de Sama, hasta los 1790 msnm en el Angosto de San Luis.

La cuenca del Río Guadalquivir es tributaria del río Bermejo, afluente del río Paraguay y perteneciente a la gran cuenca hidrográfica del río La Plata. El río Guadalquivir nace del aporte de ríos como el Chamata y Trancas, entre otros cuerpos de agua que emergen en las faldas de la serranía de Sama a unos 50 km al noroeste de la ciudad de Tarija. La cuenca abarca una superficie de 1.536 km², está delimitada por la línea divisoria de las serranías que la rodean y concluye en El Angosto en el municipio de Uriondo, donde cambia de nombre a río Grande de Tarija. El río Guadalquivir en la parte alta, dentro del municipio de San Lorenzo, recibe el aporte de los ríos Calama y Erquis por la margen izquierda y de los ríos Carachimayu y Sella por la margen derecha. Aguas abajo en el límite entre San Lorenzo y Tarija, recibe las aguas del río Victoria. En la jurisdicción de Tarija las quebradas Sosa y Sagredo y el río Tolomosa, que proviene del embalse de San Jacinto, son los principales aportantes del río Guadalquivir por la margen izquierda, y las quebradas El Monte, San Pedro, Torrecillas y Cabeza de Toro por la ribera derecha.

Figura N°2 Ubicación de la cuenca con respecto al Departamento de Tarija



Fuente: Balance Hídrico Integral para la Cuenca del Río Guadalquivir

2.1.1.2. Topografía

Topográficamente es irregular con variadas altitudes, y se encuentra con frecuencia:

- Terrenos escarpados: Entre 50% y 75% de pendiente.
- Fuertemente ondulados y quebrados: Entre 12% y 25% de pendiente.
- Ligeramente ondulados: De 3% a 9% de pendiente.
- Terrenos casi planos: De 2% a 3% de pendiente.

Asimismo, las variadas altitudes dan lugar a las subformaciones:

i) Zona Alta: Presenta subformación montañosa, en su parte occidental, con elevaciones montañosas como las serranías de Alto de Cajas, Jarcas, Alto España, León Cancha, San Lorencito, Trancas y Tomatas. Esta zona presenta mesetas onduladas donde predominan las Serranías o Colinas, con cimas redondeadas, de disección moderada a fuerte y con interfluvios estrechos.

ii) Zona Baja: De Valles comprendiendo al sector que se extiende entre los cantones de Tomatitas, La Calama, Sella Méndez, San Lorenzo y Canasmoro. Esta Zona presenta un Valle de Origen Fluvio Lacustre, con predominio de Terrazas Aluviales, además de pie de monte, abanicos de origen pluvial y planicies.

2.1.2. ESTUDIOS REALIZADOS EN LA CUENCA

La Cuenca del Río Guadalquivir es la cuenca más estudiada del departamento de Tarija por la importancia para el desarrollo de la misma, se cuenta con varios estudios hidrológicos y climáticos en sus subcuencas y también en sus afluentes principales.

Se realiza un breve resumen de dos de estos estudios asumiendo que son los más importantes:

2.1.2.1 Valoración hidrológica de las cuencas de los ríos Tolomosa y la Victoria

En el año 2002 se propuso un estudio realizado en las cuencas de los ríos Tolomosa y la Victoria, el cual, tenía cuatro componentes principales: análisis de escenarios, estudio de

oferta y demanda de agua, estudio de aporte de sedimentos y estudio de crecidas con una metodología propuesta en forma independiente para cada uno de los componentes.

Los objetivos de este estudio fueron evaluar a nivel mensual el aporte líquido (oferta de agua) y el balance hídrico de dos de las subcuencas más importantes de la cuenca del río Guadalquivir como son las subcuencas de los ríos Tolomosa y La Victoria. Tanto para las condiciones actuales como para escenarios hipotéticos. Otro de los objetivos fue evaluar los requerimientos de riego por cultivo en ambas cuencas y estimar el aporte total medio de sedimentos de esas cuencas debido a la erosión del suelo por lavado.

2.1.2.2. Balance hídrico integral para la Cuenca del Río Guadalquivir

En el año 2016 se realizó el estudio más reciente y completo que se llevó a cabo en la cuenca del Río Guadalquivir, éste fue elaborado por “Helvetas” que realizó el balance hídrico integral de la Cuenca del Río Guadalquivir caracterizando la misma en base a la información que fue recopilada del SENAMHI, se utilizó 27 estaciones pluviométricas (16 dentro la cuenca en estudio y 11 fuera de la Cuenca del Río Guadalquivir), 14 estaciones de temperatura, 7 de humedad relativa, 7 de evaporación y 7 con información de viento.

El periodo de estudio se definió en función a un diagnóstico de la información hidrometeorológica (climática) a nivel mensual disponible en la Base de datos (SENAMHI).

Para definir el periodo del balance hídrico de la Cuenca del Río Guadalquivir se consideraron dos criterios: el periodo común de datos y el número de estaciones en funcionamiento para un periodo. Es importante aclarar que se cuentan con muy pocas estaciones con más de 25 años de registro continuos. El segundo criterio fue el número de estaciones, que cuentan con datos entre los periodos predefinidos desde 1980 a 2014.

2.1.2.3. Plan de manejo integral de la Alta Cuenca del Río Guadalquivir

El objetivo principal del plan fue promover el desarrollo sostenible de la Alta Cuenca del Río Guadalquivir, a través del uso y manejo ordenado de los recursos naturales y la protección y conservación del medio ambiente de la Cuenca Alta del Río Guadalquivir. Los objetivos específicos fueron:

1. Implementar medidas y prácticas de control de la erosión y conservación de suelos.
2. Fomentar prácticas de aprovechamiento y manejo eficiente del agua.
3. Fomentar prácticas de cultivos agrícolas en áreas a secano y riego.
4. Establecer una red de monitoreo de la calidad de agua de los ríos e implementar sistemas de mitigación de aguas contaminadas.
5. Implementar obras de control de inundaciones.
6. Fomentar la protección de áreas naturales y la biodiversidad.

Los principales problemas de degradación ambiental identificados en el diagnóstico de la Cuenca Alta del Río Guadalquivir fueron:

1. Erosión de los suelos, fenómeno que afecta a toda la cuenca con diferentes grados de intensidad y reduce paulatinamente la capacidad de producción de la tierra. Los factores naturales que coadyuvan a este problema tienen que ver con la inestabilidad de las formaciones geológicas, el relieve montañoso con fuertes pendientes, las altas intensidades y concentración de lluvias en tres meses del año, y los factores antrópicos, principalmente la pérdida de la cobertura vegetal del suelo debido a la deforestación, sobrepastoreo, la destrucción de la cobertura vegetal para habilitar terrenos de cultivo y el inadecuado uso del suelo.
2. Limitado y deficiente aprovechamiento del recurso hídrico; la fuerte estacionalidad de la precipitación pluvial resulta en la baja disponibilidad de agua en la época seca del año, problema que se agrava debido a la insuficiente y deficiente infraestructura de regulación del agua.
3. Baja producción y productividad agropecuaria; el principal factor antrópico para este problema es el uso de prácticas agropecuarias y forestales inapropiadas (monocultivo, cultivo en laderas, sobrepastoreo y deforestación) y entre los factores naturales se tiene la baja capacidad de producción natural del suelo, la escasa superficie de suelos aptos para la actividad agropecuaria, el corto período de lluvias y las frecuentes sequías, heladas, granizadas y crecidas de los ríos.
4. Contaminación del agua; se presenta en el tramo del río Guadalquivir a su paso por la Ciudad de Tarija la que es causada, por una parte, por el vertido directo al río de

aguas servidas sin tratamiento, y, por otra, los desechos sólidos. El resultado de esta contaminación se manifiesta en el incremento de enfermedades infectocontagiosas, disminución de la calidad de agua para consumo humano y riego, disminución de la flora y fauna acuática y deterioro de áreas recreativas.

5. Pérdida de la biodiversidad, se manifiesta en la paulatina disminución de la población de especies de aves, mamíferos y peces y por la pérdida de ecosistemas naturales causado por destrucción de hábitats y la caza y pesca indiscriminada.
6. Inundaciones, heladas granizadas y sequías, son fenómenos frecuentes que afectan áreas rurales y urbanas causando pérdidas de cosechas en el área rural. Asimismo, las inundaciones afectan la infraestructura urbana, principalmente, la de la ciudad de Tarija

Descripción del proyecto. - El proyecto propone un programa de 8 años para implementar acciones integrales en un 70% del territorio de la Alta Cuenca, incluyendo los siguientes componentes:

1. Control de la erosión, manejo y conservación de los recursos naturales
2. Aprovechamiento y manejo del recurso hídrico
3. Mejoramiento de prácticas de producción agropecuaria
4. Control de inundaciones y saneamiento ambiental del río Guadalquivir
5. Protección de áreas naturales y la biodiversidad

Con un manejo de manera integral y sostenible, se logrará un equilibrio entre la protección del medio ambiente y el uso racional de los recursos naturales, permitiendo elevar el nivel de vida de los habitantes y desarrollo sostenible de la cuenca.

Con el proyecto se beneficiará a más de 2.800 familias en el área rural por el mejoramiento de las condiciones medio ambientales, la construcción de infraestructura de protección de inundaciones y obras de regulación de caudales.

2.1.2.4. Diagnóstico Ambiental de la Alta Cuenca del Río Guadalquivir

Este estudio concluyó señalando que los resultados de la gestión ambiental de las entidades evaluadas el Ministerio de Medio Ambiente y Agua, el Gobierno Autónomo Departamental de Tarija, los Gobiernos Autónomos Municipales de San Lorenzo, Tarija, Uriondo y Padcaya, la Entidad Ejecutora de Medio Ambiente y Agua, la Oficina Técnica Nacional de los ríos Pilcomayo y Bermejo y la Cooperativa de Servicio de Agua y Alcantarillado Tarija, no han logrado mitigar los impactos ambientales negativos generados en la cuenca del río Guadalquivir.

Esta situación se ha visto reflejada en un ecosistema dañado, conformado por cuerpos de agua contaminados, la manifestación de consecuencias reales y riesgos potenciales importantes para la biodiversidad de los cuerpos de agua de la cuenca y para la salud de la población implícitamente afectada por las prácticas que realizan en la zona como la agricultura y la recreación.

Para contribuir a revertir esta situación, la Contraloría General del Estado ha formulado cuarenta y siete recomendaciones, orientadas a mejorar la gestión ambiental de las entidades involucradas y vinculadas de manera directa con la cuenca del río Guadalquivir, a fin de lograr su recuperación y restauración y atenuar los riesgos potenciales presentes para la salud de la población expuesta.

2.2. MARCO CONCEPTUAL

2.2.1. COMPLETADO DE DATOS

Los vacíos de datos en un registro meteorológico se pueden rellenar o completar mediante técnicas que combinen el conocimiento de la evolución de cada variable meteorológica, de sus relaciones con las registradas en el entorno y de su relación con otras variables registradas bajo las mismas condiciones. Para ello, se puede hacer uso del conocimiento físico de la variable y de sus relaciones; y de técnicas de interpolación y estadísticas. La revisión de los métodos para el completado de datos de una serie y una variable se han ordenado de la siguiente manera:

1. Aprovechando únicamente la información contenida en el resto de la propia serie temporal. *Métodos univariados*.
2. En función de series de la misma variable registradas en “otros puntos bajo condiciones climáticas similares”. *Métodos multivariantes aplicados a una única variable meteorológica*.

2.2.1.1. Métodos univariados

Los métodos de completado univariado reciben este nombre porque para completar las lagunas existentes en una serie se utilizan los datos de una única serie y en este caso particular los datos de la misma serie que se quiere completar. Estos métodos tienen en cuenta la posible correlación entre los datos de la serie (autocorrelación) o el comportamiento físico de la variable (modelización física).

2.2.1.1.1 Autocorrelación

Definición: La autocorrelación es el estudio de la independencia de una serie de datos entre sí o estudiar si los datos de una serie son aleatorios entre sí. También se denomina *correlación serial o dependencia temporal*. Esto es, que exista dependencia entre un valor y el siguiente en la escala temporal. Este hecho es el que sucede en la gran mayoría de series meteorológicas, pero depende de la escala temporal de trabajo. A escalas diaria y horaria la autocorrelación suele ser alta, en cambio, para escalas mensuales y anuales la autocorrelación suele ser baja.

La autocorrelación se basa en las correlaciones lineales entre la serie original y la misma serie desplazada en 1, 2, 3, ..., k posiciones. A estas correlaciones se les llama autocorrelaciones de orden 1, 2, 3, ..., k respectivamente. Si los datos fueran independientes, todas las posibles autocorrelaciones serían iguales a cero y, en consecuencia, la función de autocorrelación, representación de las autocorrelaciones en función del desplazamiento, sería nula. Una manera de observar esto fácilmente es la representación del correlograma de la serie. Éste es una representación gráfica del valor de la función de autocorrelación respecto al desplazamiento en la serie. La gráfica debe mostrar una tendencia a acercarse a cero a medida que crece el número de paso o desplazamiento. Una vez estudiada la autocorrelación de la

serie, las lagunas que presenta se pueden sustituir básicamente a través de una serie de métodos sencillos, que se explican detalladamente más adelante, los cuales son los siguientes:

1. Criterio del dato anterior o posterior

Este método consiste en sustituir la falta de un dato por el valor correspondiente al registro anterior o posterior, en la escala temporal dada. Para escalas diarias, por el registro del día anterior o posterior; para escalas mensuales o anuales por el registro del mes o año anterior o posterior. Este método es de esperar que sea más útil para escalas temporales reducidas (diaria y horaria) que para escalas mensuales y anuales y para series que no tengan un ciclo muy marcado en la escala de estudio. Su utilización depende del estudio de los coeficientes de autocorrelación (que deben ser altos) o de otra forma que la serie presente dependencia temporal y además no tenga un ciclo muy marcado en la escala de estudio.

2. Criterio del valor medio

Este método consiste en sustituir la carencia de un registro por el valor medio de la serie, o del mes en cuestión si se trata de valores mensuales. Si la serie que se está completando es una serie estacionaria, este método produce un efecto perturbador mínimo o muy pequeño, en magnitud, frente a las anomalías de la variable (Rodríguez *et al.*, 2000). Es de esperar que este método sólo sea útil y fiable para series cuyos coeficientes de variación sean lo más pequeño posible, inferiores al 5%, o lo que es lo mismo, para series con desviaciones típicas mucho menores que su valor medio.

3. Criterio de las diferencias

El presente método consiste en sustituir la falta de un registro mensual por el valor del mes anterior, añadiendo el valor medio de las diferencias entre el mes anterior y el mes posterior. Este método es de esperar que sea válido únicamente para series que presenten una alta autocorrelación.

2.2.1.2. Métodos multivariantes aplicados a una única variable meteorológica

Los métodos incluidos en este apartado tienen en cuenta para rellenar los datos de una serie de una cierta variable meteorológica medida en un observatorio, series de la misma variable,

pero medida en diferentes observatorios. Todos los métodos se basan en el estudio de las correlaciones lineales individuales entre las series (correlación de Pearson). El algoritmo que se aplica para el rellenado es muy variado, así como el número de series que se utilizan. Los métodos que se engloban dentro de este grupo son los que aparecen a continuación, ordenados según su sencillez de aplicación:

1. Criterio de Karl (Karl y Williams, 1987)

Este método consiste en sustituir la carencia de un registro en una estación dada por el valor correspondiente de otra estación vecina (estación próxima a la estación problema) que presente un alto índice de correlación lineal. Este método requiere realizar los cálculos con los valores de las series previamente normalizados mensualmente y el coeficiente de correlación calculado es el de Pearson.

2. Razón normal (Paulus y Kohler, 1952)

Este método consiste en calcular el dato incompleto, $x(t)$, de una serie a partir de los datos de las series de tres estaciones vecinas y contemporáneas, que presenten un alto grado de correlación con la serie a completar, a través de la expresión:

$$x(t) = \frac{1}{3} \left[\frac{\bar{x}}{\bar{x}_1} x_1(t) + \frac{\bar{x}}{\bar{x}_2} x_2(t) + \frac{\bar{x}}{\bar{x}_3} x_3(t) \right] \quad \dots (1)$$

donde \bar{x} , \bar{x}_1 , \bar{x}_2 y \bar{x}_3 son las medias de las variables en cuestión de la serie incompleta y de las tres series vecinas, respectivamente; y $x_1(t)$, $x_2(t)$ y $x_3(t)$ son los datos correspondientes a las series vecinas, respectivamente. Este método juega con la variabilidad registrada en otras estaciones y con la razón proporcional entre ellas. Al ser tres estaciones se suaviza la influencia que podría tener un error en una de ellas.

3. Combinación lineal ponderada (CLP)

Este método consiste en sustituir la falta de datos a partir de los datos de series estadísticamente próximas, que son conocidas como vecinas. De tal manera que cada dato

incompleto se obtenga mediante la combinación lineal ponderada de los datos de las series que se van a utilizar para el completado. Dichos datos tienen un peso en la CLP proporcional al coeficiente de correlación de Pearson con la serie incompleta, siempre que éste sea superior a un valor crítico aceptable.

El valor que se suele coger depende del tipo de variable que se quiera completar, afinidad climática y alcance de la correlación espacial según el tipo de variable. Por ejemplo, para la precipitación, que es una variable que presenta, sobre todo en ambientes mediterráneos, una baja correlación entre estaciones vecinas (*Rodríguez et al., 1999*) el valor que se suele escoger es entre: $0,7 < r < 0,8$ (*González Hidalgo et al., 1999*). En cambio, para variables como la temperatura, la insolación o la presión atmosférica se pueden escoger valores de $r > 0,8$; ya que son variables que presentan mayores correlaciones con estaciones vecinas.

Por lo tanto, se tiene que para un mes t determinado, el dato incompleto $x(t)$ se puede expresar como:

$$x(t) = \frac{r_1 \cdot x_1(t) + r_2 \cdot x_2(t) + r_3 \cdot x_3(t) + \dots}{r_1 + r_2 + r_3 + \dots} \quad \dots (2)$$

donde, en general: r_i es el coeficiente de correlación de Pearson entre la serie i -ésima y la serie incompleta y $x_i(t)$ es el valor del instante t de la serie i -ésima. El número de series que se utilizan para el completado es arbitrario en principio. El valor que se suele coger está en torno a cinco y nunca inferior a dos (*Peterson y Easterling, 1994*).

Los coeficientes de correlación de Pearson deben calcularse con los valores de las series normalizados mensualmente.

4. Criterio de la correlación lineal

Este método consiste en el cálculo de los valores inexistentes por el ajuste, a través del método de los mínimos cuadrados, de una recta de regresión lineal del tipo $y = a \cdot x + b$; entre dos estaciones: una incompleta, cuya serie es la variable dependiente (y), y otra completa que proporciona la variable independiente (x) de la cual depende la anterior. La base teórica del cálculo admite como hipótesis de partida que entre las dos series a comparar existe una correlación apreciable, de tal modo que las variaciones que se producen en ambas son

similares. Esto quiere decir que la estación que se utiliza de referencia presenta un coeficiente de correlación lineal lo más cercano posible al valor 1. El coeficiente de correlación lineal que se calcula es el de Pearson. Para calcularlo, previamente hay que normalizar las series mensualmente.

Para que el ajuste sea correcto se han de cumplir una serie de requisitos como son: Fernández García y Galán (1994) o Fernández García (1995):

- a) *La existencia de un número de años comunes en ambas series.* Según los autores el número de años mínimos en la serie a completar oscila entre 5 y 15 años; pero cuantos más años con datos en común tengan ambas series, más creíble será la correlación encontrada entre ambas series. El número mínimo de años que se requieren, también depende de la variable meteorológica que se tenga en cuenta: por ejemplo, en el caso de series térmicas, más regulares, 5 años puede ser un umbral mínimo aceptable.
- b) *La proximidad geográfica de los observatorios* es necesaria, pero no suficiente, debido a las modificaciones que pueden introducir los factores locales del clima. Por lo tanto, es necesario que las dos estaciones presenten rasgos climáticos similares y por todo ello debe tenerse en cuenta, además, la correlación entre observatorios.
- c) *Comprobación de que la nueva serie resultante aporta un incremento a la información original.* Esto se consigue cuando el cociente entre la varianza de la muestra original y la muestra completada, es superior o igual a la unidad o lo que es lo mismo, cuando el coeficiente de determinación de la regresión, r^2 , cumple que:

$$r^2 \geq \frac{1}{n_1 - 2} = R^2 \quad \dots (3)$$

siendo n_1 la longitud de la serie incompleta.

5. Correlación múltiple (CORMUL)

Este método es válido para series homogéneas y tal que la autocorrelación temporal de los datos sea baja. Se utilizan los datos de series que presenten una alta correlación con la serie

de datos que se quiere completar. Cuanto mayor sea el número de series, menor será la sensibilidad al error en una de ellas. El método suele utilizarse para el completado de datos mensuales y por lo tanto requiere que se trabaje con valores normalizados mensualmente. El caso más sencillo de aplicar es en el que se escogen dos series de referencia (caso particular de la *regresión bivariada*).

5.1 Regresión bivariada

El proceso que hay que seguir para la realización de la regresión bivariada es el siguiente:

Establecimiento de la ecuación de regresión, a través de la cual se puede expresar la dependencia de los datos de la serie incompleta como una función lineal de los datos de las dos series de referencia. Esta ecuación, que recibe el nombre de ecuación *estocástica de regresión bivariada*, tiene el siguiente aspecto:

$$z_{ij} = a_1 \cdot z_{ij}^{(1)} + a_2 \cdot z_{ij}^{(2)} + \varepsilon_{ij} \quad \dots (4)$$

donde z_{ij} es el valor normalizado de la variable en cuestión del año i -ésimo y mes j -ésimo; los superíndices (1) y (2) hacen referencia a los datos de las dos estaciones que se utilizan en el proceso de completado y los coeficientes a_1 y a_2 son los coeficientes de la regresión parcial que se pueden expresar en función de los coeficientes de correlación simple r_1 , r_2 y r_{12} al añadir la condición de error cuadrático mínimo y ε_{ij} es un ruido independiente y normalmente distribuido de media 0 y desviación típica S_ε (ruido blanco).

De acuerdo a la mencionada condición de error cuadrático mínimo, las expresiones de los coeficientes de regresión a_1 y a_2 son:

$$a_1 = \left(\frac{r_1 - r_2 \cdot r_{12}}{1 - r_{12}^2} \right) \quad \dots (5)$$

$$a_2 = \left(\frac{r_2 - r_1 \cdot r_{12}}{1 - r_{12}^2} \right)$$

donde r_1 , r_2 y r_{12} son los coeficientes de correlación simple entre los valores normalizados de las estaciones 1 e incompleta; 2 e incompleta y entre 1 y 2, respectivamente. Los coeficientes de correlación simple se calculan a partir de la expresión de Pearson.

Se puede comprobar que la media del ruido ε_{ij} es cero, mientras que su desviación típica es función del error cuadrático cometido en la ecuación de regresión. Se puede estimar a partir de la expresión:

$$S_{\varepsilon}^2 = s^2 \cdot \left[1 - \left(R_{12}^{(i)} \right)^2 \right] \quad \dots (6)$$

donde s es la desviación típica de la serie incompleta y $R_{12}^{(i)}$ es el coeficiente de correlación múltiple entre los valores normalizados de las estaciones 1 y 2 con la serie incompleta y cuya expresión para calcularlo es:

$$R_{12}^{(i)} = \sqrt{\frac{r_1^2 + r_2^2 - 2 \cdot r_{12} \cdot r_1 \cdot r_2}{1 - r_{12}^2}} \quad \dots (7)$$

con valores comprendidos entre cero y uno.

Una vez elegidas las estaciones por su correlación con la dada y por su disponibilidad de datos, se ajusta la ecuación de regresión. En el momento de rellenar un dato se debe elegir aquella pareja de estaciones que proporcione el completado más satisfactorio. El criterio que se debe seguir para realizar tal elección se basa en la formación de una *matriz de priorización* para cada estación a completar, función de los coeficientes de correlación múltiple entre las series normalizadas y del número de datos comunes entre las tres estaciones (la pareja que se utiliza en el completado y la serie incompleta que se quiere rellenar). Entonces, en general, se tiene que los términos de la matriz de priorización para la estación a completar k , para la pareja de estaciones de referencia m y n , se expresan de la siguiente manera:

$$P_{mn} = R_{mn}^k \cdot \left(\frac{N_{mnk}}{12 \cdot N} \right)^a \quad \dots (8)$$

donde P_{mn} es el elemento (m,n) de la matriz de priorización correspondiente a la estación a completar k ; R_{mn}^k el coeficiente de correlación múltiple entre los valores normalizados de la estación k y las estaciones m y n de referencia; N_{mnk} el número de datos mensuales en común (registrados a la vez) de las tres estaciones m , n y k ; N el número de años de la serie incompleta; y a , el *exponente de priorización*. Tal exponente suele ser un dato que se introduce para ponderar la importancia que tiene en el completado el número de datos

comunes entre las tres estaciones, aunque se puede determinar experimentalmente bajo el criterio de generar series con mínimo error cuadrático.

Halladas las matrices de priorización de cada estación, se puede elegir cuál es la mejor pareja para rellenar un mes en cuestión. En principio se escogerá para cada estación k la pareja de referencia con un mayor valor de priorización, pero la existencia de los valores en el mes de completado en esas estaciones n y m , origina que se tengan que elegir sucesivas parejas de estaciones de acuerdo a valores decrecientes de priorización hasta un valor mínimo llamado *umbral de priorización*. Debido a que las series que se han utilizado en la ecuación de regresión están normalizadas, los valores que se obtienen para rellenar los huecos también están normalizados y, por lo tanto, hay que deshacer la transformación de los resultados. Esto se consigue aplicando la siguiente expresión:

$$x_{ij} = \bar{x}_j + s_j \cdot z_{ij} \quad \dots (9)$$

donde x_{ij} y z_{ij} son, respectivamente, los valores completos y normalizados correspondientes al mes j -ésimo del año i -ésimo de la estación que se ha completado; y \bar{x}_j y s_j son el valor medio y la desviación estándar del mes j -ésimo de la serie antes de completarla.

2.2.2. CONTROL DE CALIDAD DE DATOS

Los registros climáticos instrumentales históricos de alta calidad constituyen una pieza clave para la realización de estudios robustos, consistentes y fiables que permitan la mejor detección, comprensión, predicción y respuesta a los impactos asociados a la variabilidad y cambio climático. La obtención de bases de datos climáticos de alta calidad y alta resolución temporal y espacial es esencial para los estudios de calibración de datos de satélite o la generación de productos en malla y de re-análisis.

El control de calidad de los datos climáticos es una obligación tanto para los servicios meteorológicos responsables de su registro y custodia, que deberían ofrecer mediciones correctas en un alto porcentaje o, en caso contrario, advertir al usuario de los problemas existentes, identificando los intervalos no utilizables, como para los investigadores del clima que requieren validar esa información antes de usarla, para no introducir señales no

climáticas en sus resultados y prevenir errores potenciales que puedan comprometer los estudios realizados.

En una serie temporal de observaciones meteorológicas existen varias fuentes potenciales de errores y/o problemas tales como:

- 1) Ausencia sistemática de datos.
- 2) Datos observados/registrados incorrectamente .
- 3) Errores de transcripción durante la digitalización, etc.

Algunas prácticas siguen un patrón, como la ausencia de datos en fines de semana o feriados (*Nicholls, 1996; Revfeim, 1990*), que influyen en gran medida en el cálculo de los valores acumulados y extremos, usualmente forzando al investigador a reformular su tratamiento.

Algunos errores pueden ocurrir en el proceso de registro, manipulación, formateo, transmisión y archivo de datos y metadatos.

Un conjunto de datos no confiables o que no se está disponible cuando se necesita, no sirve para ningún propósito. Tener datos “malos”, es decir, de mala calidad, es como carecer por completo de información.

O peor, sin control de calidad, los errores podrían pasar desapercibidos y dar como resultado productos erróneos que podrían generar decisiones equivocadas.

Para ello conviene efectuar una adecuada elección de los métodos de control. De esto depende en gran manera el control de calidad y los datos disponibles en el momento de la comprobación. Casi todos los métodos de control de calidad de las observaciones meteorológicas tienen carácter estadístico y siempre habrá cierto riesgo de rechazar valores correctos (y de aceptar valores erróneos). Por ello es necesario que el control de calidad sea ejecutado por un experto, que pueda evaluar cada caso donde el sistema de control automático detecte un dato “dudoso”.

El conocimiento de los procedimientos aplicados permitiría la evaluación de la validez de las observaciones y la mejora en la calidad de la información (*WMO, 2011*).

El control no consiste únicamente en la identificación de posibles errores, sino que los datos que no superen las diferentes pruebas del mismo deben ser etiquetados como sospechosos y debe consultarse en la fuente original si se han introducido o importado correctamente, lo que permite recuperar una fracción importante de los datos sospechosos de ser erróneos, ya sea validándolos o sustituyéndolos por los valores correctos. Únicamente aquellos datos sospechosos que no puedan validarse o recuperarse son eliminados de la base de datos.

Los métodos de verificación de la coherencia interna y de la coherencia temporal pueden parecer más complejos, pero no resultan laboriosos una vez automatizados.

Cada uno de los pasos se describe a continuación:

1. Verificaciones de los límites para detectar errores graves.

Se verifica que los valores reportados no salgan de un intervalo dado, que teóricamente contiene todo el rango posible de observaciones. Se aconseja que estos límites se definan mensualmente o como mínimo cada estación del año. Por ejemplo, para cierta localidad, se puede establecer que en verano los valores de temperatura no pueden estar por debajo de 0°C ni por encima de 50°C.

Las comparaciones con límites que dependen de la zona y del momento del año pueden realizarse para la velocidad del viento, la temperatura del aire, la temperatura del punto de rocío, la presión en la estación, entre otras. Los límites pueden modificarse y tal vez reducirse sobre la base de estadísticas climatológicas mejores o de la experiencia del técnico a cargo.

Estas verificaciones de límites solo permiten detectar errores de observación muy toscos. Sin embargo, ofrecen la ventaja de que requieren recursos limitados. El resultado de estas verificaciones puede ser una clasificación en valores correctos, sospechosos o erróneos.

2. Verificaciones de la coherencia interna

Se realizan a partir de la comparación entre datos del mismo reporte (momento) de una estación. Los distintos parámetros o valores de los datos reportados por las estaciones se verifican entre sí y se consideran sospechosos o erróneos los que no concuerdan con el resto de la información. Por ejemplo, es sospechoso que, al estar lloviendo, la humedad relativa sea del 50 %. O es erróneo que la temperatura máxima de la última hora sea menor que la mínima.

3. Verificaciones de la coherencia temporal

Para verificar la coherencia en relación con el tiempo, se utiliza la información de reportes consecutivos procedentes de la misma fuente. Se establece una tolerancia para la diferencia entre el valor de una variable en un momento de observación y el valor de esta en la observación siguiente. Si la diferencia es demasiado grande, el dato se considera sospechoso. Por ejemplo, será sospechoso si en un reporte dado la temperatura es de 22°C y en el reporte de la hora siguiente es de 38°C.

4. Verificaciones de la coherencia espacial (prueba de contigüidad)

La información de estaciones meteorológicas cercanas puede utilizarse para realizar comparaciones. Se compara cada par de observaciones próximas: si son similares, es probable que sean ambas correctas o ambas erróneas; si difieren mucho, una de ellas es probablemente correcta y la otra errónea.

En los métodos de interpolación estadística, es posible interpolar o analizar cada una de las observaciones de una zona utilizando las observaciones del resto del área. En el procedimiento de interpolación se excluye la observación por analizar y se obtiene un valor interpolado. Este valor preliminar debería ser similar a la observación. Si no, es muy probable que la observación sea errónea.

2.2.2.1 Control de Calidad de Datos de lluvia

El control de la calidad de los datos de las precipitaciones es semiautomático; es decir, tiene una primera fase automatizada de detección de valores “dudosos” y una segunda fase manual que consiste en la revisión de los datos dudosos identificados por un experto debidamente capacitado. Se procede a corregir los errores obvios de los datos de precipitación, en la medida de lo posible. A los datos que son indiscutiblemente incorrectos se les asigna la clave con que se identifican los valores faltantes, como por ejemplo “-1” o “999”.

La parte automática del procedimiento de control no ha sido concebida para la corrección de los datos, sino para reducir el número de datos que requieren control visual. Un control de la calidad totalmente automático eliminaría todos los datos dudosos y, dada la elevada

variabilidad de la precipitación, se estarían eliminando seguramente también gran cantidad de datos correctos, en particular los valores extremos.

Para conservar los datos extremos verdaderos y, al mismo tiempo, eliminar los datos que claramente son incorrectos del sistema de análisis, evidentemente es imprescindible una verificación visual de los datos dudosos, aunque resulte muy laboriosa.

2.2.2.1.1 Tipos de errores detectados

A menudo, los errores en los metadatos se manifiestan como errores en los datos. Por ejemplo, un identificador de estación incorrecto puede significar que los datos correspondientes a una estación aparentemente proceden de otra, o un sello con la fecha incorrecta puede implicar que los datos aparentemente fueron observados en otro momento.

La indisponibilidad de los datos correspondientes a la estación y al momento correcto debería detectarse mediante pruebas que verifiquen si figuran todos los datos; los datos que se hayan atribuido a un emplazamiento y un momento incorrectos deberán detectarse mediante pruebas de coherencia y de tolerancia.

Los errores en los datos pueden surgir principalmente como resultado de errores atribuibles a los instrumentos, al observador o a los procesos de transmisión de datos, la introducción de datos mediante teclado y su validación, así como a la modificación de los formatos de datos y los problemas que implica resumirlos.

Al establecer un conjunto de procedimientos de control de calidad, deberían considerarse todos los tipos, fuentes y causas posibles de errores y debería hacerse todo lo posible para reducirlos.

2.2.2.1.2 Pruebas de Coherencia

Los cuatro tipos principales de prueba de coherencia son:

- Prueba de coherencia interna.
- Prueba de coherencia temporal.
- Prueba de coherencia espacial.

- Prueba de coherencia sumaria.

Dado que los valores de los datos guardan una relación mutua tanto temporal como espacial, debería elaborarse un procedimiento integrado para examinar la coherencia.

Todas las pruebas de coherencia deberían documentarse por completo mediante procedimientos, fórmulas y criterios de decisión.

1. La coherencia interna, se basa en las relaciones físicas que existen entre los elementos climatológicos. Todos los elementos deberían verificarse exhaustivamente comparándolos con otros elementos afines dentro de cada observación. Por ejemplo, deberían verificarse todos los datos psicométricos para garantizar que la temperatura de bulbo seco notificada sea igual o superior a la temperatura de bulbo húmedo notificada. Del mismo modo, deberá comprobarse que la relación entre la visibilidad y el tiempo reinante se ajuste a las prácticas de observación normalizadas.

Se debería verificar la coherencia entre los datos y las definiciones. Por ejemplo, un valor máximo debe ser igual o mayor que un valor mínimo. Los límites físicos imponen reglas para las verificaciones de la coherencia interna. Por ejemplo, la precipitación no puede ser negativa.

2. Las pruebas de coherencia temporal examinan la variación de un elemento en el tiempo. Muchos conjuntos de datos climatológicos muestran una importante correlación en serie. Debería hacerse un control comparando la observación en cuestión con la anterior y la posterior en el ordenamiento.

Basándose en la experiencia o en métodos analíticos o estadísticos, quienes examinen los datos pueden establecer la magnitud del cambio que podría experimentar un determinado elemento en cualquier intervalo de tiempo. Por lo general, este cambio depende del elemento, la estación del año, la ubicación y el lapso de tiempo transcurrido entre dos observaciones sucesivas. Por ejemplo, un descenso de 10 °C en la temperatura al cabo de una hora puede ser sospechoso, pero podría resultar muy factible si se le relacionara con el paso de un frente frío o de una brisa marina.

Las presentaciones gráficas también son una herramienta excelente para la verificación. Para facilitar el diagnóstico, deberán visualizarse varios elementos al

mismo tiempo. Por ejemplo, será más fácil validar un descenso de la temperatura si también se dispone de la información que muestra el cambio de dirección de los vientos relacionado con el paso de un frente frío o la lluvia fuerte de una tormenta.

3. La coherencia espacial, compara cada observación con observaciones efectuadas al mismo tiempo en otras estaciones de la zona. Cada observación puede compararse con las que cabría esperar en el emplazamiento tomando en cuenta las observaciones de las estaciones vecinas. Aquellos datos que difieren notablemente entre las observaciones esperadas y las que, de hecho, se obtienen deberán marcarse para ser sometidos a una revisión, corrección o eliminación, según proceda.

Las pruebas de reducción de datos son unas de las más fáciles de efectuar. Mediante la comparación de diferentes resúmenes de datos, pueden detectarse errores en cada valor o en cada resumen. Por ejemplo, pueden calcularse las sumas y los promedios de los valores diarios correspondientes a diversos períodos como semanas, meses o años.

2.2.2.1.3 Pruebas de Tolerancia

Las pruebas de tolerancia establecen límites superiores o inferiores sobre los posibles valores de un elemento climatológico (tal como la dirección del viento, la nubosidad y las condiciones meteorológicas pasadas y presentes), o en aquellos casos en que la gama teórica de valores es infinita y es poco probable que una medición se sitúe fuera de sus límites.

En el segundo caso, por lo general, los límites suelen depender del tiempo y la ubicación y deberían establecerse recurriendo a los valores históricos o mediante métodos de interpolación espacial.

También es importante determinar y remediar rápidamente los sesgos sistemáticos que presenta la información proporcionada por los instrumentos.

Debe conservarse documentación con información de las pruebas de tolerancia que se han efectuado, los límites climáticos establecidos respecto de cada elemento estudiado y las razones fundamentales para determinar dichos límites.

En general, las pruebas de tolerancia comparan un valor en cuestión con algún patrón que utilice un umbral estadístico. Algunas de las pruebas de tolerancia simples consisten en comparar un valor observado con el valor extremo o récord o con algún múltiplo de desviaciones típicas en torno al valor promedio correspondiente a esa fecha.

En el segundo caso, se ha de tener en cuenta la posibilidad de que tal vez el elemento no tenga necesariamente una distribución simétrica o gaussiana, o que algunos valores extremos determinados a partir del multiplicador de la desviación típica puedan ser incorrectos.

Cuando se utilicen datos históricos de largo período para el control de calidad, es preferible utilizar una referencia normalizada (por ejemplo, las desviaciones típicas o una estadística de clasificación por orden no paramétrica) en lugar de una referencia absoluta.

2.2.2.2 RCLIMDEX

ClimDex es un programa basado en Microsoft Excel que proporciona un paquete computacional fácil de usar para el cálculo de Índices de extremos climáticos para monitorear y detectar cambio climático. Fue desarrollado por Byron Gleason del National Climate Data Centre (NCDC) de NOAA, y ha sido usado en talleres CCI/CLIVAR sobre índices climáticos desde el 2001.

El objetivo original fue el de colocar ClimDex en un ambiente que no dependa de un sistema operativo determinado. Fue muy natural usar R como nuestra plataforma, dado que R es gratis; pero al mismo tiempo, es un software muy robusto y poderoso para gráficos y análisis estadístico. Puede ser ejecutado tanto en ambientes Windows como Unix. En el 2003 se descubrió que el método utilizado por ClimDex y otros programas para calcular índices de temperatura basados en percentiles resultaba en inhomogeneidad en las series de índices. Una solución a este problema requiere un procedimiento "bootstrap" que se hace imposible de implementar en un ambiente Excel. Esto ha hecho aún más urgente la necesidad de desarrollar este paquete basado en R.

RClimDex está diseñado para proporcionar una interfase amigable para calcular índices de extremos climáticos. Calcula todos los 27 índices básicos recomendados por el Equipo de Expertos de CCI/CLIVAR para "Climate Change Detection Monitoring and Índices"

(ETCCDMI) así como también otros índices de temperatura y precipitación con límites definidos por el usuario.

Actualmente RClimDex incluye solamente un procedimiento simple de control de calidad de datos que fue proporcionado por ClimDex. Como en ClimDex, se requiere que los datos tengan un control de calidad antes de que los índices sean calculados.

2.2.2.2.1. Control de Calidad de RClimdex

El control de calidad (CC) de RClimDex desarrolla el siguiente procedimiento:

- 1) Reemplaza todos los datos faltantes (actualmente codificados como -99.9) en un formato interno que reconoce R (i.e. NA, no disponible).
- 2) Remplaza todos los valores no razonables por NA. Estos valores incluyen:
 - a) Cantidades de precipitación diarias menores que cero.
 - b) Temperatura máxima diaria menor que temperatura mínima diaria.

Adicionalmente, CC también identifica valores extremos (“outliers”) en temperaturas diarias máximas y mínimas. Los valores extremos son valores diarios que se encuentran fuera de una región definida por el usuario. Actualmente esta región se define como n veces la desviación estándar del valor del día, esto es, $[media - n*std, media+n*std]$. Donde std representa la desviación estándar para el día y n es una entrada del usuario.

Una vez completado el control de calidad con esta herramienta se generarán dos archivos que contengan información sobre valores no razonables para temperatura y precipitación, un tercer archivo marcará todos los valores extremos posibles en temperatura diaria junto con las fechas en que aquellos valores ocurrieron y un último archivo contendrá los datos una vez hecho el control de calidad y que serán usados para posteriores cálculos. En el mismo, se puede notar que solamente los valores faltantes y no razonables son reemplazados por NA y los posibles valores extremos que sean marcados NO serán cambiados. Para una visualización más fácil, se generarán cuatro archivos PDF que contendrán gráficos de las series de tiempo (valores faltantes graficados como puntos rojos) de precipitación diaria, y las temperaturas diarias máximas y mínimas también serán guardadas.

Para que se puedan obtener los resultados esperados de esta herramienta, se necesita que los archivos de datos de entrada que serán leídos estén en el formato que se muestra a continuación, además de cumplir los siguientes requisitos:

1. Archivo de texto
2. Columnas en las siguientes secuencias: Año, Mes, Día, PRCP, TMAX, TMIN.
(NOTA: unidades de PRCP= milímetros y unidades de Temperatura= grados Celsius)
3. El formato descrito arriba debe ser delimitado por espacios (cada elemento separado por uno o más espacios).
4. Para los registros de datos, los datos faltantes deben ser codificados como -99.9; los registros de datos deben estar en orden cronológico. Se permiten fechas faltantes.

Ejemplo de formato de datos para el archivo de datos inicial (usados en el paso de Control de Calidad):

1901	1	1	-99.9	-3.1	-6.8
1901	1	2	-99.9	-1.3	-3.6
1901	1	3	-99.9	-0.5	-7.9
1901	1	4	-99.9	-1	-9.1
1901	1	7	-99.9	-1.8	-8.4

2.2.2.2.2. Análisis de los resultados de RCLimindex

Para realizar una interpretación de los resultados generados por RCLimindex se necesita tener conocimiento acerca de los diagramas de caja, los mismos son una presentación visual que describe varias características importantes, al mismo tiempo, tales como la dispersión y simetría.

Para su realización se representan los tres cuartiles y los valores mínimo y máximo de los datos, sobre un rectángulo, alineado horizontal o verticalmente.

La construcción de una gráfica de este tipo (diagrama de caja) consiste en una caja rectangular, donde los lados más largos muestran el recorrido intercuartílico.

Este rectángulo está dividido por un segmento vertical que indica dónde se posiciona la mediana y por lo tanto su relación con los cuartiles primero y tercero (recordemos que el segundo cuartil coincide con la mediana).

Esta caja se ubica a escala sobre un segmento que tiene como extremos los valores mínimo y máximo de la variable. Las líneas que sobresalen de la caja se llaman bigotes. Estos bigotes tienen un límite de prolongación, de modo que cualquier dato o caso que no se encuentre dentro de este rango es marcado e identificado individualmente.

Los pasos para interpretar un diagrama de caja son:

1. Evaluar las características clave

Se debe examinar el centro y la dispersión de la distribución, además de evaluar cómo el tamaño de la muestra puede afectar la apariencia de la gráfica de caja.

Centro y dispersión. - Examine los siguientes elementos para conocer más acerca del centro y la dispersión de sus datos de muestra.

Mediana. - La mediana está representada por la línea en la caja. La mediana es una medida común del centro de sus datos. La mitad de las observaciones es menor que o igual al valor y la mitad es mayor que o igual al valor.

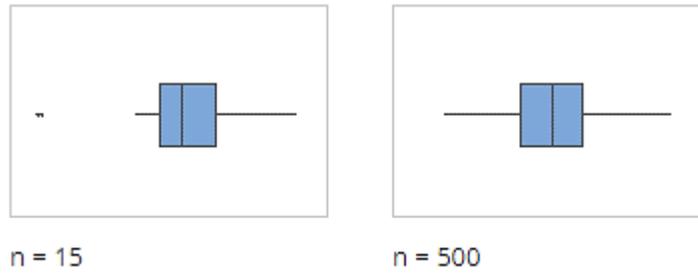
Caja de rango intercuartil. - La caja de rango intercuartil representa el 50% intermedio de los datos. Muestra la distancia entre el primer cuartil y el tercer cuartil ($Q_3 - Q_1$).

Bigotes. - Los bigotes se extienden de cualquier lado de la caja. Los bigotes representan los rangos del 25 % de valores de datos de la parte inferior y el 25 % de la parte superior, excluyendo los valores atípicos.

Tamaño de la muestra (n). - El tamaño de la muestra puede afectar la apariencia de la gráfica.

Por ejemplo, aunque estas gráficas de caja parecen ser muy diferentes, ambas se crearon utilizando muestras seleccionadas aleatoriamente a partir de la misma población.

Figura N°3 Tamaño de la muestra



Una gráfica de caja funciona mejor cuando el tamaño de la muestra es al menos de 20. Si el tamaño de la muestra es demasiado pequeño, los cuartiles y los valores atípicos que la gráfica de caja muestra pueden no ser significativos.

2. Buscar indicadores de datos inusuales o no normales

Los datos asimétricos indican que los datos podrían ser no normales. Los valores atípicos pueden indicar otras condiciones en sus datos.

Datos asimétricos. - Cuando los datos son asimétricos, la mayoría de los datos se ubican en la parte superior o inferior de la gráfica. La asimetría indica que los datos pueden no estar distribuidos normalmente.

Estas gráficas de caja ilustran los datos asimétricos. La gráfica de caja con datos asimétricos hacia la derecha muestra tiempos de espera. La mayoría de los tiempos de espera son relativamente cortos y solo unos pocos tiempos de espera son largos. La gráfica de caja con datos asimétricos hacia la izquierda muestra datos de tiempo de falla. Unos pocos elementos fallan inmediatamente y muchos más fallan posteriormente.

Figura N°4 Datos asimétricos

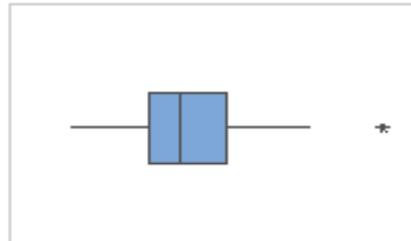


Si se está en conocimiento de que sus datos no son naturalmente asimétricos, investigue las posibles causas.

Valores atípicos. - Los valores atípicos, que son valores de datos que están muy alejados de otros valores de datos, pueden afectar fuertemente sus resultados. Frecuentemente, es más fácil identificar los valores atípicos en una gráfica de caja.

En una gráfica de caja, los valores atípicos se identifican con asteriscos (*).

Figura N°5 Valores Atípicos



Se debe identificar la causa de cualquier valor atípico y corregir cualquier error de entrada de datos o de medición. Se debe considerar eliminar los valores de datos que estén asociados con eventos anormales y únicos (causas especiales).

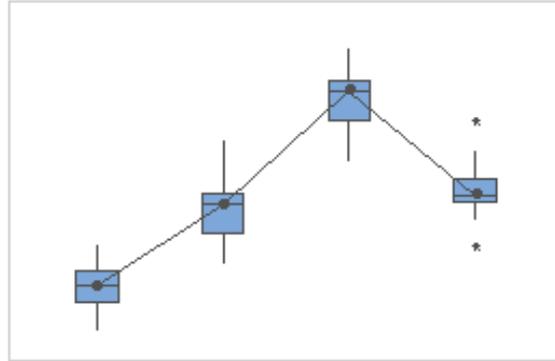
3. Evaluar y comparar los grupos

Cuando la gráfica de caja tiene grupos, se debe evaluar y comparar el centro y la dispersión de los grupos.

Centros. – Para el análisis de los mismos se debe buscar diferencias entre los centros de los grupos.

Por ejemplo, en la Figura N°6 el diagrama de caja muestra el grosor de cable producido por cuatro proveedores. Los grosores medios de algunos grupos parecen ser diferentes.

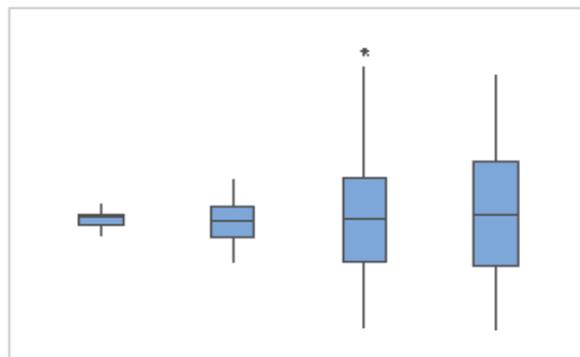
Figura N°6 Centros de los grupos



Dispersiones. –Se debe buscar diferencias entre las dispersiones de los grupos.

Por ejemplo, en la Figura N°7 el diagrama de caja muestra los pesos de llenado de las cajas de cereales de cuatro líneas de producción. Los pesos medios de los grupos de cajas de cereales son similares, pero los pesos de algunos de los grupos son más variables que los otros.

Figura N°7 Dispersiones de los grupos

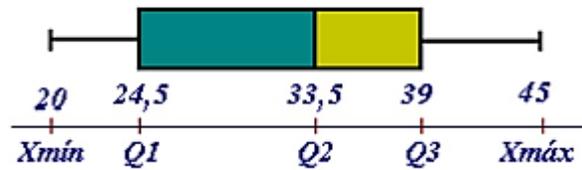


Para determinar si una diferencia en la dispersión (varianza) es estadísticamente significativa, se puede realizar uno de los siguientes procedimientos:

- Prueba de dos varianzas si solamente se tiene dos grupos.
- Prueba de varianzas iguales si solamente tiene tres o más grupos.

A continuación, se muestra un ejemplo de análisis de un diagrama de caja (Figura N°8).

Figura N°8 Diagrama de Caja



De la gráfica, el bigote de la izquierda representa al colectivo de datos (X_{\min} , Q_1), la primera parte de la caja a (Q_1 , Q_2), la segunda parte de la caja a (Q_2 , Q_3) y el bigote de la derecha viene dado por (Q_3 , X_{\max}).

Se puede obtener abundante información de una distribución a partir de estas representaciones. De la figura se puede analizar lo siguiente:

- La parte izquierda de la caja es mayor que la de la derecha; ello quiere decir que los datos comprendidos entre el 25% y el 50% de la población está más dispersa que entre el 50% y el 75%.
- El bigote de la izquierda (X_{\min} , Q_1) es más corto que el de la derecha; por ello el 25% de los datos menores están más concentrados que el 25% de los mayores.
- El rango intercuartílico= $Q_3 - Q_1 = 14,5$; es decir, el 50% de la población está comprendido en 14,5.

2.2.2.3. DECADE

Es una “herramienta” dedicada netamente al control de calidad visual. La misma es una rutina que corre bajo la plataforma R, desarrollado por Stefan Hunziker para el control de calidad de las series diarias de precipitación (PRCP), temperatura máxima (TX) y temperatura mínima (TN) del proyecto DECADE.

Los resultados que se obtendrán para su posterior análisis visual son los siguientes:

- Gráficos de puntos de las series temporales (PRCP; TX; TN)

Permitirá que veamos cada dato y año registrado, representado con un punto en la gráfica.

- Histograma de decimales de la serie de datos (PRCP; TX; TN)

Veremos en diferentes colores los decimales de los datos lo que nos permitirá establecer la frecuencia con la que aparecen los mismos.

- Histograma temporal (años) de datos faltantes (PRCP; TX; TN)

Con esto podremos ver los datos faltantes por año, considerando tanto datos consecutivos como no consecutivos, se nos permitirá establecer la cantidad de días faltantes por año.

- Diagrama de barras de días que probablemente sean indicadores de ciclos semanales (PRCP)

Podremos observar el día que sería el candidato a producir ciclos semanales en la serie de acuerdo al test binomial aplicado a cada día de la semana para toda la serie, esto se observa para cada estación que se introducirá.

- Archivos de estadística de datos faltantes (PRCP; TX; TN)

Se generarán archivos para cada una de las variables a estudiar (PRCP; TX; TN) en el que se podrá verificar cuantas estaciones se han evaluado y cuantas de ellas tienen más de veinte años de registro o un registro inferior a lo antes mencionado.

Para que se puedan obtener los resultados esperados de esta herramienta se necesita que los archivos de datos de entrada que serán leídos estén en el formato que se muestra a continuación, además de cumplir los siguientes requisitos:

1. Archivo de texto.
2. Columnas en las siguientes secuencias: Día juliano, Año, Mes, Día, Dato, COD1, COD2, COD3.
3. El formato descrito arriba debe ser delimitado por espacios (cada elemento separado por uno o más espacios).
4. Para los registros de datos, los datos faltantes deben ser codificados como -999.9 = NA, además serán codificados, en temperatura los valores $< -60^{\circ}\text{C}$ y los valores $> 60^{\circ}\text{C}$ y en precipitación los valores < 0 mm (valores negativos).

Ejemplo de formato de datos para el archivo de datos inicial

2431091	1944	1	1	-999.900	000	00	000
2431092	1944	1	2	-999.900	000	00	000
2431093	1944	1	3	-999.900	000	00	000
2431094	1944	1	4	-999.900	000	00	000
2431095	1944	1	5	-999.900	000	00	000
2431096	1944	1	6	-999.900	000	00	000

2.2.3. TEST DE HOMOGENEIZACIÓN

La información climática puede no representar la variación del tiempo actual debido a fallas en los instrumentos de medición, errores ocasionados accidentalmente por la persona responsable en la toma de datos, la ubicación en la que se encuentra la estación, entre otras. Esto ocasiona variaciones en los datos reales, provocando que el usuario de información climática obtenga resultados o realice inferencias poco confiables. Es por ello que contar con un conjunto de datos meteorológicos de largo plazo homogéneos es actualmente de interés para la comunidad científica.

La fiabilidad de la información climática, así como su precisión resulta de gran importancia en la actualidad para el desarrollo de estudios relacionados con el cambio climático, por ello ésta debe ser validada para desarrollar cualquier aplicación. Para lograr este cometido es necesario aplicar metodologías de verificación y tratamiento que permitan establecer qué estaciones cumplen el supuesto de homogeneidad en variables de temperatura mínima, temperatura máxima y precipitación.

El fin es discriminar entre las homogeneidades de tipo “climatológico” debidas a las variaciones de la atmósfera, de las homogeneidades “no climatológicas” observadas en las series de datos debidas a la alteración de la observación, las cuales causan rupturas abruptas en los datos: cambio de instrumentos medidores; de observador; de instalación; de lugar; de modificaciones próximas del entorno; caída de la calidad de las observaciones, etc. (Martínez Molina 1986).

Según Conrad y Pollack (1962); “una muestra es homogénea si sus variaciones responden exclusivamente a las variaciones de la atmósfera”.

Al realizar los test expone; “muchas alternancias indican oscilación y muchas permanencias indican una tendencia o cambio”. Así un número de rachas elevados indican oscilaciones en torno a la mediana de la serie, mientras que un número de rachas bajo indica tendencias.

Para diferenciar las inhomogeneidades “climáticas” de las restantes, se realiza una comparativa de las observaciones frente a los observatorios y colaboradores más cercanos.

Para su análisis, existen diferentes test de control de homogeneidad, aplicados a series anuales, comparaciones entre dos estaciones cercanas para encontrar ciertos factores de calidad, los mismos se explican a continuación:

2.2.3.1. Test del recorrido

Este método consiste en evaluar los cambios de signo entre cada uno de los elementos de la serie en relación a la mediana y compara el número de rachas o cambio con los umbrales correspondientes a una distribución de frecuencias de una serie aleatoria cuyos valores son conocidos. Si el número de rachas está dentro del límite establecido, la serie se considera homogénea, si existen muchas rachas indica una oscilación en la mediana durante el periodo de la muestra. (Fernández, 1996).

Etapas del Procedimiento:

1. Estimación del valor medio de la serie (usamos la mediana si la muestra es chica).
2. Cálculo de los desvíos de cada elemento respecto al valor medio. Se asigna a cada valor de la serie el signo correspondiente, (+) si está el valor de la serie por encima de la media y (-) si está por debajo.

Para probar si la serie climática es homogénea se utiliza el criterio de Doorembos, y según la Tabla N°2 si el número de cambios está dentro del rango admitido la serie analizada es homogénea, en caso contrario no es homogénea.

Tabla N° 2 Criterio de Doorembos

N° observ.	Intervalo	N° observ.	Intervalo
12	5-8	32	13-20
14	5-10	34	14-21
16	6-11	36	15-22
18	7-12	38	16-23
20	8-13	40	16-25
22	9-14	50	22-30
24	9-16	60	26-36
26	10-17	70	31-41
28	11-18	80	35-47
30	12-19	90	40-52
		100	45-57

Fuente: Gomez, L., Medina M., *Guía de Métodos Estadísticos en Climatología*

Si al aplicar la prueba de recorrido vemos que la serie no es homogénea, debe buscarse la causa y ya no se aplica ninguna otra prueba. Las pruebas no son condiciones necesarias y suficientes para afirmar que una serie sea homogénea. Mientras no haya una prueba que diga que una serie no es homogénea, será homogénea. Las pruebas paramétricas son más potentes que las no paramétricas, pero presuponen cierta distribución de la serie.

Para aplicar este método es necesario verificar que cuente con los datos, esto incluye que la serie esté completa (un 100 y 90 %), además se requieren datos del promedio mensual de la precipitación.

2.2.3.2. Método de los residuales acumulados

Cuando se relacionan dos registros de datos provenientes de dos estaciones meteorológicas distintas, de las cuales la primera se considera homogénea, el registro de la segunda estación se puede considerar homogéneo si los residuales acumulados del segundo registro a partir de una línea de regresión basada en el primer registro no se desvían notoriamente. El grado de desviación se puede probar para una probabilidad dada p . Esto se lleva adelante verificando si las residuales pueden ser incluidas dentro de una elipse con ejes α y β . Las magnitudes α y β dependen de la longitud del registro, de la desviación estándar de la muestra que está siendo analizada y de la probabilidad p usada.

Esta prueba utiliza resultados de residuales de la regresión lineal de Y y X. Los residuales deben seguir una distribución normal con media cero y desviación estándar s_y, x , es decir el error $\epsilon_i \in N(0, s_{y,x})$. Los residuales de la regresión deben ser considerados como variables independientes aleatorias (es decir, deben exhibir homoscedasticidad).

El procedimiento para analizar la homogeneidad de un registro climático Y_i recogido en una estación meteorológica dada, puede ser resumido como sigue:

1. Selección de una estación meteorológica de referencia que se sabe que posee un registro homogéneo X_i de la variable analizada dentro de la misma región climática. Como alternativa, se puede construir un registro homogéneo regional haciendo un promedio de observaciones de varias estaciones meteorológicas en la misma región.
2. Organización de los registros x_i y y_i en orden cronológico $i = 1, 2, \dots, n$, de tal manera que la fecha de inicio de registro, así como el lapso entre datos sea idéntico en ambos.
3. Para ambos registros, se debe calcular la media y la desviación estándar para la variable homogénea (x_i) y para la variable analizada (y_i).
4. Cálculo y trazado de la línea de regresión entre las dos variables y_i y x_i y el coeficiente de correlación asociado. La ecuación de regresión entre ambos sistemas se expresa como

$$\hat{y}_i = af + bf x_i \quad \dots(10)$$

donde el subíndice f se refiere al sistema completo. Siempre que sea posible, se debe trazar el diagrama x_i, y_i y la línea de regresión para verificar visualmente si la hipótesis de homoscedasticidad puede ser aceptada.

5. Cálculo de los residuales de los valores observados y_i con respecto a la línea de regresión, de la desviación estándar s_y, x de los residuales y del E_i residual acumulativo correspondiente:

$$\epsilon_i = Y_i - \hat{Y}_i \quad \dots (11)$$

$$S_{y,x} = S_Y (n - r^2)^{1/2} \quad \dots (12)$$

$$E_i = \varepsilon_i + \sum_{j=1}^{i-1} \varepsilon_j \quad \dots (13)$$

6. Selección de un nivel de probabilidad p para aceptar la hipótesis de homogeneidad. El valor p = 80% es comúnmente utilizado. Con esto se calcula la ecuación de la elipse con ejes:

$$\alpha = n/2 \quad \dots (14)$$

$$\beta = \frac{n}{(n-1)^{1/2}} Z_p S_{y,x} \quad \dots (15)$$

Donde:

n: tamaño de la muestra bajo análisis

Z_p : variante estándar normal para la probabilidad p (generalmente p = 80% para no excedencia).

$s_{y,x}$: desviación estándar de los residuales

La ecuación paramétrica de la elipse entonces será:

$$X = \alpha \cos(\theta) \quad \dots (16)$$

$$Y = \beta \cos(\theta) \quad \dots (17)$$

con θ [rad] variando de 0 a 2π .

Puede por lo tanto concluirse que en el nivel de probabilidad p, no hay desviaciones en la distribución de residuales, y, por tanto, el registro y_i puede ser considerado homogéneo, cuando los valores calculados de E_i se encuentran dentro de la elipse.

2.2.3.3. Prueba de Helmert

La prueba de Helmert es bastante sencilla y consiste en analizar el signo de las desviaciones de cada evento de la serie con respecto a la media. Si una desviación de un cierto signo es seguida de otra del mismo signo, se presenta una secuencia (S), si a una desviación de cierto signo le sigue otra de signo contrario se tiene un cambio (C).

Si la serie es homogénea, la diferencia entre el número de secuencias y el número de cambios en el registro debe ser cero, dentro de los límites de un error probable, el cual depende de la longitud del registro, de acuerdo a la siguiente relación.

$$-\sqrt{n-1} \leq S - C \leq \sqrt{n-1}$$

Donde:

S: Número de secuencias.

C: Número de cambios.

n: Número de observaciones.

2.2.3.4. Método de las alternancias.

El método consiste en determinar grupos de datos que sean superiores o inferiores a la media, se los señala con signos (-) si son inferiores a la media y (+) si son superiores, luego se determina el número de alternancias (U), una vez determinado este valor, se calculan los límites superior e inferior del rango al que debe pertenecer U para que la información sea homogénea, si el valor de U está fuera de estos límites entonces la serie no es homogénea o es muy variable.

Estos límites están tabulados o se pueden calcular mediante las fórmulas:

$$\text{Límite superior } LS = 1 + 1.1 \frac{n}{2}$$

$$\text{Límite inferior } LI = -1 + 0.9 \frac{n}{2}$$

2.2.4. GRILLA METEOROLÓGICA GMET

La Herramienta de Ensamble Meteorológico de Rejilla (GMET, por sus siglas en inglés) es un conjunto cuadrículado de precipitación y temperatura basado en observaciones, en un incremento diario durante un período determinado. Esto permite la estimación de la precipitación y la incertidumbre de la temperatura en el modelado hidrológico y la

asimilación de datos mediante el uso de la varianza del conjunto. La verificación estadística del conjunto indica que generalmente tiene buena confiabilidad y discriminación de eventos de diferentes magnitudes, pero tiene un ligero sesgo de humedad para eventos de umbral alto (> 50 mm). La media del conjunto es similar a otro conjunto de datos hidrometeorológicos ampliamente utilizados, pero con algunas diferencias importantes. El producto conjunto produce una ocurrencia más realista de estadísticas de precipitación (fracción de día húmedo), lo que afecta la derivación empírica de otros campos utilizados en el modelado hidrológico y de la superficie terrestre. Las tasas de lapso de elevación de la temperatura se derivan directamente de las observaciones, lo que resulta en temperaturas de invierno más altas en elevaciones altas en el oeste de EE. UU.

El desarrollo de Gridded Meteorological Ensemble Tool (GMET) permite la cuantificación de la incertidumbre para los conjuntos de datos de temperatura y precipitación cuadrículados basados en estaciones, específicamente. GMET completa los muchos pasos necesarios para construir conjuntos meteorológicos cuadrículados:

- Ingiere datos de estación.
- Garantiza que las estaciones se completen en serie (rechaza los registros demasiado cortos para ser validados y llena los vacíos de datos usando métodos bien establecidos).
- Interpola las series de tiempo de precipitación y temperatura completas en serie de los datos de la estación para estimar la probabilidad de ocurrencia de la precipitación y las distribuciones de probabilidad de las cantidades de precipitación.
- Genera campos espaciales cuadrículados utilizando campos aleatorios espaciados espacial y temporalmente para muestrear las distribuciones estimadas generadas en el paso anterior, lo que permite estimaciones constantes de incertidumbre Tanto en el espacio como en el tiempo. El enfoque de interpolación probabilística se detalla en Clark y Slater [2006].

GMET se ha utilizado para desarrollar un conjunto de datos de precipitación y temperatura diarios de primera clase en su clase a una resolución espacial de un octavo grado para los Estados Unidos contiguos desde 1980 hasta 2012. El conjunto de datos de 100 miembros,

descrito en Newman et al. [2015], se compara bien con los productos de datos establecidos, refleja incertidumbre de observación e interpolación.

Los conjuntos producidos por GMET se han utilizado en la asimilación de datos hidrológicos (por ejemplo, para evaluar el potencial de la asimilación de datos de nieve para mejorar la predicción del flujo de la corriente en todo el oeste de los EE. UU.) Y la evaluación del modelo de clima regional de alta resolución.

La metodología de GMET descrita en detalle en (Newman et al., 2015), reconoce la incertidumbre inherente de los productos de interpolación de precipitación y temperatura debido a las observaciones escasas, representatividad de las observaciones y errores de medición.

Reconociendo esta incertidumbre, el método propuesto produce un ensamble de precipitación y temperatura a paso de tiempo diario basado en las observaciones existentes. La generación de ensambles permite la estimación de una varianza para determinar la incertidumbre, y además durante la validación del producto se realizó una validación cruzada con datos de estaciones observadas. La verificación estadística del ensamble permite identificar su confiabilidad, aunque su promedio es similar a otros sets de datos existentes a nivel global. El producto de GMET es una representación más realista de la estadística de la precipitación por considerar parámetros espaciales invariantes en el tiempo y una red de estaciones de medición puntual, lo cual impacta en la derivación empírica de otros campos usados en modelación hidrológica. La habilidad para la modelación hidrológica de la grilla es corroborada durante la ejecución de los balances hídricos.

Los algoritmos usados en GMET son una extensión del trabajo realizado por Clark and Slater (2006) para desarrollar un ensamble diario de precipitación y temperatura basado en estaciones para USA. El propósito principal de este set de datos es ser usado como entrada para la modelación de superficie de suelo e hidrología y para estudios de asimilación de datos, pero también puede ser usado para la validación de modelos atmosféricos. La descripción del método se basa principalmente en la explicación del sistema de interpolación probabilística.

La interpolación probabilística en la que se basa GMET tiene dos pasos principales:

- La interpolación espacial.

- La generación del ensamble.

La interpolación espacial usa series de datos completa de precipitación y temperatura. Este set de datos cubre el territorio para el cual se quieren generar los resultados.

Para la generación del ensamble, se usan campos aleatorios correlacionados espacialmente (SCRFs por sus siglas en inglés – Spatially Correlated Random Fields). El método para generar los SCRf genera números progresivamente para cada punto de la grilla, condicionando los números aleatorios de acuerdo con los números generados previamente. El SCRf se genera para cada día de forma independiente para cada una de las variables: precipitación, temperatura media y rango diurno (diferencia entre temperatura máxima diaria y mínima).

Un resultado importante de uso de SCRf para generar los valores puntuales es que *permite que sucedan eventos más extremos que las observaciones en secciones de la grilla donde no hay observaciones*. Este es un punto clave con respecto a cualquier esquema basado en interpolaciones en que cada punto de la grilla es limitado por el valor máximo de las observaciones, lo cual no siempre es correcto y peor aún en terrenos con bastante influencia orográfica.

En total se desarrollan 30 miembros del ensamble mensual para el periodo 1980-2016 con su promedio y desviación estándar. El promedio del ensamble es el candidato ideal para utilizar como entrada a posteriores estudios y la desviación estándar, la representación de la incertidumbre en los datos de entrada de clima (precipitación y temperatura específicamente).

2.2.5. MODELOS DIGITALES DE ELEVACIÓN

Un modelo digital de elevación es una representación visual y matemática de los valores de altura con respecto al nivel medio del mar, que permite caracterizar las formas del relieve y los elementos u objetos presentes en el mismo.

Estos valores están contenidos en un archivo de tipo ráster con estructura regular, el cual se genera utilizando equipo de cómputo y software especializados.

En los modelos digitales de elevación existen dos cualidades esenciales que son la exactitud y la resolución horizontal o grado de detalle digital de representación en formato digital, las

cuales varían dependiendo del método que se emplea para generarlos y para el caso de los que son generados con tecnología LIDAR se obtienen modelos de alta resolución y gran exactitud (valores submétricos).

Los datos digitales más comunes de la forma de la superficie de la tierra son los modelos digitales de elevación (DEM) basados en celdas. Estos datos se utilizan como entrada para cuantificar las características de la superficie del suelo.

Un DEM es una representación de ráster de una superficie continua, que en general hace referencia a la superficie de la tierra. La precisión de estos datos se determina principalmente por la resolución (la distancia entre los puntos de muestra). Otros factores que afectan la precisión son el tipo de datos (entero o punto flotante) y el muestreo real de la superficie cuando se crea el DEM original.

Los Modelos Digitales de Elevación son ampliamente utilizados en aplicaciones relacionadas con el uso y manejo de recursos naturales, de las cuales pueden distinguirse grandes categorías de aplicaciones que utilizan los modelos como son geodesia y fotogrametría, ingeniería civil, planeación y manejo de recursos naturales, ciencias de la tierra, en aplicaciones militares, cartografía especializada, prevención y atención a desastres naturales, entre otras:

- Geodesia y fotogrametría. - En éstos campos, el propósito principal es el de producir modelos de alta calidad para otras aplicaciones como ingeniería civil, y cartografía, además de emplearse en la captura de datos fotogramétricos, determinación del geoide, métodos de control de calidad, como fuente de comprobación de mediciones del terreno, edición de datos del terreno, producción de ortofotografías, cartografía topográfica, etc.
- Ingeniería civil. - Los Modelos Digitales de Elevación pueden ser usados, en ingeniería civil, en aplicaciones tales como en el diseño para la construcción de infraestructura diversa, minas a cielo abierto, el cálculo de perfiles (secciones de perfil) y los cálculos de volúmenes (llamado también de “corte y relleno”). Son usados en aplicaciones tales como diseño de carreteras, presas y otro tipo de infraestructura.

- Manejo y planeación de recursos naturales. - Este campo, el de mayor uso de los Modelos Digitales de Elevación, involucra disciplinas tales como planeación urbana y ambiental, teledetección, ciencias del suelo, agricultura, meteorología y climatología. Aplicaciones típicas serían, entre otras: estudios de impacto ambiental, prevención de desastres, localización de sitios industriales, corrección geométrica y auxiliar en la clasificación de imágenes de satélite, estudios de rentabilidad, desarrollo de estrategias de cosecha, modelos de flujo de viento y dispersión de contaminantes.
- Ciencias de la tierra. - Las aplicaciones en Ciencias de la Tierra (geología, hidrología, geomorfología, glaciología) requieren de funciones específicas para el modelaje de las discontinuidades del terreno, principalmente redes de drenaje, de las que se requiere una representación muy precisa de ellos. Algunas aplicaciones serían: monitoreo de cuencas de drenaje para monitoreo de inundaciones y control de contaminantes, modelamiento de flujos hidrológicos, simulaciones para la creación de cuencas hidrológicas, interpretación y cartografía geológica.

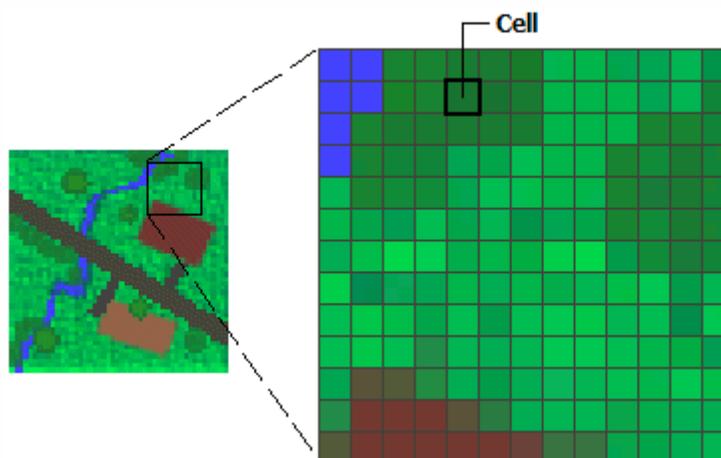
2.2.6. DATOS RÁSTER

En su forma más simple, un ráster consta de una matriz de celdas (o píxeles) organizadas en filas y columnas (o una cuadrícula) en la que cada celda contiene un valor que representa información, como la temperatura. Los rásteres son fotografías aéreas digitales, imágenes de satélite, imágenes digitales o incluso mapas escaneados.

Los datos almacenados en formato ráster representan fenómenos del mundo real:

- Los datos temáticos (también conocidos como discretos) representan entidades como datos de la tierra o de uso de la tierra.
- Los datos continuos representan fenómenos como la temperatura, la elevación o datos espectrales, entre ellos imágenes satelitales y fotografías aéreas.

Figura N°9 Datos Ráster



- Las imágenes incluyen mapas escaneados o dibujos y fotografías de edificios. Los rásteres temáticos y continuos se pueden visualizar en el mapa en forma de capas de datos junto con otros datos geográficos, pero a menudo se utilizan como datos de origen para el análisis espacial con la extensión de ArcGIS Spatial Analyst extension. Los rásteres de imágenes suelen utilizarse como atributos en tablas: pueden visualizarse con datos geográficos y se utilizan para transmitir información adicional acerca de las entidades geográficas de mapas.

Si bien la estructura de datos ráster es simple, es excepcionalmente útil para una amplia variedad de aplicaciones. En un SIG, los usos de los datos ráster se pueden dividir en cuatro categorías principales:

- Rásteres en forma de mapas base. - Un uso común de los datos ráster en un SIG es en forma de visualización de fondo para otras capas de entidades. Por ejemplo, las ortofotografías que se visualizan debajo de otras capas ofrecen al usuario de mapas la garantía de que las capas de mapa se alinean espacialmente y representan tanto objetos reales como información adicional. Las tres fuentes principales de mapas base ráster son las ortofotografías de fotografías aéreas, imágenes de satélite y mapas escaneados.
- Rásteres en forma de mapas de superficie. - Los rásteres son apropiados para representar datos que cambian continuamente en un entorno (superficie). Ofrecen un método efectivo para almacenar la continuidad en forma de superficie. También

proporcionan una representación de superficies con espacios regulares. Los valores de elevación que se miden desde la superficie de la Tierra son la aplicación más común de los mapas de superficie, pero otros valores, como las precipitaciones, la temperatura, la concentración y la densidad de población, también pueden definir superficies que se pueden analizar espacialmente. En el siguiente ráster se visualiza la elevación: se utiliza el color verde para mostrar una elevación menor y celdas de color rojo, rosa y blanco para mostrar elevaciones mayores.

- Rásteres en forma de mapas temáticos. - Los rásteres que representan datos temáticos se pueden derivar al analizar otros datos. Una aplicación de análisis común consiste en clasificar una imagen de satélite por categorías de cobertura de suelo. Básicamente, esta actividad agrupa los valores de datos multiespectrales en clases (como tipo de vegetación) y asigna un valor categórico. También es posible obtener mapas temáticos a partir de operaciones de geoprocésamiento que combinen datos de varias fuentes como, por ejemplo, datos vectoriales, ráster y de terreno. Por ejemplo, puede procesar datos por medio de un modelo de geoprocésamiento para crear un dataset ráster apropiado para una actividad específica.
- Rásteres en forma de atributos de una entidad. - Los rásteres utilizados como atributos de una entidad pueden ser fotografías digitales, documentos escaneados o dibujos escaneados relacionados con un objeto o ubicación geográfica. Una capa de parcela podría tener documentos legales escaneados que identifiquen la transacción más reciente de dicha parcela, o una capa que represente las entradas a una cueva que podría incluir imágenes de las entradas reales a las cuevas asociadas a las entidades de puntos.

Las ventajas de almacenar los datos en forma de ráster son las siguientes:

- Estructura de datos simple: matriz de celdas con valores que representan una coordenada y que, en ocasiones, se encuentra vinculada a una tabla de atributos.
- Formato potente para análisis espacial y estadístico avanzado.
- Capacidad de representar superficies continuas y llevar a cabo análisis de superficie.
- Capacidad de almacenar puntos, líneas, polígonos y superficies de manera uniforme.
- Capacidad de llevar a cabo superposiciones rápidas con datasets complejos.

Existen otras consideraciones para almacenar los datos en forma de ráster que podrían convencerle para que utilice una opción de almacenamiento basada en vectores. Por ejemplo:

- Podrían producirse imprecisiones espaciales debido a los límites impuestos por las dimensiones de celda del dataset ráster.
- Los datasets ráster son potencialmente de gran tamaño. La resolución aumenta a medida que disminuye el tamaño de la celda. Sin embargo, el coste también aumenta en el espacio en disco y en las velocidades de procesamiento. Para un área determinada, el cambio de celdas a la mitad del tamaño actual requerirá cuatro veces más espacio de almacenamiento, dependiendo del tipo de datos y las técnicas de almacenamiento utilizadas.

CAPÍTULO III

METODOLOGÍA Y DESARROLLO

3.1. Recopilación de datos meteorológicos

En el Departamento de Tarija se cuenta con un total de 56 estaciones meteorológicas en funcionamiento de las cuales 22 (Tabla N°3) están dentro de la Cuenca del Río Guadalquivir o en zonas cercanas, podemos observar su ubicación geográfica en la Figura N°10 (Ubicación Geográfica Estaciones Activas).

De la totalidad de estaciones disponibles en la cuenca, se observa que su registro es discontinuo y el periodo de medición es variable, ya que se presentan estaciones con periodos de mediciones mayores a 30 años; sin embargo, otras cuentan con menos de 10 años de registro como se puede apreciar en las Tablas N°3.

Tabla N° 3 Estaciones Activas

N°	Estación	Provincia	Latitud	Longitud	Elevación msnm	Periodo histórico	Record (años)	Estado Actual
1	Tarija Aeropuerto	Cercado	-21.55	-64.71	1875	1944 - 2015	71	Activa
2	El Tejar Tarija	Cercado	-21.54	-64.72	1859	1970 - 2015	41	Activa
3	San Jacinto Sud	Cercado	-21.61	-64.72	1820	1975 - 2015	40	Activa
4	San Andrés	Cercado	-21.62	-64.82	1987	1975 - 2015	40	Activa
5	Juntas	Avilés	-21.81	-64.8	1900	1975 - 2015	39	Activa
6	Cañas	Arce	-21.9	-64.85	2030	1977 - 2015	37	Activa
7	Tucumillas	MéndeZ	-21.46	-64.83	2560	1978 - 2015	37	Activa
8	Yesera Norte	Cercado	-21.37	-64.55	2200	1977 - 2015	36	Activa
9	Coimata	MéndeZ	-21.5	-64.79	2027	1980 - 2015	36	Activa
10	Sella Quebrada	MéndeZ	-21.38	-64.7	2080	1980 - 2015	34	Activa
11	Tomayapo Pueblo	MéndeZ	-21.27	-65.05	2720	1981 - 2015	34	Activa
12	Trancas	MéndeZ	-21.31	-64.82	2200	1984 - 2015	29	Activa
13	Ciudad-Tja	Cercado	-21.54	-64.73	1864	1985 - 2014	28	Activa
14	Ce.Na.Vit.	Avilés	-21.73	-64.66	1730	1989 - 2015	25	Activa
15	Campanario	MéndeZ	-21.51	-64.98	3460	1990 - 2015	25	Activa
16	Alto Cajas	MéndeZ	-21.3	-64.47	2980	1977 - 2015	24	Activa
17	Pinos Sud	Cercado	-21.76	-64.88	2100	1975 - 2015	17	Activa

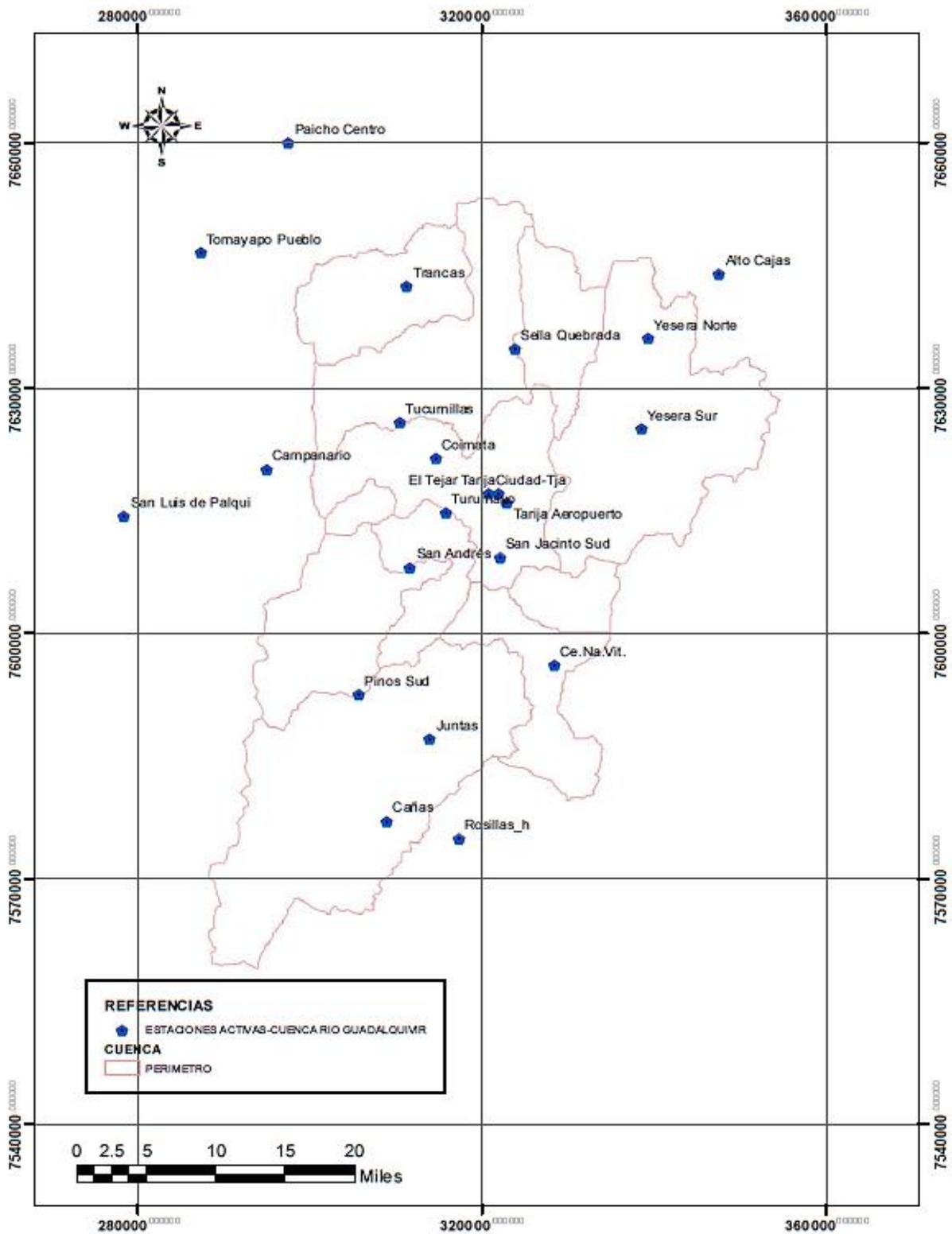
18	Turumayo	Cercado	-21.56	-64.78	1978	1999 - 2015	17	Activa
19	Yesera Sur	Cercado	-21.47	-64.56	2092	2002 - 2015	13	Activa
20	Paicho Centro	Méndez	-21.15	-64.95	2570	2005 - 2015	8	Activa
21	San Luis de Palqui	Avilés	-21.56	-65.14	3321	2005 - 2015	7	Activa
22	Rosillas_h	Arce	-21.92	-64.77	2050	1989 - 2015	5	Activa

Fuente: Elaboración Propia

En el Departamento se cuenta también con 87 estaciones cerradas de las cuales 55 están ubicadas en la cuenca y zonas cercanas a la misma, hecho que se puede observar en la Figura N°11 (Ubicación Geográfica Estaciones Cerradas). Como se muestra en la Tabla N°4 algunas presentan un periodo de registro de mediciones mayor a los 20 años pese a que actualmente no se encuentran en funcionamiento. Por lo tanto, se realizará el correspondiente análisis para la incorporación y generación de la grilla meteorológica GMET.

Esta información de las estaciones presentes en la cuenca y sus alrededores que se analizarán en el presente estudio, además de los datos tanto de precipitación, como de temperatura diaria que serán analizados, fueron obtenidos del Servicio Nacional de Meteorología e Hidrología de Bolivia (SENAMHI).

Figura N°10 Ubicación Geográfica Estaciones Activas



Fuente: Elaboración Propia

Tabla N° 4 Estaciones Cerradas

N°	Estación	Provincia	Latitud	Longitud	Elevación	Periodo histórico	Record (Años)	Estado Actual
1	León Cancha	Méndez	-21.18	-64.72	2600	1976 - 2011	35	Inactiva
2	Chocloca	Avilés	-21.75	-64.73	1795	1975 - 2010	32	Inactiva
3	Canasmoro	Méndez	-21.35	-64.75	2080	1944 - 2003	32	Inactiva
4	Calderillas	Avilés	-21.75	-64.95	2800	1975 - 2003	28	Inactiva
5	Pampa Redonda_H	Cercado	-21.7	-64.82	1950	1966 - 1994	26	Inactiva
6	Junacas	Cercado	-21.43	-64.47	2300	1977 - 2003	25	Inactiva
7	Tolomosa Grande	Cercado	-21.62	-64.77	1900	1968 - 1993	25	Inactiva
8	Padcaya	Arce	-21.88	-64.72	2010	1943 - 1992	24	Inactiva
9	Gamoneda	Cercado	-21.48	-64.63	1150	1979 - 2003	24	Inactiva
10	Rejara	Arce	-22.02	-64.98	3000	1980 - 2003	23	Inactiva
11	Ladera Centro	Avilés	-21.65	-64.53	2080	1979 - 2003	22	Inactiva
12	El Molino Tomayapo	Méndez	-21.37	-64.95	3410	1981 - 2003	21	Inactiva
13	Iscaiyachi - San Antonio	Méndez	-21.39	-64.95	3450	1975 - 1994	19	Inactiva
14	Sama Cumbre	Méndez	-21.49	-64.98	3820	1980 - 1999	19	Inactiva
15	Canchasmayu	Arce	-21.88	-64.88	1120	1975 - 1993	18	Inactiva
16	San Nicolás	Avilés	-21.72	-64.68	1800	1986 - 2004	18	Inactiva
17	Calamuchita	Avilés	-21.7	-64.63	1680	1975 - 1994	18	Inactiva
18	San Lorenzo	Cercado	-21.42	-64.75	1900	1975 - 1993	18	Inactiva
19	Copacabana Taxara	Avilés	-21.8	-65.1	3800	1970 - 1993	17	Inactiva
20	Rincón Cañas H	Arce	-21.97	-64.9	2280	1977 - 1993	16	Inactiva
21	Alizos	Avilés	-21.82	-64.87	2040	1977 - 1993	15	Inactiva
22	Colon Sud	Avilés	-21.75	-64.65	2000	1977 - 1993	15	Inactiva
23	San Pedro Buena Vista_tja	Cercado	-21.43	-64.67	2195	1979 - 2001	15	Inactiva
24	Sama Iscaiyachi	Méndez	-21.47	-64.95	3440	1979 - 1993	14	Inactiva
25	Concepcion_tj	Avilés	-21.7	-64.62	1710	1975 - 1987	13	Inactiva
26	San Agustin_tja	Cercado	-21.53	-64.52	2210	1979 - 1992	13	Inactiva
27	Tarija_SENAMHI_h	Cercado	-21.55	-64.73	1875	1974 - 1985	12	Inactiva
28	Tomatitas	Méndez	-21.49	-64.76	1928	1999 - 2013	12	Inactiva
29	Santa Ana Puente	Cercado	-21.52	-64.57	1950	1977 - 2000	11	Inactiva
30	Tarija Cancha	Méndez	-21.4	-64.77	2050	1979 - 1991	11	Inactiva
31	Tomatas Grande	Méndez	-21.3	-64.8	2150	1976 - 1986	10	Inactiva
32	La Angostura_C_TJA	Avilés	-21.7	-64.6	1690	1975 - 1984	9	Inactiva
33	Embalse San Jacinto	Cercado	-21.6	-64.7	1800	1990 - 1999	9	Inactiva

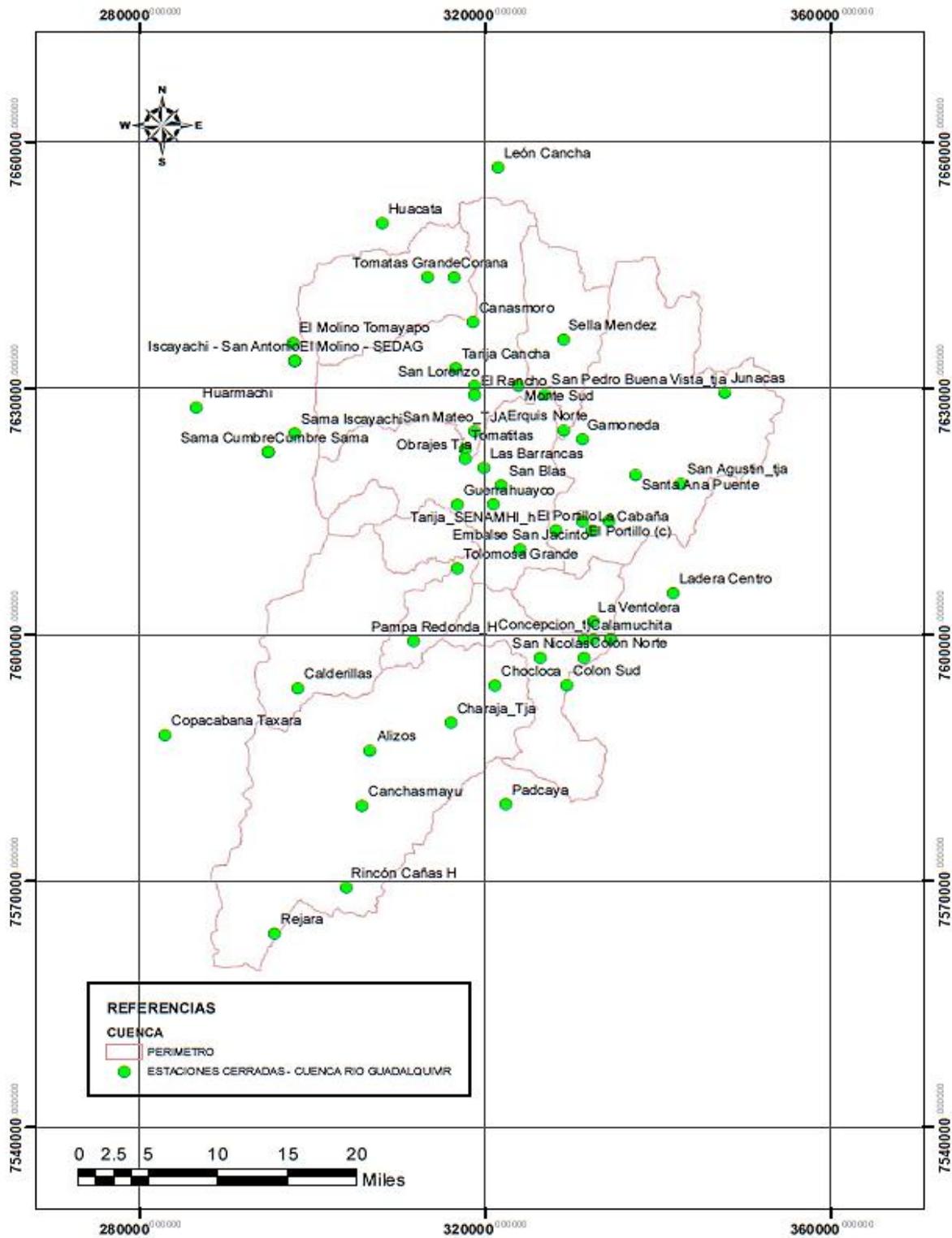
34	San Mateo_TJA	Méndez	-21.47	-64.75	1900	1977 - 1986	9	Inactiva
35	Monte Sud	Cercado	-21.42	-64.7	2005	1979 - 1987	8	Inactiva
36	Colon Norte	Avilés	-21.72	-64.63	1800	1987 - 1994	7	Inactiva
37	Santa Ana Kolhberg	Cercado	-21.57	-64.6	1870	1976 - 1985	7	Inactiva
38	La Cabaña	Cercado	-21.58	-64.62	1870	1978 - 1986	7	Inactiva
39	Obrajes Tja	Cercado	-21.5	-64.76	1890	1977 - 1985	7	Inactiva
40	Sella Méndez	Méndez	-21.37	-64.65	2060	1977 - 1984	7	Inactiva
41	Corana	Méndez	-21.3	-64.77	2120	1986 - 1994	7	Inactiva
42	Huacata	Méndez	-21.24	-64.85	2858	2001 - 2008	7	Inactiva
43	Cumbre Sama	Méndez	-21.49	-64.98	3820	1981 - 1987	7	Inactiva
44	La Ventolera	Avilés	-21.68	-64.62	1710	1987 - 1993	6	Inactiva
45	San Blas	Cercado	-21.53	-64.72	1875	1989 - 1994	5	Inactiva
46	Guerrahuayco	Cercado	-21.55	-64.77	1980	1990 - 1994	4	Inactiva
47	El Portillo (c)	Cercado	-21.57	-64.63	1825	1988 - 1992	4	Inactiva
48	Erquis Norte	Méndez	-21.47	-64.65	2062	1981 - 1985	4	Inactiva
49	El Portillo	Cercado	-21.58	-64.66	1844	2006 - 2010	3	Inactiva
50	El Molino – SEDAG	Méndez	-21.39	-64.95	3387	2006 - 2009	3	Inactiva
51	Las Barrancas	Cercado	-21.51	-64.74	1924	2006 - 2009	2	Inactiva
52	El Rancho	Méndez	-21.43	-64.75	1988	2007 - 2009	2	Inactiva
53	Huarmachi	Méndez	-21.44	-65.06	3382	2006 - 2008	2	Inactiva
54	La Angostura_Tja	Avilés	-21.7	-64.6	2700	2007 - 2007	0.5	Inactiva
55	Charaja_Tja	Avilés	-21.79	-64.78	1884	2008 - 2009	0.25	Inactiva

Fuente: Elaboración Propia

3.1.1 Periodo de Registro

El acopio de datos meteorológicos para que estos puedan ser útiles para futuras investigaciones, requiere de gran constancia para lograr series continuas y rigor en la medición, ya que las faltas en esa constancia y rigor, junto con algunos inconvenientes como ser cambios eventuales de observador o de aparato de registro, son inevitables cuando se estudian periodos amplios que provocan la obtención de periodos de registro variados y discontinuos como es el caso de las estaciones presentes en la Cuenca del Rio Guadalquivir.

Figura N°11 Ubicación Geográfica Estaciones Cerradas



Fuente: Elaboración Propia

Al recopilar la información de los datos históricos de la precipitación, temperaturas máximas y mínimas a nivel diario de las estaciones tanto activas como inactivas presentes, en la Cuenca del Río Guadalquivir y sus alrededores, se pudo detallar el registro de las mismas. En la Tabla N°5 se muestra el detalle del periodo de registro para la variable precipitación. La Tabla N°6 muestra el periodo de registro para la temperatura máxima diaria y la Tabla N°7 muestra el periodo de registro para la temperatura mínima diaria para estaciones en funcionamiento.

Además, de igual manera se detalla en las Tablas N°8, 9 y 10 los periodos en los que se registró información de precipitación, temperatura máxima y temperatura mínima a nivel diario de las estaciones inactivas en la cuenca en estudio.

En la elaboración de la grilla en el balance hidrológico superficial de Bolivia de 2017, se utilizaron varios criterios de selección de estaciones no solo estadístico sino también de cantidad de años de registro de manera continua. Esto, derivó en una menor cantidad de estaciones utilizadas ya que no cumplían con las exigencias establecidas por esta herramienta para poder ser empleadas, como ser: el contar con un lapso de registro amplio para ser validado y contar con más del 80% de completitud de datos en sus registros en el periodo que se pretende analizar.

3.2.Control de Calidad

Para desarrollar nuevos productos y servicios climáticos es indispensable contar con una base de datos robusta y confiable, en la que todos sus registros sean de buena calidad. La consistencia de los datos es indispensable para que estos nuevos productos o servicios sean confiables y puedan proveer soluciones a los usuarios.

El uso de variables climáticas para la generación de información climática útil y relevante requiere que los datos hayan sido sometidos a un proceso de control de calidad. Este proceso debe identificar valores sospechosos que podrían ser incorrectos y, en consecuencia, afectar indebidamente los productos o estadísticas derivados a partir de los datos originales. Por esta razón se realizó un análisis de control de calidad a los datos brindados por las estaciones que se emplearon en este estudio. Éste análisis se realizó con el apoyo de herramientas que tienen esta finalidad como son el RCLimDex y DECADE. En los apartados 3.2.1 y 3.2.2. se muestra el análisis realizado por ambos programas para una estación de las 52 que se analizaron.

3.2.1. Análisis de control de calidad: RCLIMDEX

RClindex es una rutina que corre sobre la plataforma de análisis estadístico R que se enfoca en el control de calidad. Para su ejecución, es necesario verificar que dentro de la plataforma R se tenga instalado y activo el paquete tkrplot. Una vez hecho esto, se carga la interfaz gráfica de RClindex; dentro de ella, la primera opción que aparece es la necesaria para cargar los datos y correr el análisis de calidad de los mismos.

Al activarla, se abre el explorador de navegación para elegir el archivo que contiene los datos por analizar. Los datos deben estar en el formato RClindex, el cual se encuentra explicado en el punto 2.2.2.2.1. Control de Calidad de RClindex. Si los datos contenidos en el archivo se encuentran en el formato correcto, el sistema envía un mensaje indicando que se han cargado de forma correcta y señalando que se continuará con el análisis de calidad. Como resultado de la ejecución de la rutina se crea una carpeta denominada “salidas” en la que se encuentran los resultados del análisis realizado.

Se muestran los resultados que se obtuvo con RClindex para la estación Campanario que es una de las estaciones que fueron sometidas al análisis de control de calidad, los resultados son los siguientes:

1. Campanario Boxes. - La figura N°12 muestra gráficas box plots arrojadas por RClindex para realizar el control de calidad, en la que contrastan los valores registrados para los doce meses del año. Lo que nos muestra este resultado son diagramas de cajas que es otra manera de representar una distribución de los valores donde la línea central que se muestra en la caja representa la mediana de los valores para el mes respectivo, y los valores extremos también llamados atípicos se muestran fuera de los extremos de la caja como es el caso de valores extremos en precipitación y temperatura diaria máxima que se pueden apreciar en los incisos a) y b) de la figura.

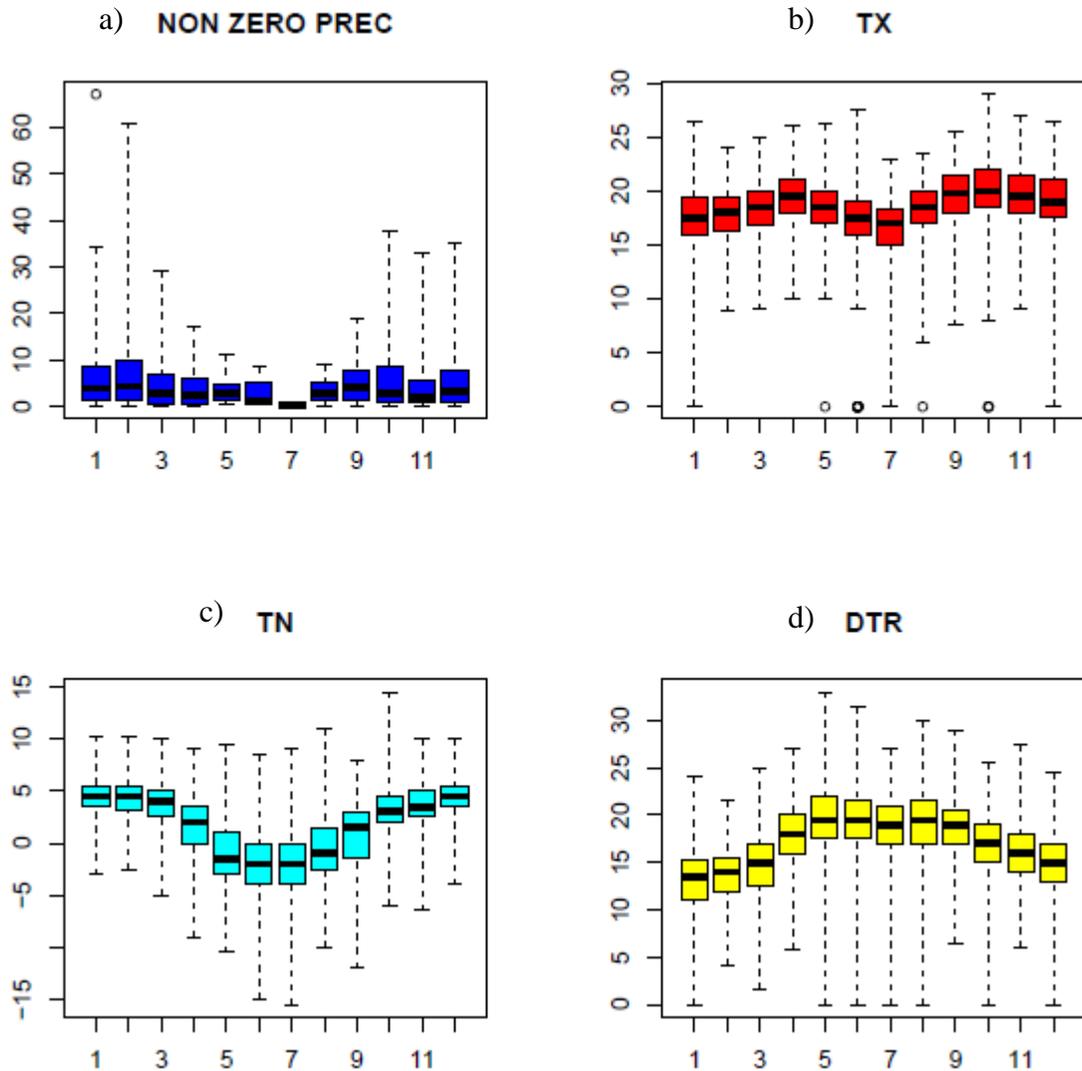
El análisis minucioso de cada una de las gráficas de la Figura N°12 sería el siguiente:

- a) NON ZERO PREC.- Las precipitaciones comprendidas entre el 25% y 50% del registro son menos dispersas que entre el 50% y el 75%; el bigote inferior (X_{min}, P) es más corto que el superior, por ello el 25% de las precipitaciones con valores menores a la mediana están más concentrados que el 25% de los mayores que están más dispersos, existen valores de hasta 60mm. correspondientes al mes de febrero que podrían deberse a eventos extremos que no son de dudosa veracidad por encontrarse en la época lluviosa. Existe un outlier mayor a 60mm. en el mes de enero que puede tratarse de un evento extremo ya que la estación más cercana como es Tucumillas registra datos similares.
- b) TX. - En estos diagramas de caja podemos ver como la mediana de los datos de temperatura están entre un rango de 15 a 20 °C y que la mayoría de los registros de temperatura máxima están entre 15 y 23°C, pero en el bigote superior vemos que hay datos de temperatura que llegan a los 29 °C y en el bigote inferior se muestra que hay registros de hasta 0°C los cuales, para esta estación, al igual que los outliers, en realidad nos están mostrando la inexistencia de datos que son representados por -99.9.
- c) TN. - En estos diagramas de caja podemos ver como la mediana de los datos de temperatura mínima están entre un rango de -3 a 4 °C y que la mayoría de los registros están entre -5 y 6°C, pero en el bigote superior se observa que hay datos

que llegan a 14°C y en el bigote inferior podemos ver que se presentan registros de -16°C.

- d) DTR. - De la inspección del diagrama de cajas de DTR podemos observar que existen oscilaciones diarias relativamente normales.

Figura N°12 Campanario Boxes



a) Precipitación, b) Tmáx, c) Tmín, d) Tmáx - Tmín

Fuente: Elaboración Propia

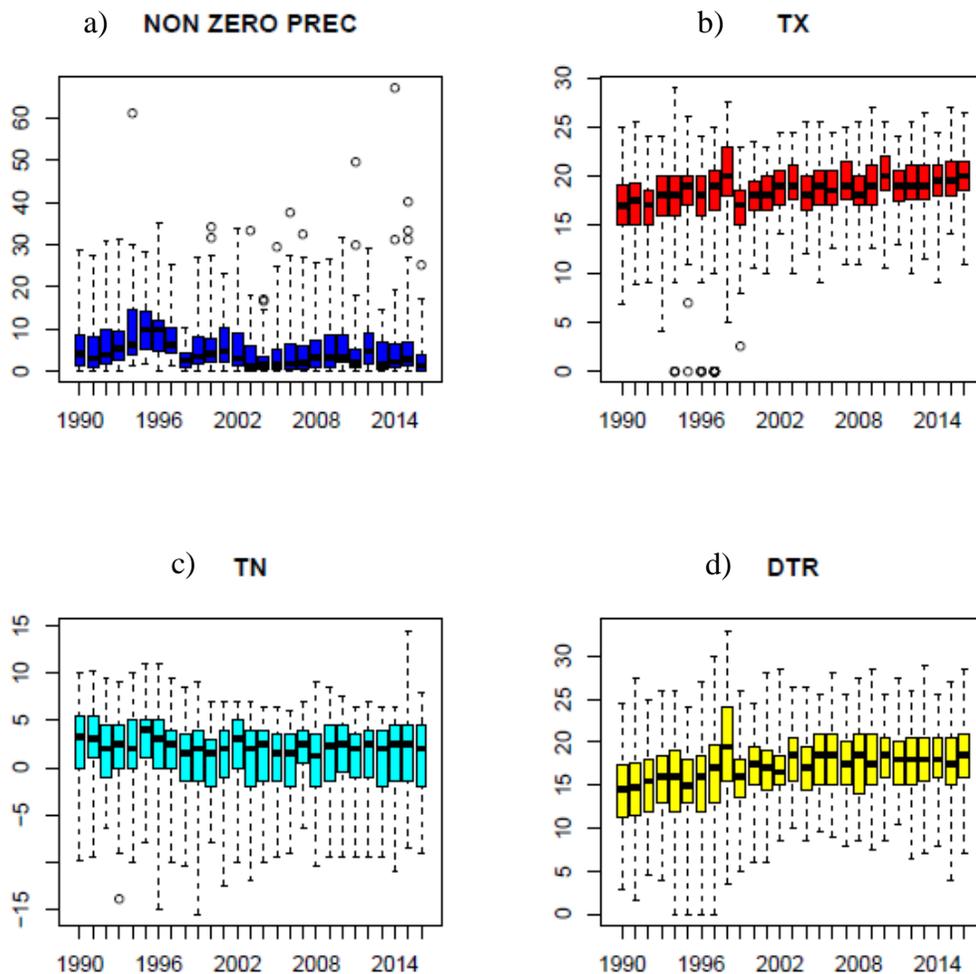
2. Campanario Boxseries. - En la Figura N°13 podemos encontrar diagramas de cajas para las mismas variables descritas en el punto anterior, pero con la diferencia que éstas no son para series mensuales sino para series anuales.

El análisis minucioso de cada una de las gráficas de la Figura N°13 sería el siguiente:

- a) NON ZERO PREC. – En los diagramas de caja para esta variable podemos ver que los bigotes inferiores son considerablemente más cortos que los bigotes superiores, de lo que podemos interpretar que el 25% de los valores mínimos están más concentrados que el 25% de los valores máximos.

La gráfica de precipitación nos advierte también la presencia de 15 outliers, los cuales se tratan de eventos extremos y los localizan en sus respectivos años de ocurrencia.

Figura N°13 Campanario Boxseries



a)Precipitación, b) Tmáx, c) Tmín, d) Tmáx - Tmín

Fuente: Elaboración Propia

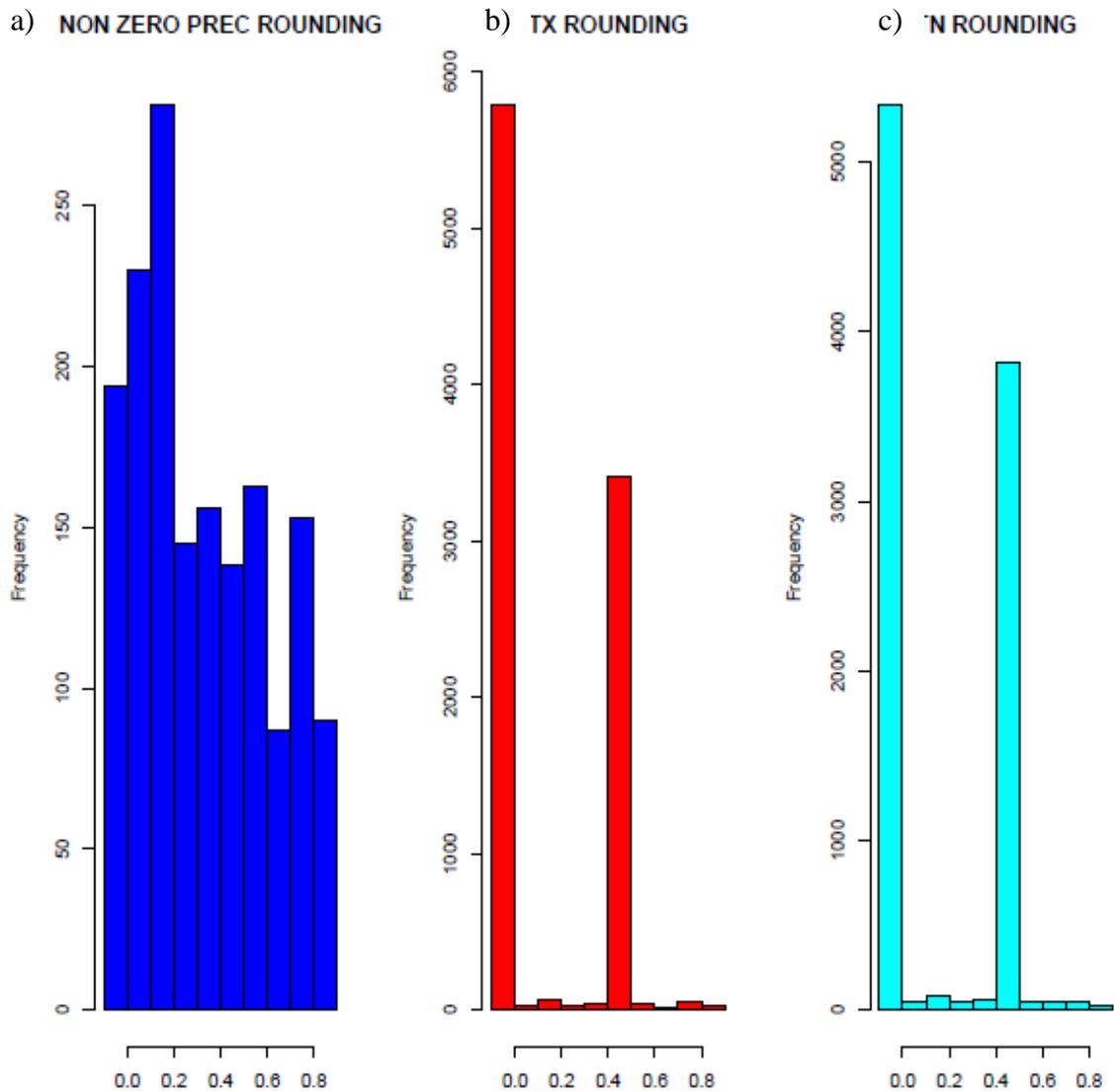
- b) TX. – En los diagramas de caja para esta variable podemos ver como la mediana de los datos de temperatura para la serie anual está en un rango de 16 a 20°C y que el 50% de los registros están comprendidos entre 15 y 23°C, podemos observar la existencia de 6 valores atípicos que son inferiores por lo que al revisar la inspección se ve que se tratan de la falta de datos en el año correspondiente y son representados por -99.9.
 - c) TN. - En los diagramas de caja para esta variable podemos ver como la mediana de los datos de temperatura mínima para la serie anual está en un rango de 2 a 4°C y que el 50% de los registros están comprendidos entre -1 y 6°C, podemos observar la existencia de un valor atípico el cual fue inspeccionado.
3. Campanario rounding. - La figura N°14 nos muestra diagramas de frecuencia de los 10 valores del punto decimal (i.e., x.0, x.1, x.2, ... x.9), nos permite comprobar si las series están redondeadas.

El análisis minucioso de cada una de las gráficas de la Figura N°14 es el siguiente:

- a) NON ZERO PREC ROUNDING. – El gráfico del redondeo nos muestra que en la precipitación la mayor parte de las observaciones se distribuyen con prevalencia en decimales.
 - b) TX ROUNDING. - Para el caso de la temperatura máxima en la gráfica se muestra que existe prevalencia de medición al entero y al redondeo de 0.5 de forma exagerada.
 - c) TN ROUNDING. – En datos de temperatura mínima se observa que al igual que para temperatura máxima existe prevalencia de mediciones redondeadas al entero y 0.5 de forma exagerada.
4. Campanario Duplicates. - Es un archivo de texto cuya principal finalidad es mostrar la posible repetición de fechas, es decir este archivo muestra la fecha duplicada de registros.

Para el caso de la estación Campanario no se encontraron datos duplicados por lo que se presenta un documento en blanco.

Figura N°14 Campanario Rounding



Fuente: Elaboración Propia

5. Campanario Outliers. - Este archivo de texto muestra los valores observados determinados como outliers (fuera de rango) y nos muestra la fecha del registro de estos valores atípicos o eventos extremos.

En la Figura N°15 en pc up nos etiqueta el dato de precipitación para el 24 de enero de 2014 haciéndonos notar que tiene un valor alto que debe ser revisado para analizar si se trata de un valor erróneo o un evento extremo.

De igual manera en el inciso b) tenemos etiquetados datos de temperatura máxima que tienen valores bajos que deben ser revisados, ya que al tratarse de temperaturas máximas no se espera valores de 0°C.

Figura N°15 Campanario Outliers

a) pc up
2014 1 24 67.1 21.5 4.5 17
b) tx low
1995 5 2 0 0 0 0
1997 6 1 NA 0 0 0
1997 10 1 0 0 -0.5 0.5

Fuente: Elaboración Propia

6. Campanario_tmaxmin: Este archivo de texto principalmente advierte fechas en las que los valores de la temperatura máxima son inferiores o iguales a la mínima y nos lo etiqueta.

Para el caso de la Estación Campanario no se encontró registros en los que la temperatura máxima sea mayor que la temperatura mínima por lo que se nos presenta un documento en blanco.

7. Campanario_tn_flatline. – Este archivo de texto está orientado hacia los valores de temperatura mínima que participan en rachas de al menos tres valores consecutivos idénticos (es decir que aparecen cuatro veces consecutivamente en la base de datos).

En la Figura N°16 tenemos el caso de cinco fechas en las que existen rachas largas de valores iguales en la temperatura mínima, podemos ver que en fechas consecutivas al 12 de julio de 1992 hay cuatro registros más que tienen un valor de -0.5 °C.

Figura N°16 Campanario_tn_flatline

1992	7	12	-0.5	5
1994	8	19	0	3
1998	3	8	3	3
2002	9	12	2.5	3
2008	1	29	5	3

Fuente: Elaboración Propia

8. Campanario_tx_flatline. - Este archivo de texto se realiza en forma simultánea al de temperatura mínima y está orientado de igual manera a indicar valores que se repitan más de cuatro veces consecutivas.

En la Figura N°17 tenemos el caso de cuatro fechas en las que existen rachas largas de valores iguales en la temperatura máxima, podemos ver que en fechas consecutivas al 04 de septiembre de 1991 hay tres registros más que tienen un valor de 19 °C.

Figura N°17 Campanario_tx_flatline

1991	9	4	19	4
1993	9	23	21	3
1993	9	30	20	3
1995	5	6	16	3

Fuente: Elaboración Propia

9. Campanario_tn_jumps: Este archivo de texto nos etiqueta las fechas en las que dos días consecutivos la diferencia de sus registros de temperatura mínima es superior a 20°C. Para el caso de la estación Campanario no se encontró registros de días consecutivos en los que la diferencia interdiaria sea mayor a 20°C.
10. Campanario_tx_jumps: Este archivo nos etiqueta al igual que el anterior, fechas en las que la diferencia interdiaria para temperatura máxima entre dos días consecutivos es mayor a 20°C.

Para la estación Campanario se puede ver en la Figura N°18 cuatro fechas en las que la diferencia interdiaria entre la fecha que se muestra y el día consecutivo a la misma fue mayor a 20°C.

Figura N°18 Campanario_tx_jumps

1994 12 9 20
1995 5 1 22
1997 6 9 22
1997 9 30 21

Fuente: Elaboración Propia

11. Campanario_toolarge. -Este archivo de texto está orientado a etiquetarnos valores considerados excesivamente elevados (valor absoluto > 50 para tx y tn y 200 mm para precipitación).

Para el caso de la estación Campanario no se presentan este tipo de registros, por lo que el documento de salida se encuentra en blanco.

3.2.2. Análisis de control de calidad: DECADE

Se aplicó esta herramienta que nos brindará resultados para las tres variables analizadas y nos permitirá realizar el control de calidad visual de las series diarias de precipitación (PRCP), temperatura máxima (TX) y temperatura mínima (TN).

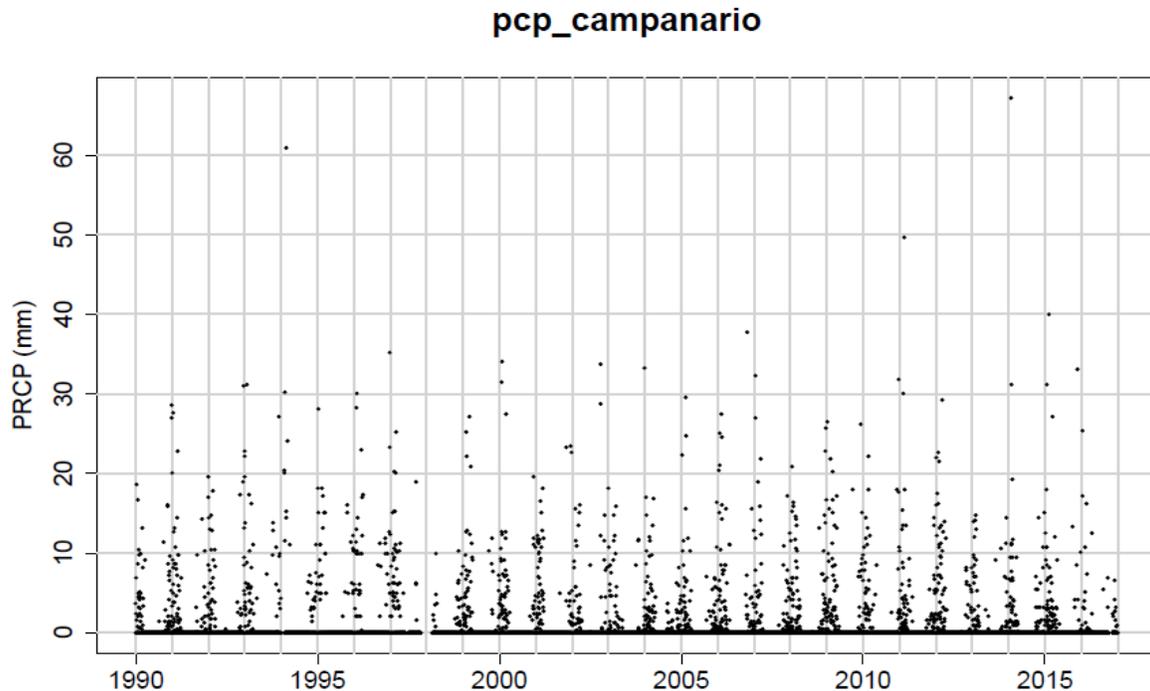
Los resultados que se obtuvieron para las tres variables que se estudiaron posteriormente son los siguientes:

1. Pointsplots
2. Decimals
3. Missing
4. Weekly Cycles
5. Statisticsç

A continuación, se muestran los resultados generados para la variable precipitación de la estación Campanario y se realiza el respectivo análisis de los mismos.

1. Pointsplots. - Es la gráfica de los puntos de las series temporales (Figura N°19), se puede ver cada dato y año registrado representado con un punto en la gráfica. En el mismo se puede observar la concentración de los registros en valores que van de 0 a 20 mm, pero se observa también la presencia de datos de dudosa veracidad o posibles eventos extremos en sus respectivos años de ocurrencia.

Figura N°19 Pointsplots Campanario

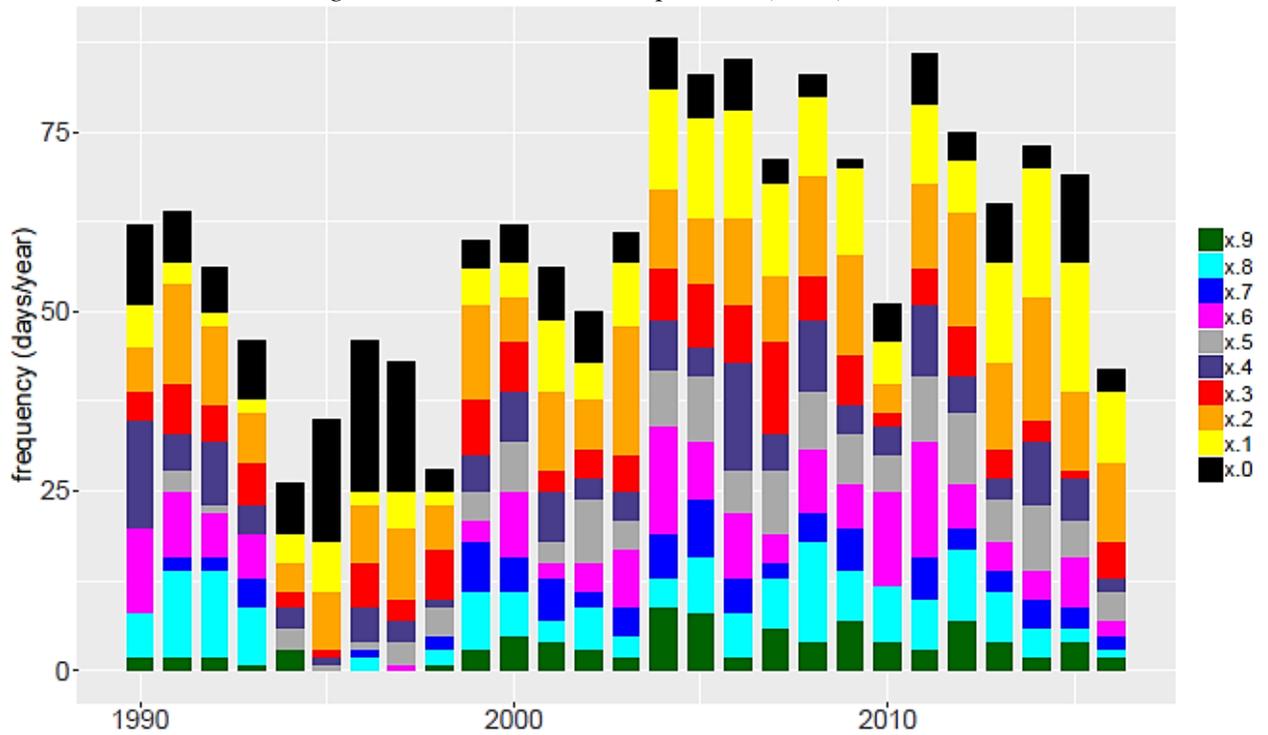


Fuente: Elaboración Propia

2. Decimals. - Es un histograma de decimales de la serie de datos para las tres variables analizadas.

En la Figura N°20 se observa en diferentes colores los decimales de los datos de precipitación, estos nos permitirán establecer la frecuencia con la que aparecen los mismos. Se puede observar que los registros de precipitación de la estación Campanario presentan una distribución normal en los decimales y no tienden a ser redondeados al entero.

Figura N°20 Decimals Campanario (PCP)



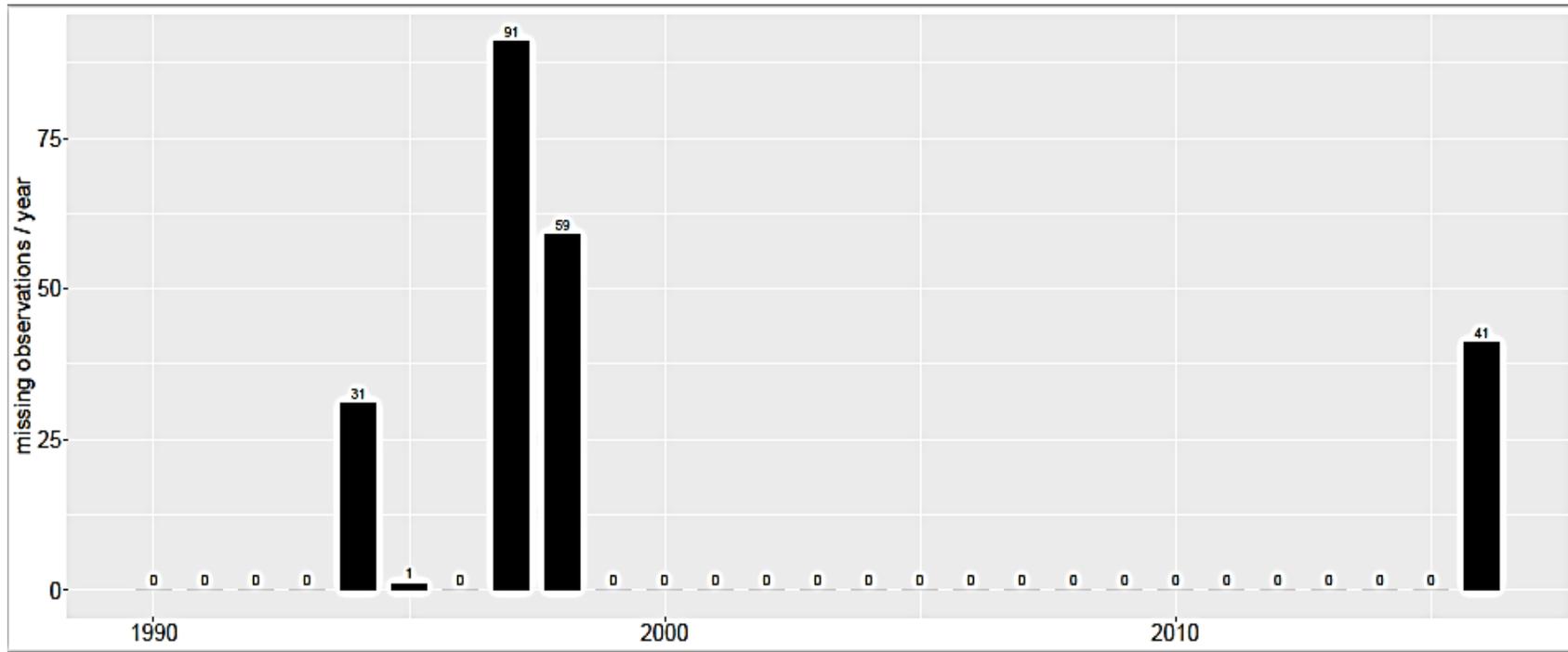
Fuente: Elaboración Propia

- Missing. - Nos presenta un histograma temporal (años) de datos faltantes para las tres variables analizadas (PCP, TX, TN), en el mismo se puede ver la cantidad de datos faltantes por año, considerando tanto datos consecutivos como no consecutivos, por lo que se nos permite establecer la cantidad de días faltantes por año.

En la Figura N°21 se puede observar la cantidad de datos faltantes por años para la variable precipitación de la estación Campanario, se puede ver que en el año 1994 no se tiene registro de 31 días, en 1995 de un día, 1997 de 91 días, 1998 de 59 días y en 2016 de 41 días y el resto de los años tiene completo su registro para esta variable.

- Weekly cycles. - Es un diagrama de barras de días que probablemente sean indicadores de ciclos semanales, en el mismo se puede observar con rojo el día que sería el candidato a producir ciclos semanales en la serie de acuerdo al test binomial aplicado a cada día de la semana para toda la serie.

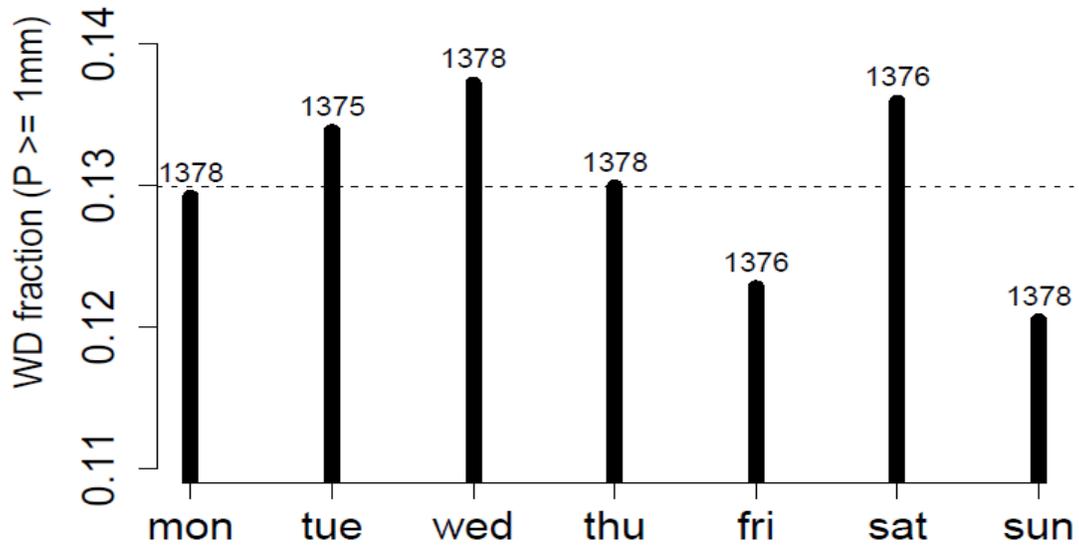
Figura N°21 Missing Campanario (PCP)



Fuente: Elaboración Propia

Para el caso de la estación Campanario como se muestra en la Figura N°22 no se presenta un día que sea indicador de un ciclo semanal ya que para todos los días de la semana se presenta la misma cantidad de registros

Figura N°22 Weekly Cycles Campanario (PCP)



Fuente: Elaboración Propia

5. Statistics. – Son archivos de estadística de datos faltantes para cada una de las variables analizadas en el que se podrá verificar cuantas estaciones se han evaluado, cuantas tienen más de veinte años de registro y cuantas tienen un registro inferior a lo antes mencionado.

Figura N°23 Statistics Campanario (PCP)

```

stations total 1
data >= 9362.25 days 1
data < 9362.25 days 0
pcp_campanario NA

```

Fuente: Elaboración Propia

En la Figura N°23 observamos en la primera fila el número de estaciones analizadas que es una, ya que se realiza el análisis de las estaciones de una en una,

en la segunda fila nos muestra la cantidad de estaciones que tienen registros mayores a 20 años que de igual manera es una. En la tercera fila se observa la cantidad de estaciones con registro menor a 20 años el cuál es cero, ya que al haber analizado una sola estación y la misma tener registro de más de 20 años, la cantidad de estaciones que tienen menos que este registro es cero y en la última fila observamos el nombre de la estación analizada seguido de “NA” que nos indica que se realizó satisfactoriamente el análisis ya que si el mismo estaría al comienzo del nombre el análisis no se habría realizado.

3.3. Criterios de Selección de Estaciones

3.3.1. Primer Criterio de Selección de Estaciones

De acuerdo al análisis realizado con la herramienta DECADE empleada en el punto anterior obtenemos también el porcentaje de completitud en los registros para cada una de las estaciones, como se muestra en la Tabla N° 11.

Tabla N° 11 Porcentaje de Completitud de datos

COD	ESTACION	PORCENTAJE DE COMPLETITUD (%)		
		PCP	TX	TN
GU001	Alto Cajas	60		
GU002	Campanario	95	95	95
GU003	Cañas	95	60	50
GU004	Ce.Na.Vit	90	90	90
GU005	Ciudad-Tja	90		
GU006	Coimata	100	85	85
GU007	El Tejar Tarija	90	90	90
GU008	Juntas	95	55	55
GU009	Paicho Centro	70	15	15
GU010	Pinos Sud	90		
GU011	Rosillas_h	20		
GU012	San Andrés	95	55	55
GU013	San Jacinto Sud	95	65	65
GU014	San Luis de Palqui	70		
GU015	Sella quebrada	90	60	60
GU016	Tarija Aeropuerto	95	75	75
GU017	Tomayapo Pueblo	95	30	30
GU018	Trancas	90	80	80

GU019	Tucumillas	95	40	40
GU020	Turumayo	100	95	95
GU021	Yesera Norte	90	60	60
GU022	Yesera Sur	100	95	95
GU023	Alizos	90		
GU024	Calamuchita	90	15	15
GU025	Calderillas	95		
GU026	Canasmoro	80	70	70
GU027	Canchasmayu	95		
GU028	Chocloca	85	40	40
GU029	Colon Sud	90		
GU030	Concepción_tj	95		
GU031	Copacabana Taxara	75		
GU032	El Molino Tomayapo	90		
GU033	Gamoneda	90		
GU034	Iscaiyachi San Antonio	90	75	75
GU035	Junacas	90		
GU036	Ladera Centro	85		
GU037	León Cancha	95		
GU038	Padcaya	95	95	95
GU039	Pampa Redonda_H	90		
GU040	Rejara	95		
GU041	Rincón Cañas_H	90		
GU042	Sama Cumbre	95		
GU043	Sama Iscaiyachi	90		
GU044	San Agustín_tja	90		
GU045	San Lorenzo	90		
GU046	San Nicolás	90	90	90
GU047	San Pedro de Buena Vista_tja	60	5	5
GU048	Santa Ana Puente	45	3	3
GU049	Tarija Cancha	80		
GU050	Tarija Senamhi_h	95		
GU051	Tolomosa Grande	95		
GU052	Tomatitas	80		

Fuente: Elaboración Propia

Flexibilizando el criterio de selección aplicado en el Balance Hídrico Superficial de Bolivia 2017 de emplear estaciones que cuenten con el 80% o más de completitud, se tomó en cuenta aquellas que cuentan con un porcentaje de completitud igual o mayor al 60% de datos en el periodo 1980-2016. Como resultado del análisis de este criterio quedaron para la continuidad

del estudio 26 estaciones que se muestran a continuación en la Tabla N°12:

Tabla N° 12 Estaciones empleadas (1980-2016)

N°	COD	ESTACIÓN
1	GU001	Alto Cajas
2	GU002	Campanario
3	GU003	Cañas
4	GU004	Ce.Na. Vit
5	GU005	Ciudad-Tja
6	GU006	Coimata
7	GU007	El Tejar Tarija
8	GU008	Juntas
9	GU010	Pinos Sud
10	GU012	San Andrés
11	GU013	San Jacinto Sud
12	GU015	Sella quebrada
13	GU016	Tarija Aeropuerto
14	GU017	Tomayapo Pueblo
15	GU018	Trancas
16	GU019	Tucumillas
17	GU021	Yesera Norte
18	GU025	Calderillas
19	GU026	Canasmoro
20	GU028	Chocloca
21	GU032	El Molino Tomayapo
22	GU033	Gamoneda
23	GU035	Junacas
24	GU036	Ladera Centro
25	GU037	León Cancha
26	GU040	Rejara

Fuente: Elaboración Propia

3.3.1.1. Relleno de series faltantes

Para el periodo 1980-2016, de las 17 estaciones abiertas que se utilizaron para este estudio se realizó el relleno de 12 de las mismas y se utilizaron cinco del Balance Hídrico Nacional 2017, además de igual forma se rellenaron cinco estaciones cerradas y se utilizaron cuatro ya completas del balance realizado (Tabla N°13).

Para completar las series de datos faltantes se aplicó la herramienta Filldata que es un software desarrollado por el Servicio Nacional de Meteorología e Hidrología (SENAMHI), este permite la reconstrucción de lagunas en las series históricas de datos meteorológicos e hidrológicos aplicando métodos recomendados por la Organización Meteorológica Mundial, que son cuatro métodos univariados y nueve métodos multivariados.

En este estudio se aplicó el método multivariado razón o ratio normal (Paulus y Kohler) que consiste en calcular el dato incompleto, $x(t)$, de una serie a partir de los datos de las series de estaciones vecinas y contemporáneas, que presenten un alto grado de correlación con la serie a completar.

Tabla N°13 Estaciones Rellenadas

COD	ESTACION	PROVINCIA	TIPO	RELLENO
GU001	Alto Cajas	MéndeZ	P	Método multivariado, razón o ratio normal
GU002	Campanario	MéndeZ	CP	Método multivariado, razón o ratio normal
GU003	Cañas	Arce	CO	Balance hídrico Nacional de Bolivia
GU004	Ce.Na.Vit	Avilés	CP	Balance hídrico Nacional de Bolivia
GU005	Ciudad-Tja	Cercado	P	Método multivariado, razón o ratio normal
GU006	Coimata	MéndeZ	CP	Método multivariado, razón o ratio normal
GU007	El Tejar Tarija	Cercado	CP	Método multivariado, razón o ratio normal
GU008	Juntas	Avilés	CP	Balance hídrico Nacional de Bolivia
GU010	Pinos Sud	Cercado	P	Método multivariado, razón o ratio normal
GU012	San Andrés	Cercado	CP	Método multivariado, razón o ratio normal
GU013	San Jacinto Sud	Cercado	P	Balance hídrico Nacional de Bolivia
GU015	Sella quebrada	MéndeZ	CO	Método multivariado, razón o ratio normal
GU016	Tarija Aeropuerto	Cercado	S	Balance hídrico Nacional de Bolivia
GU017	Tomayapo Pueblo	MéndeZ	TP	Método multivariado, razón o ratio normal
GU018	Trancas	MéndeZ	CO	Método multivariado, razón o ratio normal
GU019	Tucumillas	MéndeZ	CO	Método multivariado, razón o ratio normal
GU021	Yesera Norte	Cercado	CP	Método multivariado, razón o ratio normal
GU025	Calderillas	Avilés	P	Balance hídrico Nacional de Bolivia
GU026	Canasmoro	MéndeZ	CO	Método multivariado, razón o ratio normal
GU028	Chocloca	Avilés	CO	Balance hídrico Nacional de Bolivia

GU032	El Molino Tomayapo	Méndez	P	Método multivariado, razón o ratio normal
GU033	Gamoneda	Cercado	P	Método multivariado, razón o ratio normal
GU035	Junacas	Cercado	P	Método multivariado, razón o ratio normal
GU036	Ladera Centro	Avilés	P	Balance hídrico Nacional de Bolivia
GU037	León Cancha	Méndez	P	Método multivariado, razón o ratio normal
GU040	Rejara	Arce	P	Balance hídrico Nacional de Bolivia

<p>Referencias</p> <p>CO : Climatología Ordinaria CP : Climatología Principal P : Pluviométrica S : Sinóptica TP : Termopluiométrica</p>
--

Fuente: Elaboración Propia

3.3.1.2. Aplicación de la Grilla Meteorológica GMET

Después de la depuración y análisis de series, 26 estaciones individuales de temperatura y precipitación son consideradas para la interpolación. GMET fue corrido para la Cuenca del Río Guadalquivir usando las estaciones presentadas en la Tabla N°12, para obtener un producto final de una resolución espacial de 0.05° (celda) a escala de tiempo diario para el período 1980-2016.

El algoritmo de GMET produce 30 miembros de un ensamble de resultados; cada miembro tiene estimaciones de precipitación diaria para el periodo analizado en el dominio del estudio.

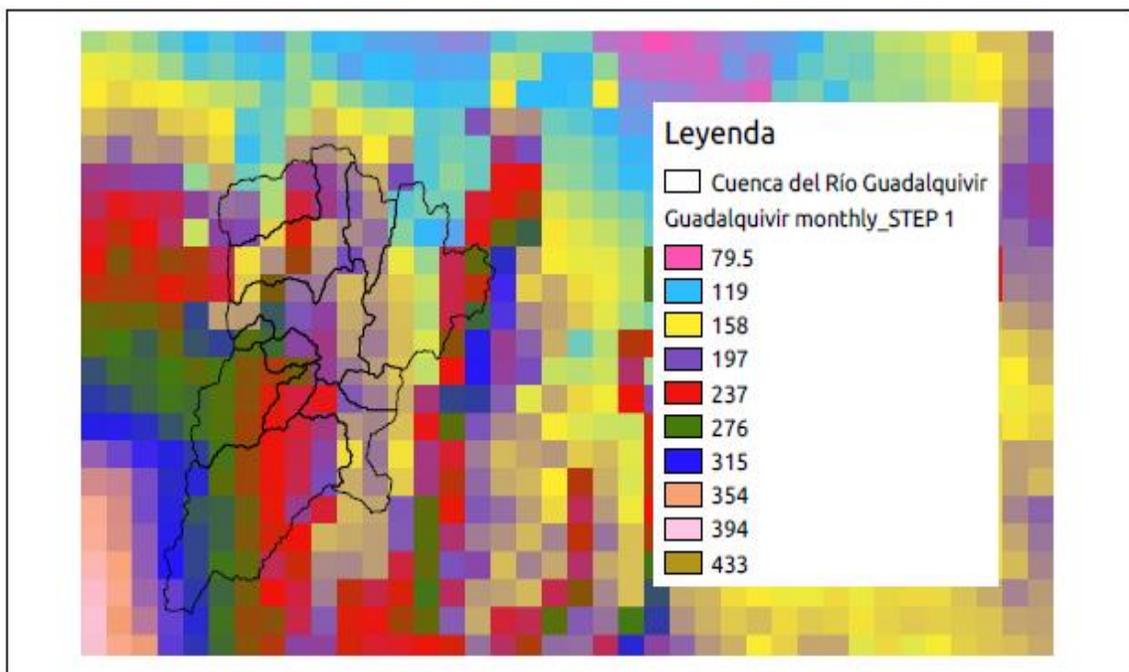
Para construir estos ensambles de grillas meteorológicas GMET completa varios pasos necesarios que son:

- Lectura de datos de estaciones.
- Se asegura que las series de tiempo de las estaciones estén completas (rechaza aquellas series que sean muy cortas para ser validadas).
- Interpola series completas de precipitación y temperatura, estima la probabilidad de ocurrencia de precipitación y la probabilidad distribución de valores de los mismos.

- Genera ensambles de campos espaciales grillados usando campos aleatorios correlacionados espacial y temporalmente para obtener valores a partir de la distribución estimada generada en el paso anterior.

Posteriormente se realizan los promedios correspondientes de los mismos, en el caso de la acumulación mensual de los ensambles se tiene 440 pasos de tiempo para el periodo antes mencionado. En las figuras a continuación, se muestran como ejemplo los resultados en algunos pasos de tiempo:

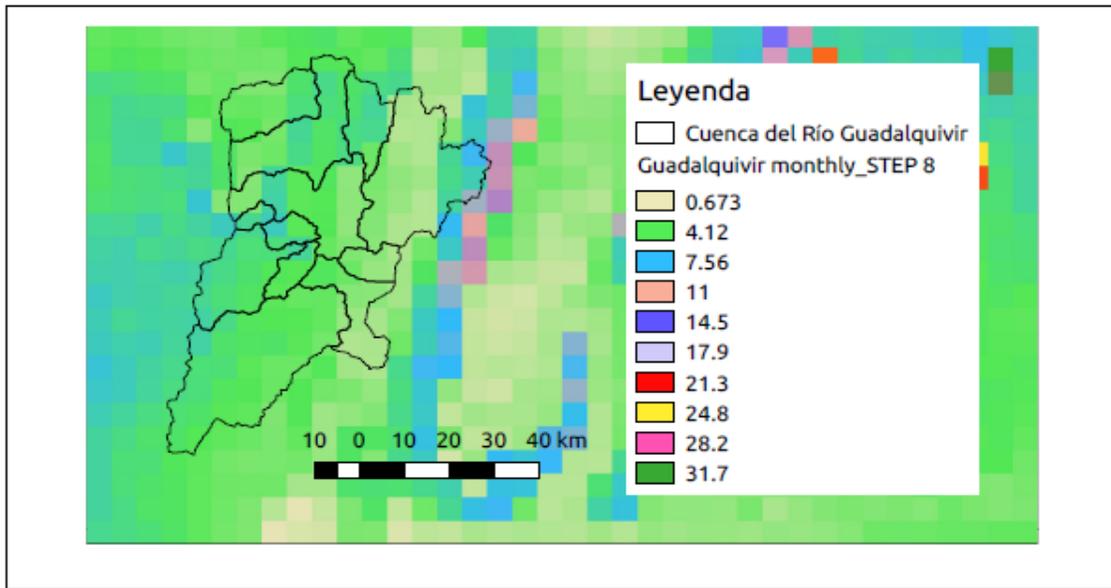
Figura N°24 Guadalquivir_monthly_STEP 1



Monsum		
STEP 1, Enero 1980	Desde	Hasta
DEM 90, resolución espacial 0.05	119	315

Fuente: Elaboración Propia

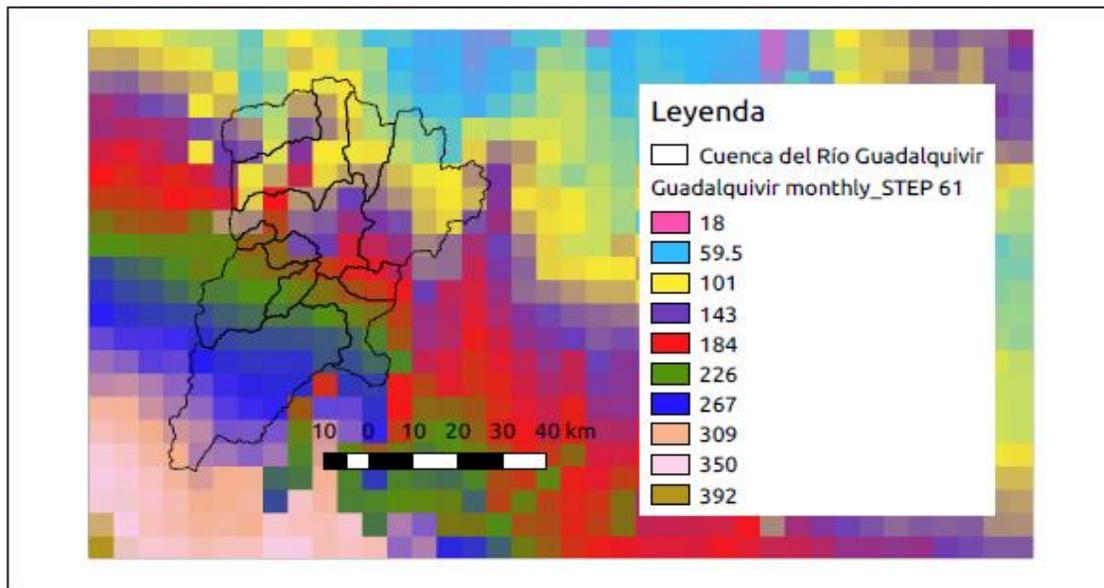
Figura N°25 Guadalquivir_monthly_STEP 8



Monsum			
STEP 8, Agosto 1980	Desde	Hasta	
DEM 90, resolución espacial 0.05	0.67	11	

Fuente: Elaboración Propia

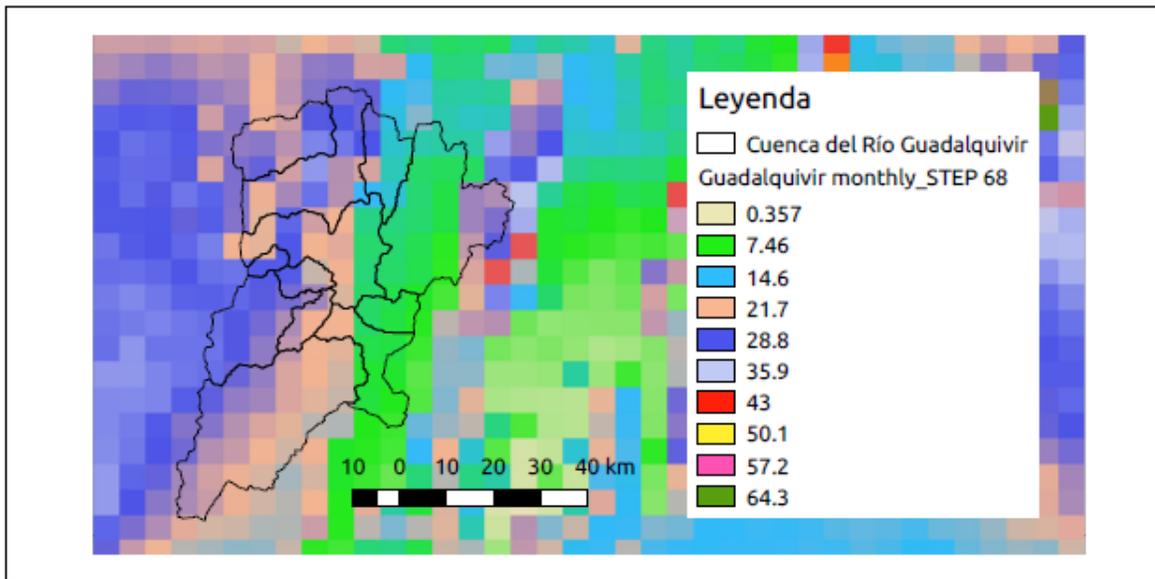
Figura N°26 Guadalquivir_monthly_STEP 61



Monsum			
STEP 61, Enero 1985	Desde	Hasta	
DEM 90, resolución espacial 0.05	59.5	309	

Fuente: Elaboración Propia

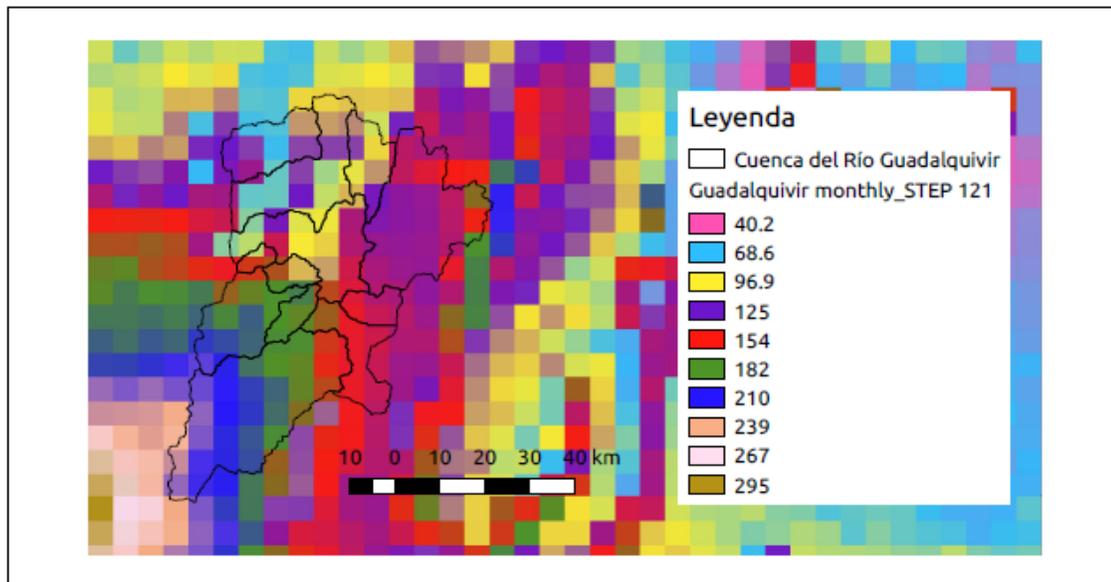
Figura N°27 Guadalquivir_monthly_STEP 68



Monsum		
STEP 68, Agosto 1985	Desde	Hasta
DEM 90, resolución espacial 0.05	0.36	35.9

Fuente: Elaboración Propia

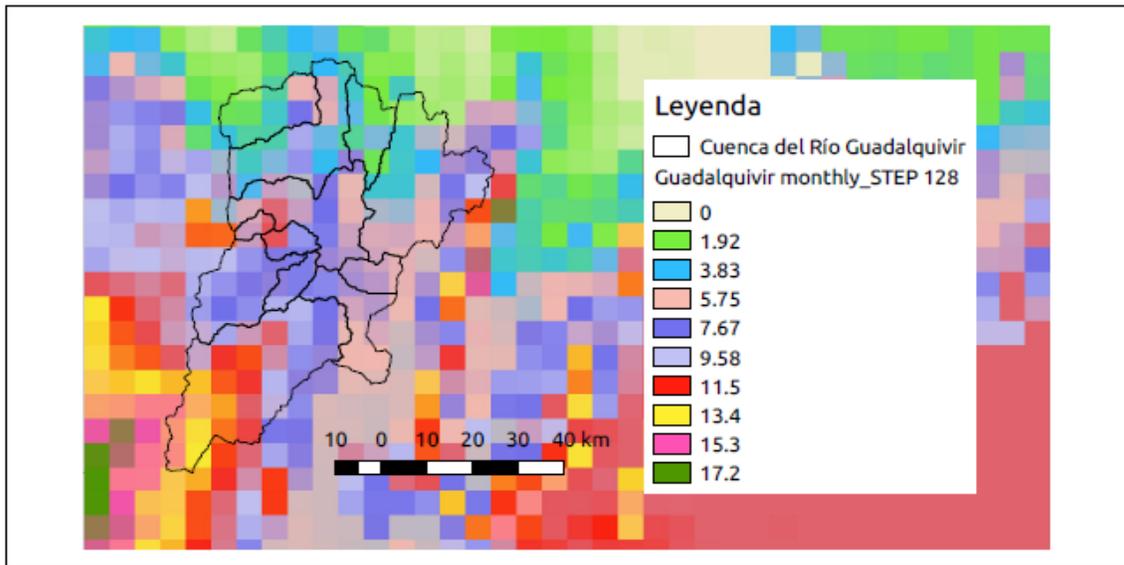
Figura N°28 Guadalquivir_monthly_STEP 121



Monsum		
STEP 121, Enero 1990	Desde	Hasta
DEM 90, resolución espacial 0.05	68.6	239

Fuente: Elaboración Propia

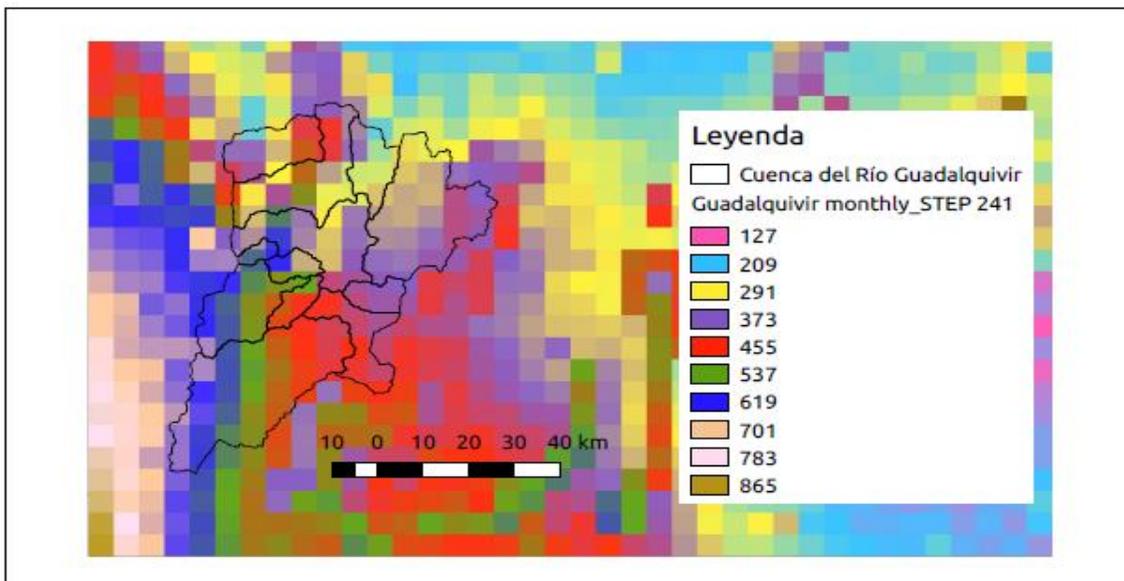
Figura N°29 Guadalquivir_monthly_STEP 128



Monsum		
STEP 128, Agosto 1990	Desde	Hasta
DEM 90, resolución espacial 0.05	0.0	15.3

Fuente: Elaboración Propia

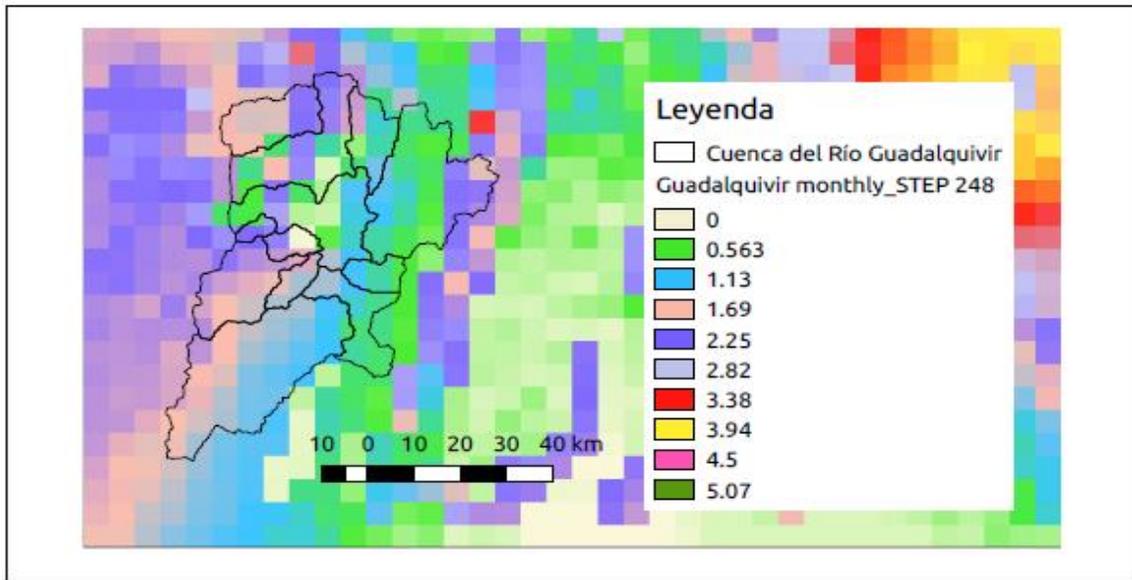
Figura N°30 Guadalquivir_monthly_STEP 241



Monsum		
STEP 241, Enero 2000	Desde	Hasta
DEM 90, resolución espacial 0.05	127	619

Fuente: Elaboración Propia

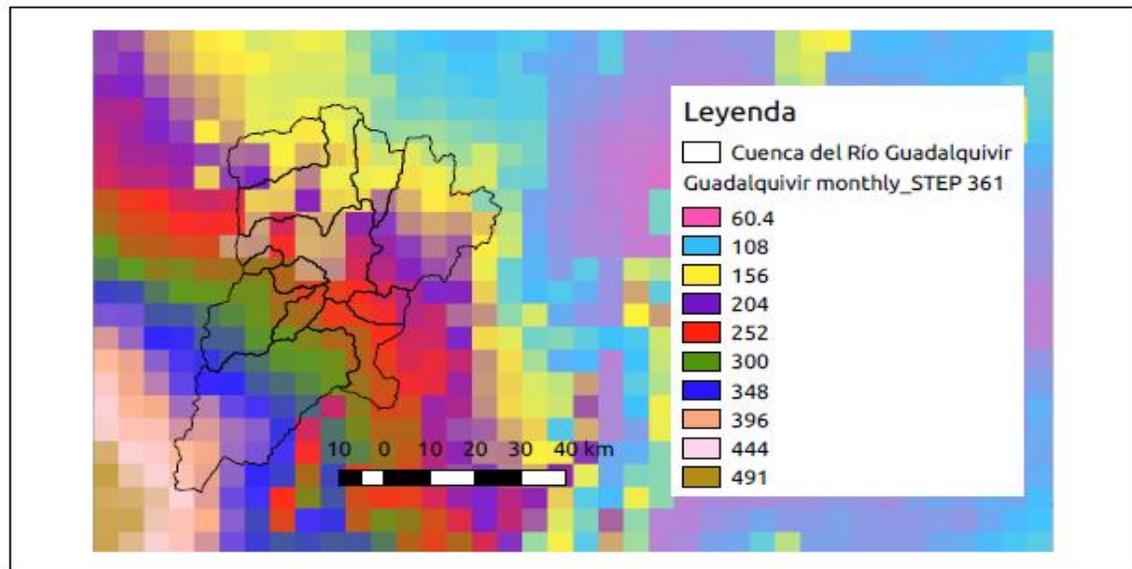
Figura N°31 Guadalquivir_monthly_STEP 248



Monsum		
STEP 248, Agosto 2000	Desde	Hasta
DEM 90, resolución espacial 0.05	0.0	2.82

Fuente: Elaboración Propia

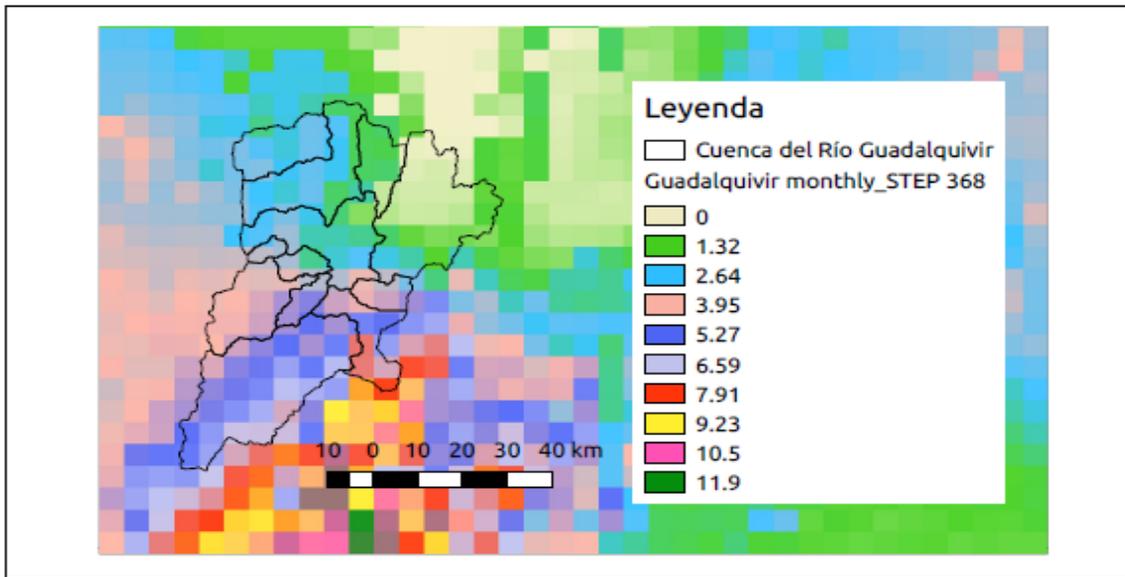
Figura N°32 Guadalquivir_monthly_STEP 361



Monsum		
STEP 361, Enero 2010	Desde	Hasta
DEM 90, resolución espacial 0.05	60.4	396

Fuente: Elaboración Propia

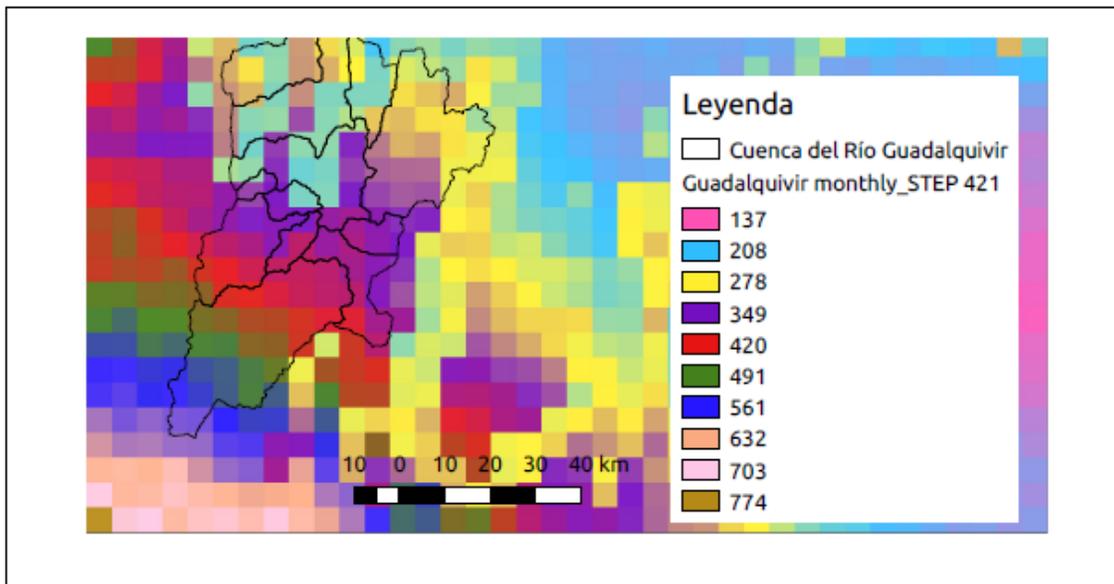
Figura N°33 Guadalquivir_monthly_STEP 368



Monsum		
STEP 368, Agosto 2010	Desde	Hasta
DEM 90, resolución espacial 0.05	0.0	9.23

Fuente: Elaboración Propia

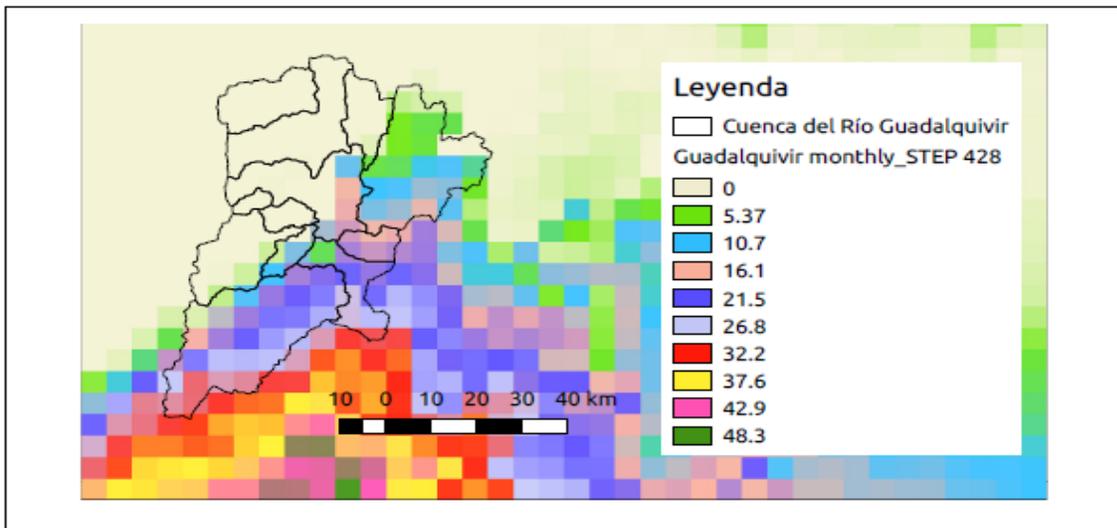
Figura N°34 Guadalquivir_monthly_STEP 421



Monsum		
STEP 421, Enero 2015	Desde	Hasta
DEM 90, resolución espacial 0.05	137	561

Fuente: Elaboración Propia

Figura N°35 Guadalquivir_monthly_STEP 428

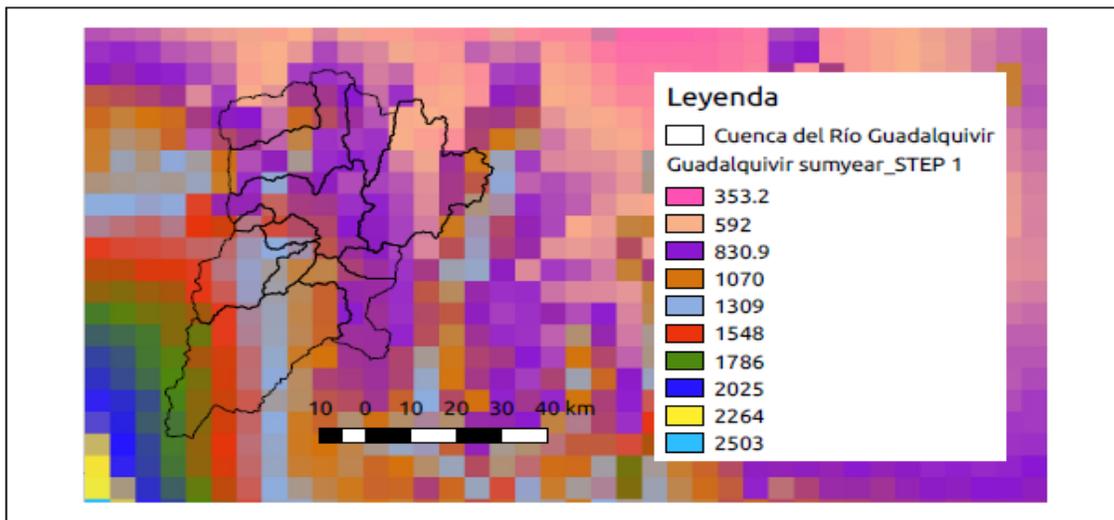


Monsum		
STEP 428, Agosto 2015	Desde	Hasta
DEM 90, resolución espacial 0.05	0.0	32.2

Fuente: Elaboración Propia

Para el promedio anual de los ensambles generados se tiene 37 pasos de tiempo. A continuación, se muestran como ejemplo los resultados en algunos pasos de tiempo:

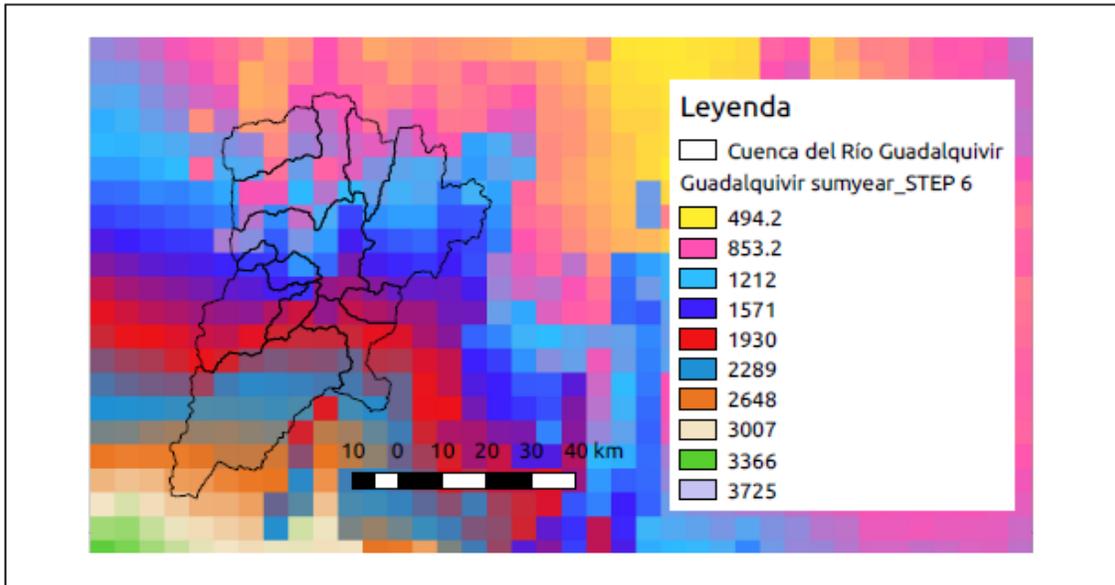
Figura 36 Guadalquivir_Sumyear_STEP 1



Sumyear		
STEP 1, 1980	Desde	Hasta
DEM 90, resolución espacial 0.05	353.2	1786

Fuente: Elaboración Propia

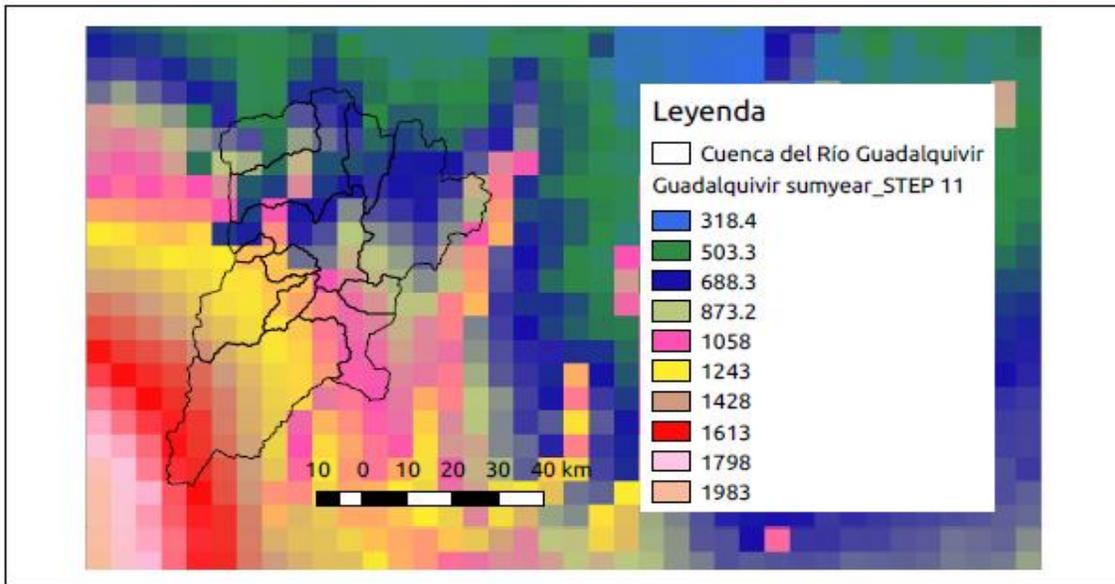
Figura N°37 Guadalquivir_Sumyear_STEP 6



Sumyear			
STEP 6, 1985	Desde	Hasta	
DEM 90, resolución espacial 0.05	494.2	2648	

Fuente: Elaboración Propia

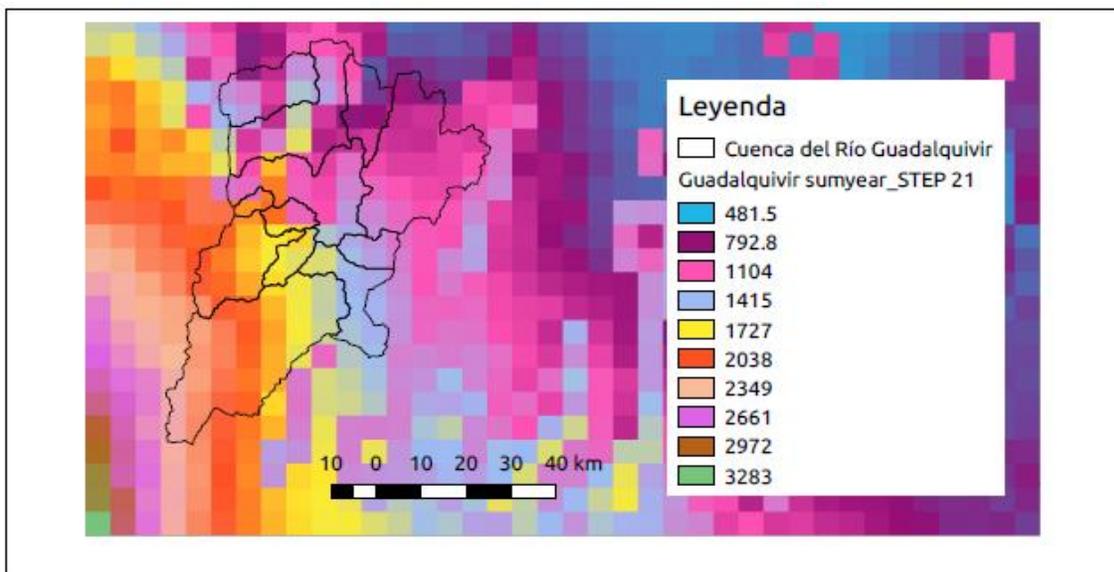
Figura N°38 Guadalquivir_Sumyear_STEP 11



Sumyear			
STEP 11, 1990	Desde	Hasta	
DEM 90, resolución espacial 0.05	318.4	1613	

Fuente: Elaboración Propia

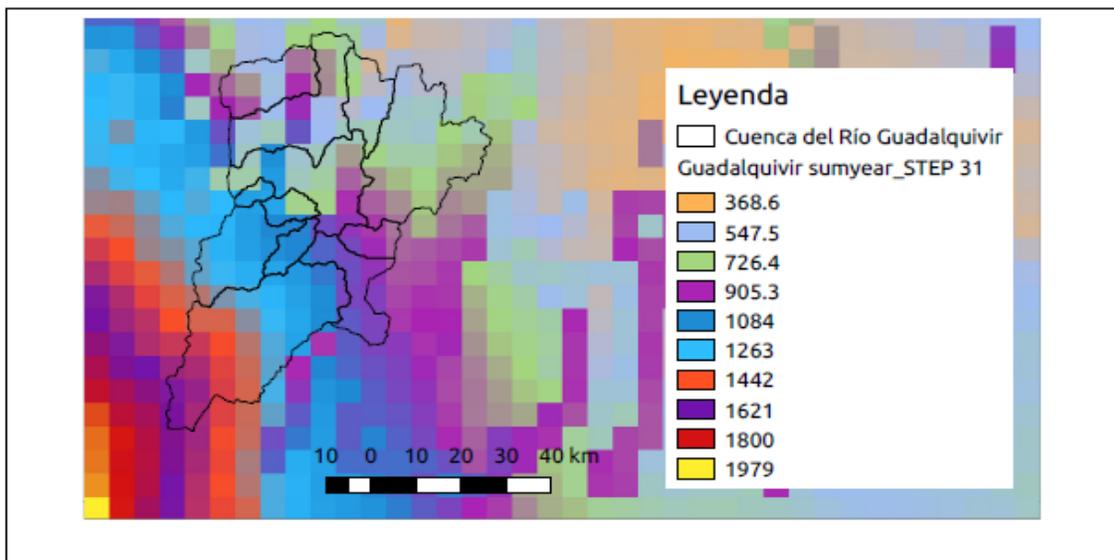
Figura N°39 Guadalquivir_Sumyear_STEP 21



Sumyear			
STEP 21, 2000		Desde	Hasta
DEM 90, resolución espacial 0.05		481.5	2661

Fuente: Elaboración Propia

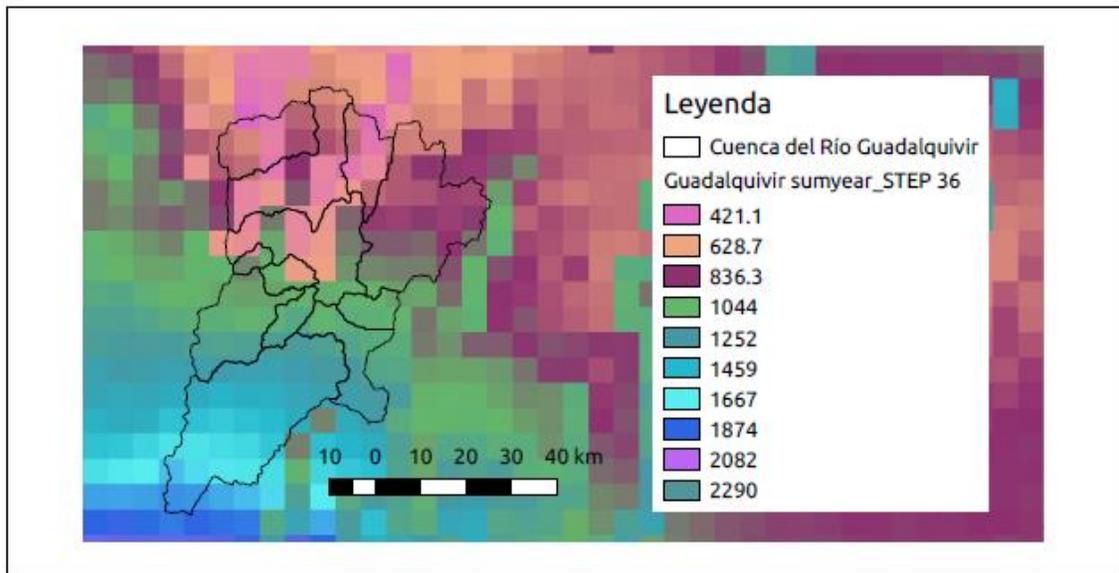
Figura N°40 Guadalquivir_Sumyear_STEP 31



Sumyear			
STEP 31, 2010		Desde	Hasta
DEM 90, resolución espacial 0.05		368.6	1621

Fuente: Elaboración Propia

Figura N°41 Guadalquivir_Sumyear_STEP 35



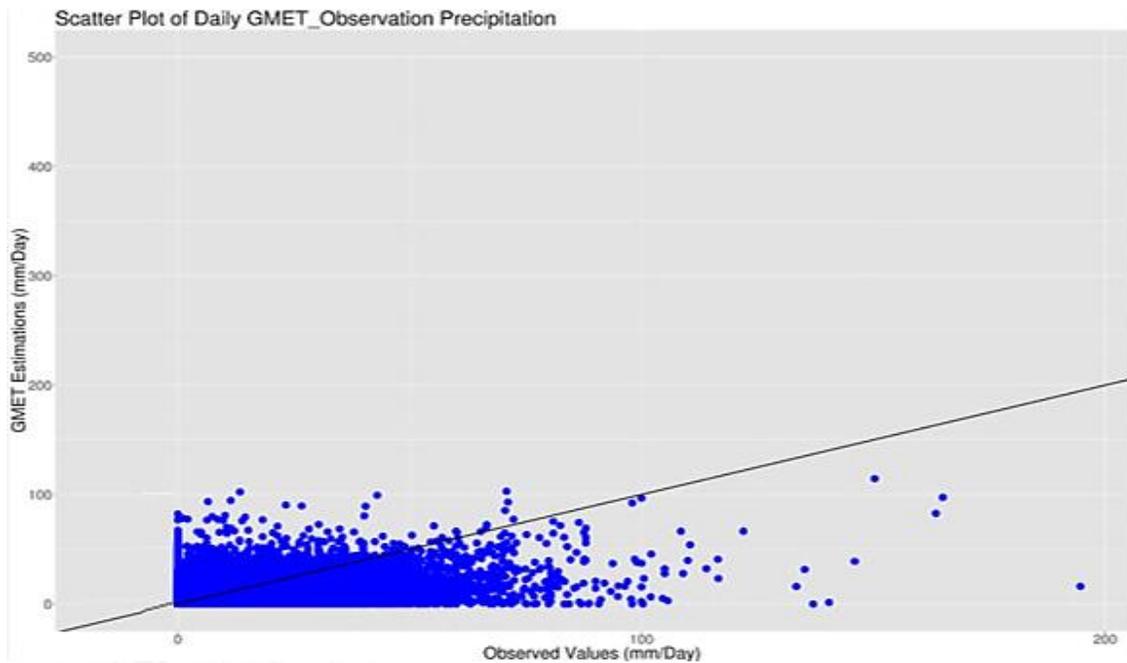
Sumyear			
STEP 36, 2015	Desde	Hasta	
DEM 90, resolución espacial 0.05	628.7	1874.0	

Fuente: Elaboración Propia

3.3.1.3. Validación de la grilla meteorológica GMET

La validación de la grilla se realiza de acuerdo a la evaluación del desempeño del modelo, usando un enfoque estadístico y de descomposición del error, la Figura N°42 muestra el diagrama de dispersión de los valores estimados por GMET respecto a los valores observados, el mismo sería perfectamente correcto cuando la línea de tendencia de los puntos estimados se alinea (estimados=observados).

Figura N°42 Diagrama de dispersión de GMET vs valores observados.



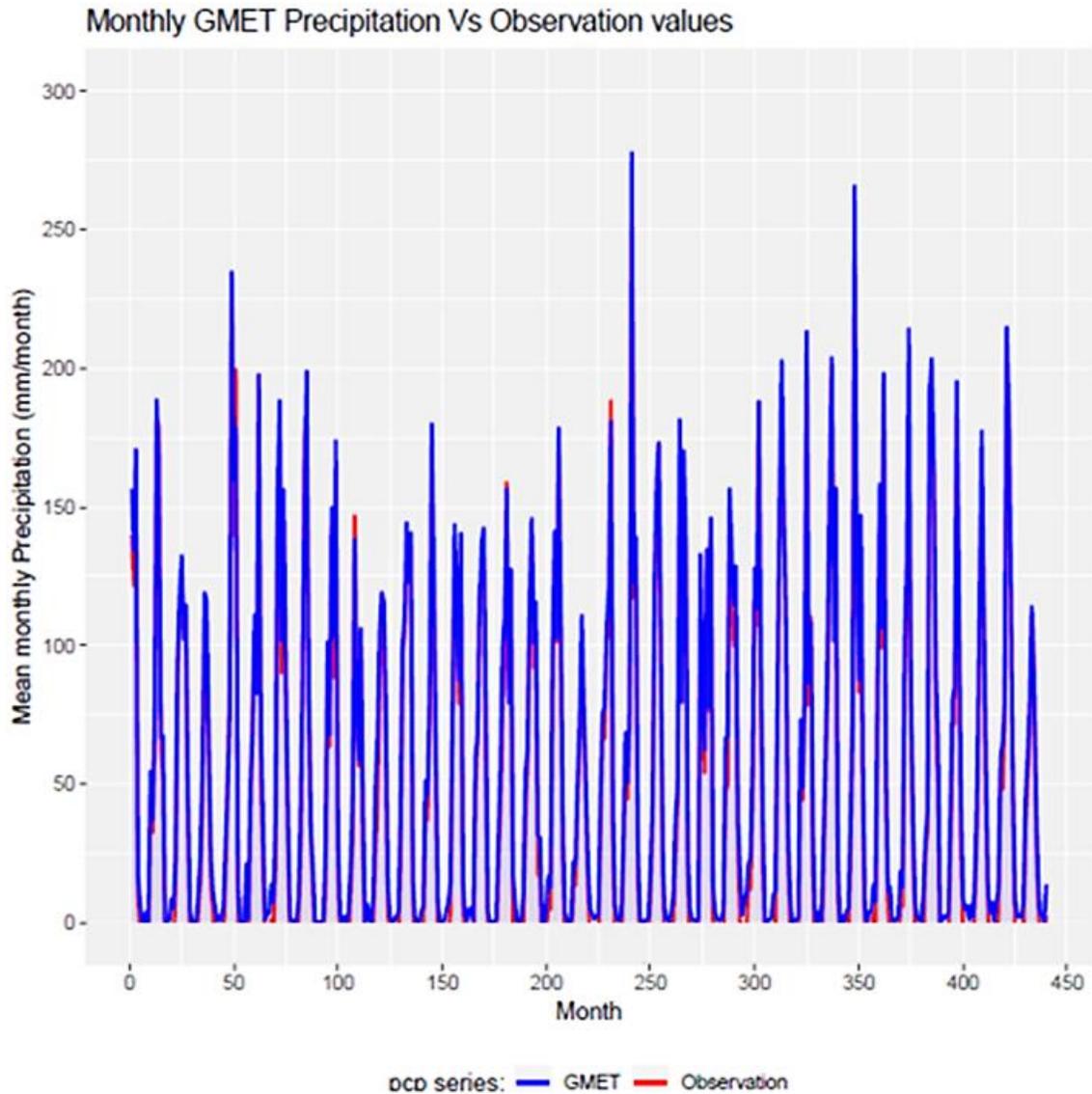
Fuente: Elaboración Propia

A continuación, se presentan los diferentes pasos de evaluación de la grilla generada:

3.3.1.3.1. Visualización de los datos

El primer paso en la comparación de los sets de datos es simplemente graficarlos para la escala temporal mensual, se preparó la Figura N°43 que compara las series mensuales de GMET y de las observaciones en la Cuenca del Río Guadalquivir. Se observa que GMET reproduce bien los patrones mensuales de variación local de la red de datos y está estimando de forma cercana a los puntos de validación.

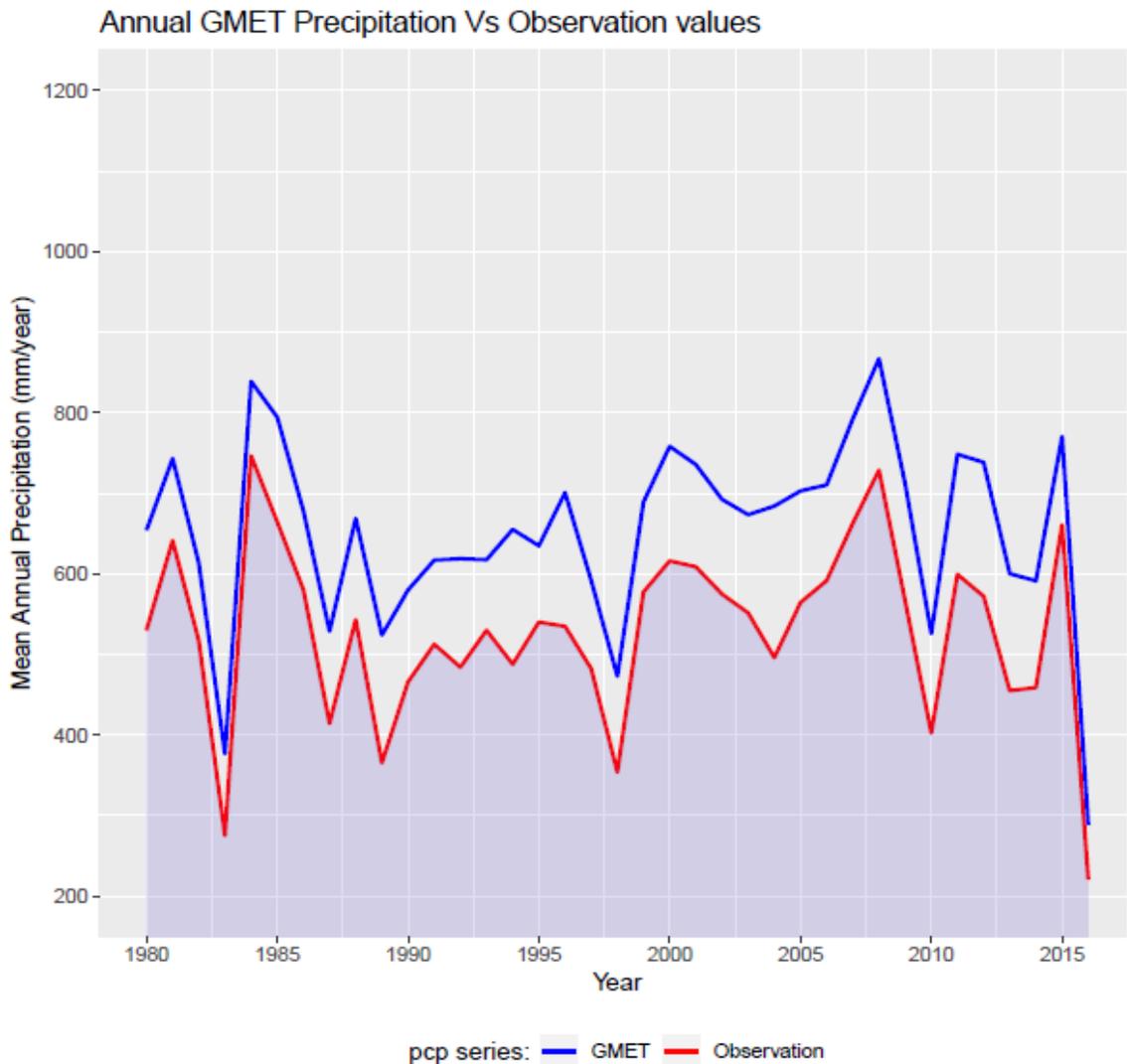
Figura N°43 Series Mensuales de GMET vs estaciones en la Cuenca del Río Guadalquivir



Fuente: Elaboración Propia

Además, se generó la gráfica para datos anuales como se observa en la Figura N°44, GMET reproduce los patrones anuales de variación local de la red de datos de la Cuenca del Río Guadalquivir, pero presenta una sobrestimación respecto a los datos observados en todo el periodo analizado mostrando el cambio anual de la precipitación sobre la cuenca en estudio.

Figura N°44 Series Anuales de GMET vs estaciones en la Cuenca del Río Guadalquivir



Fuente: Elaboración Propia

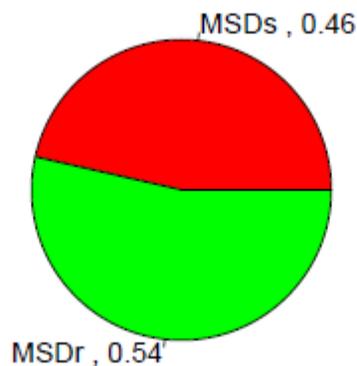
3.3.1.3.2. Descomposición del error

En la Figura N°45 se muestran los componentes de error sistemático y aleatorio para las series de tiempo diarias, los resultados indican que en general, el 46% del error asociado con GMET es causado por situaciones del sistema de estimación y que el 54% es causado por errores aleatorios. (Componente sistemático MSD_s y un componente aleatorio MSD_r).

Al calcular las porciones del error sistemático y aleatorio es posible identificar que tanto GMET está capturando el proceso físico de la precipitación en la zona de la cuenca del Río Guadalquivir. Si el error sistemático es mayor que el componente del error aleatorio esto

significa que GMET debe ser ajustado basado en los patrones hidrológicos de la cuenca. De otra forma si el error aleatorio es dominante, esto significa que la mayoría de los errores son causa de la incertidumbre que es inherente al fenómeno de la precipitación.

Figura N°45 Componentes del error sistemático y aleatorio



Fuente: Elaboración Propia

3.3.2. Segundo Criterio de Selección de Estaciones

De acuerdo al análisis visual de la Tabla N°4 y 7 el segundo criterio de selección fue el número de estaciones, que cuentan con una buena cantidad de registros en el periodo 1980-1992, por lo que se definió el mismo como un segundo periodo de análisis, tomando en cuenta estaciones cerradas y abiertas se alcanza un total de 40 estaciones. De las cuales tres no pasaron el control de calidad realizado por RClimdex; es por esto, que la generación de la grilla meteorológica GMET se realizó con un total de 37 estaciones que se muestran a continuación en la Tabla N°14:

Tabla N° 14 Estaciones Empleadas (1980-1992)

COD	ESTACIÓN	Latitud	Longitud	Elevación
GU001	Alto Cajas	-21.3	-64.47	2980
GU002	Campanario	-21.51	-64.98	3460
GU003	Cañas	-21.9	-64.85	2030
GU004	Ce.Na.Vit	-21.73	-64.66	1730
GU005	Ciudad-Tja	-21.54	-64.73	1864
GU006	Coimata	-21.5	-64.79	2027
GU007	El Tejar Tarija	-21.54	-64.72	1859

GU008	Juntas	-21.81	-64.8	1900
GU010	Pinos Sud	-21.76	-64.88	2100
GU012	San Andrés	-21.62	-64.82	1987
GU013	San Jacinto Sud	-21.61	-64.72	1820
GU015	Sella quebrada	-21.38	-64.7	2080
GU016	Tarija Aeropuerto	-21.55	-64.71	1875
GU017	Tomayapo Pueblo	-21.27	-65.05	2720
GU018	Trancas	-21.31	-64.82	2200
GU019	Tucumillas	-21.46	-64.83	2560
GU021	Yesera Norte	-21.37	-64.55	2200
GU023	Alizos	-21.82	-64.87	2040
GU025	Calderillas	-21.75	-64.95	2800
GU026	Canasmoro	-21.35	-64.75	2080
GU027	Canchasmayu	-21.88	-64.88	1120
GU028	Chocloca	-21.75	-64.73	1795
GU031	Copacabana Taxara	-21.8	-65.1	3800
GU032	El Molino Tomayapo	-21.37	-64.95	3410
GU033	Gamoneda	-21.48	-64.63	1150
GU034	Iscayachi San Antonio	-21.39	-64.95	3450
GU035	Junacas	-21.43	-64.47	2300
GU036	Ladera Centro	-21.65	-64.53	2080
GU037	León Cancha	-21.18	-64.72	2600
GU038	Padcaya	-21.88	-64.72	2010
GU039	Pampa Redonda_h	-21.7	-64.82	1950
GU040	Rejara	-22.02	-64.98	3000
GU041	Rincón Cañas_h	-21.97	-64.9	2280
GU042	Sama Cumbre	-21.49	-64.98	3820
GU044	San Agustín_tja	-21.53	-64.52	2210
GU045	San Lorenzo	-21.42	-64.75	1900
GU051	Tolomosa Grande	-21.62	-64.77	1900

Fuente: Elaboración Propia

3.3.2.1. Relleno de series faltantes

Para el periodo 1980-1992, de las 11 estaciones cerradas que se pudieron adicionar a las analizadas en el periodo 1980-2016, se realizó el relleno de cinco de las mismas, las cuales se muestran en la Tabla N°15.

Tabla N° 15 Estaciones Rellenadas (1980-1992)

COD	ESTACION	PROVINCIA	TIPO	RELLENO
GU026	Canasmoro	Méndez	CO	Método multivariado, razón o ratio normal
GU031	Copacabana Taxara	Avilés	P	Método multivariado, razón o ratio normal
GU038	Padcaya	Arce	CO	Método multivariado, razón o ratio normal
GU042	Sama Cumbre	Méndez	P	Método multivariado, razón o ratio normal
GU044	San Agustín_tja	Cercado	P	Método multivariado, razón o ratio normal

Referencias

CO : Climatología Ordinaria

P : Pluviométrica

Fuente: Elaboración Propia

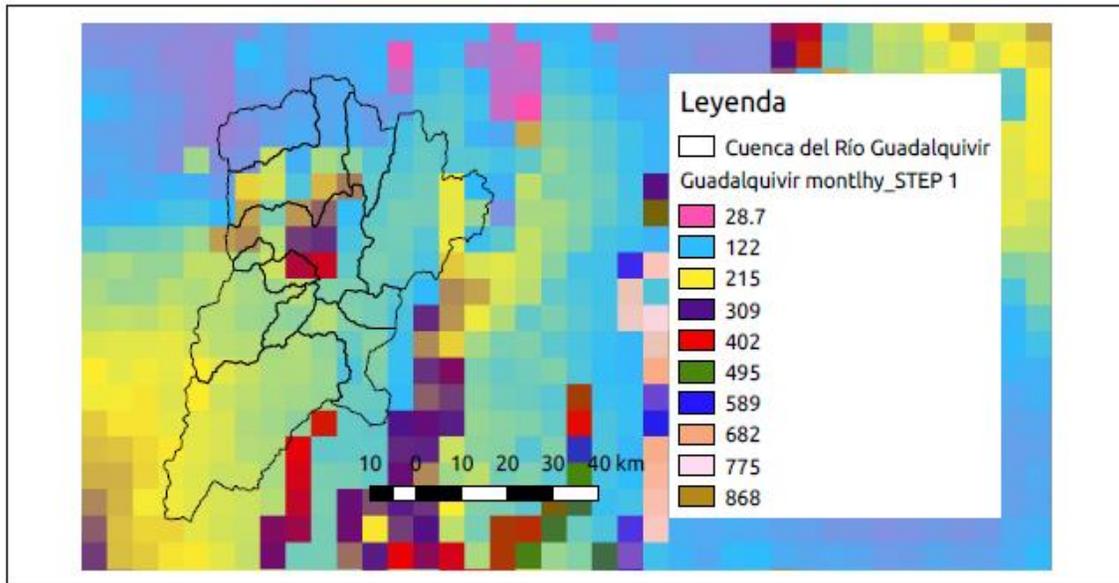
Para completar la serie de datos faltantes en este periodo se aplicó de igual manera la herramienta Filldata con el método multivariado razón o ratio normal (Paulus y Kohler) a partir de los datos de las series de estaciones vecinas, metodología que se encuentra detallada en el punto 3.3.1.1.

3.3.2.2. Aplicación de la Grilla Meteorológica GMET

Con la selección de las 37 estaciones mostradas en el punto 3.3.2 y realizado el control de calidad de los registros de las mismas se corrió GMET siguiendo la metodología aplicada en el punto 3.3.1.2.

En las figuras a continuación, se muestran como ejemplo los resultados en algunos de los 156 pasos de tiempo para la acumulación mensual de los ensambles en este periodo:

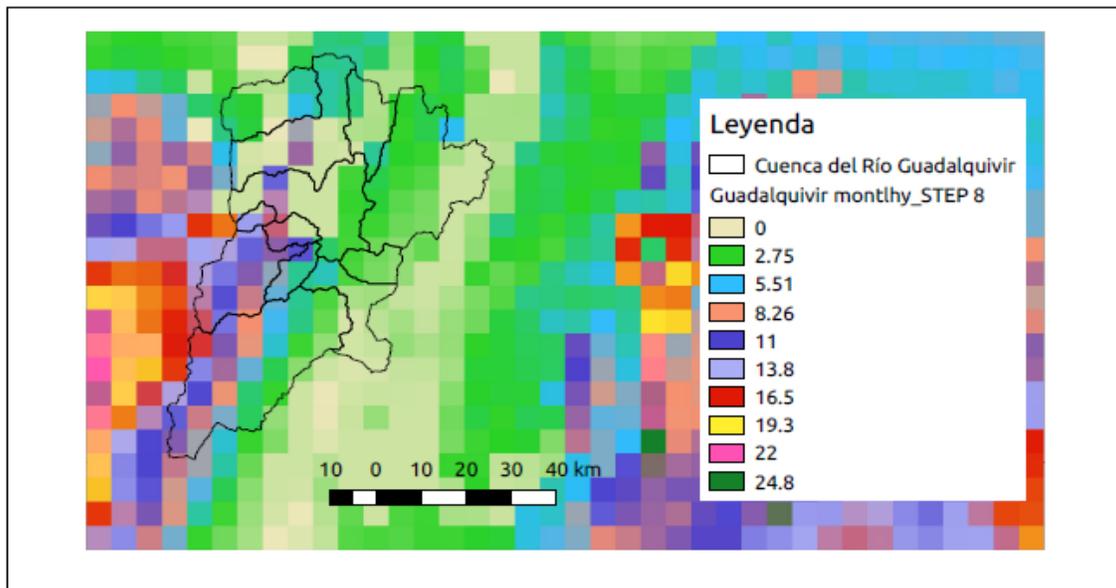
Figura N°46 Guadalquivir monthy_STEP 1



Monsum		
STEP 1, Enero 1980	Desde	Hasta
DEM 90, resolución espacial 0.05	122	402

Fuente: Elaboración Propia

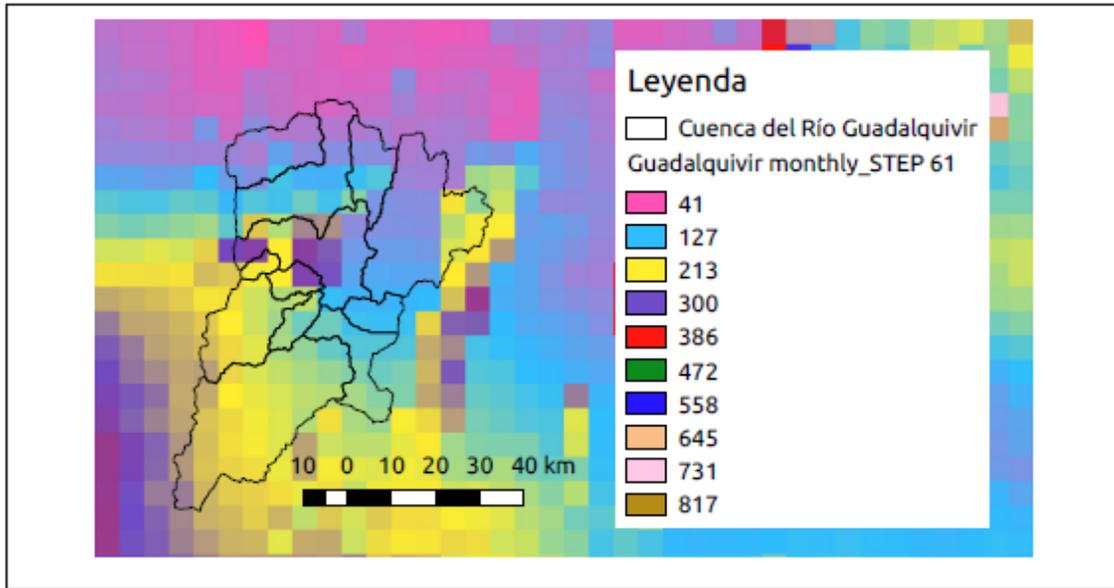
Figura N°47 Guadalquivir monthy_STEP 8



Monsum		
STEP 8, Agosto 1980	Desde	Hasta
DEM 90, resolución espacial 0.05	0.0	16.5

Fuente: Elaboración Propia

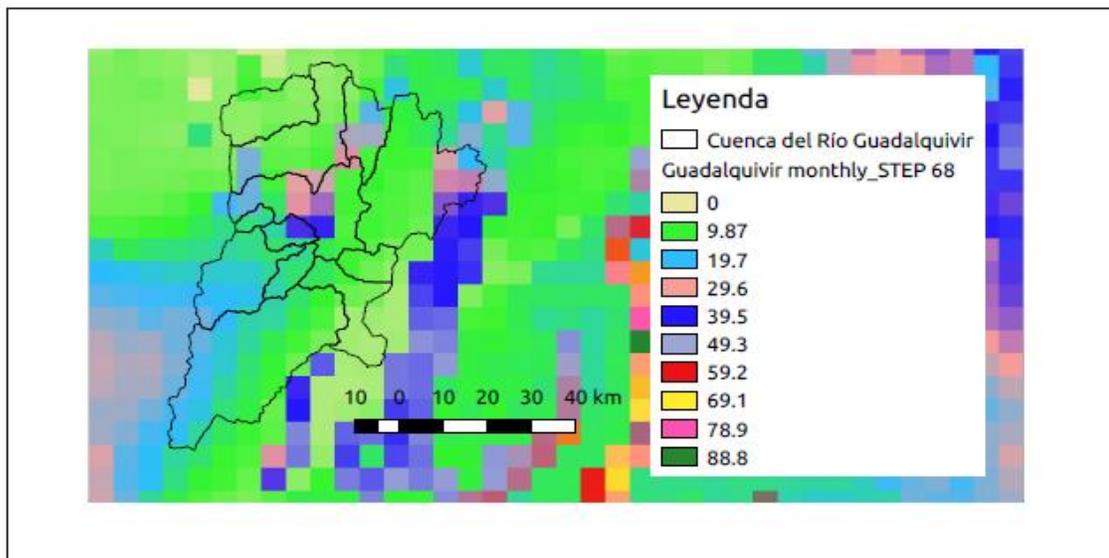
Figura N°48 Guadalquivir monthly_STEP 61



Monsum		
STEP 61, Enero 1985	Desde	Hasta
DEM 90, resolución espacial 0.05	41.0	300

Fuente: Elaboración Propia

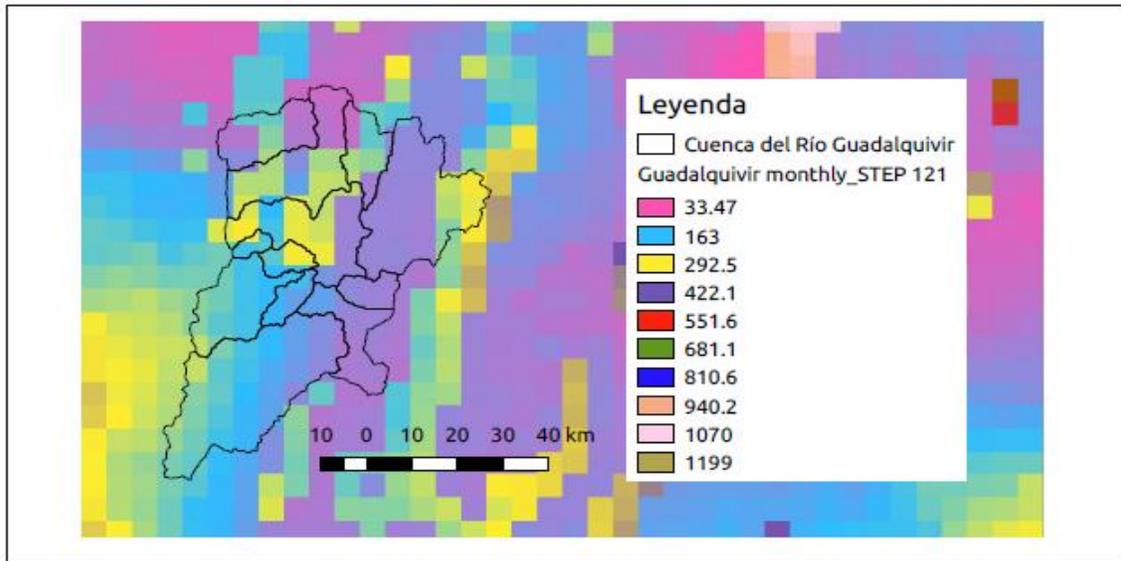
Figura N°49 Guadalquivir monthy_STEP 68



Monsum		
STEP 68, Agosto 1985	Desde	Hasta
DEM 90, resolución espacial 0.05	0.0	39.5

Fuente: Elaboración Propia

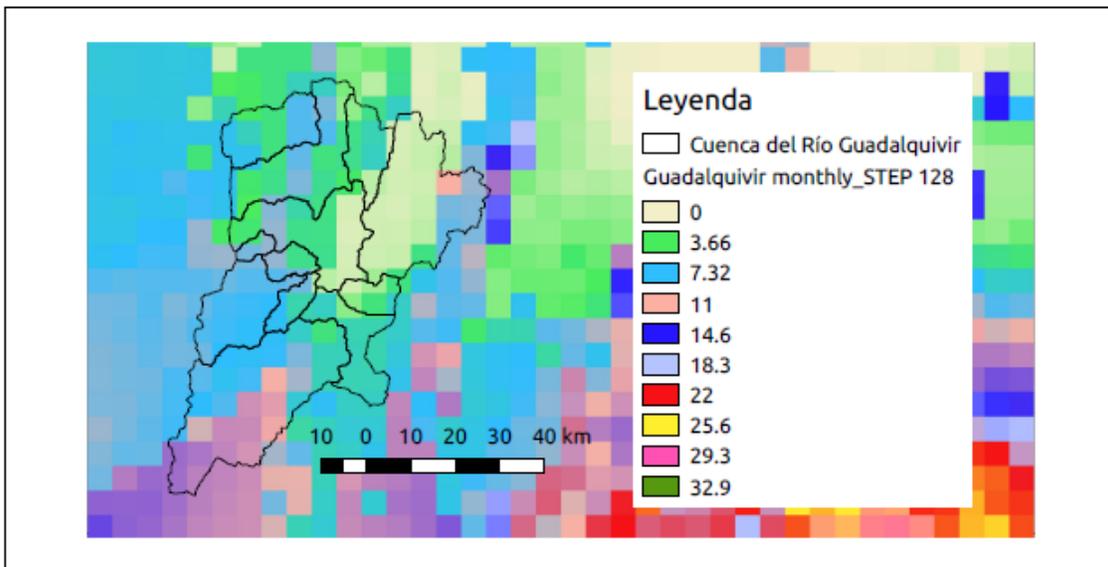
Figura N°50 Guadalquivir monthy_STEP 121



Monsum		
STEP 121, Enero 1990	Desde	Hasta
DEM 90, resolución espacial 0.05	33.5	292.5

Fuente: Elaboración Propia

Figura N°51 Guadalquivir monthy_STEP 128

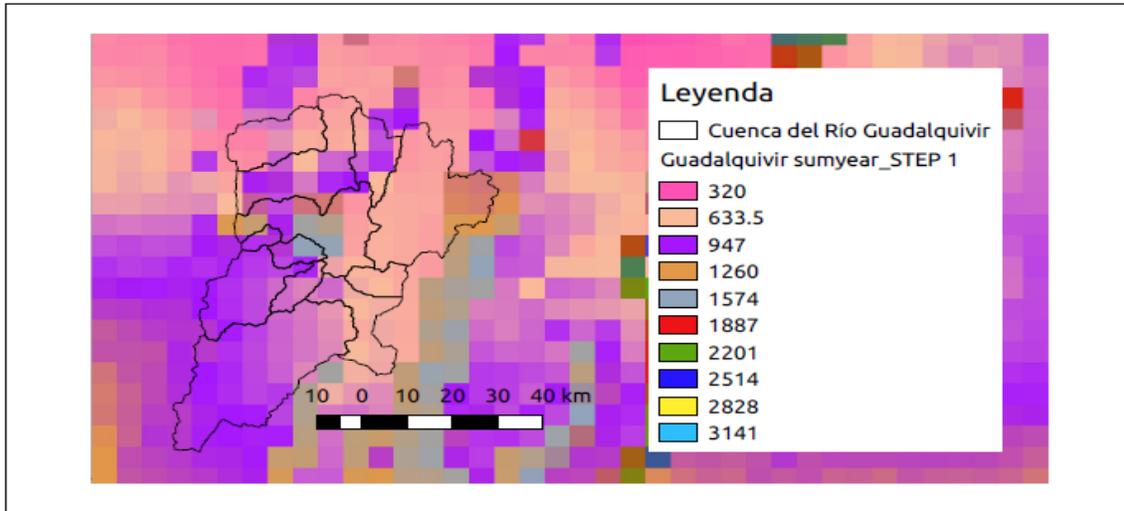


Monsum		
STEP 128, Agosto 1990	Desde	Hasta
DEM 90, resolución espacial 0.05	0.0	14.6

Fuente: Elaboración Propia

Para el promedio anual de los ensambles generados se tiene 13 pasos de tiempo. A continuación, se muestran como ejemplo los resultados en algunos pasos de tiempo:

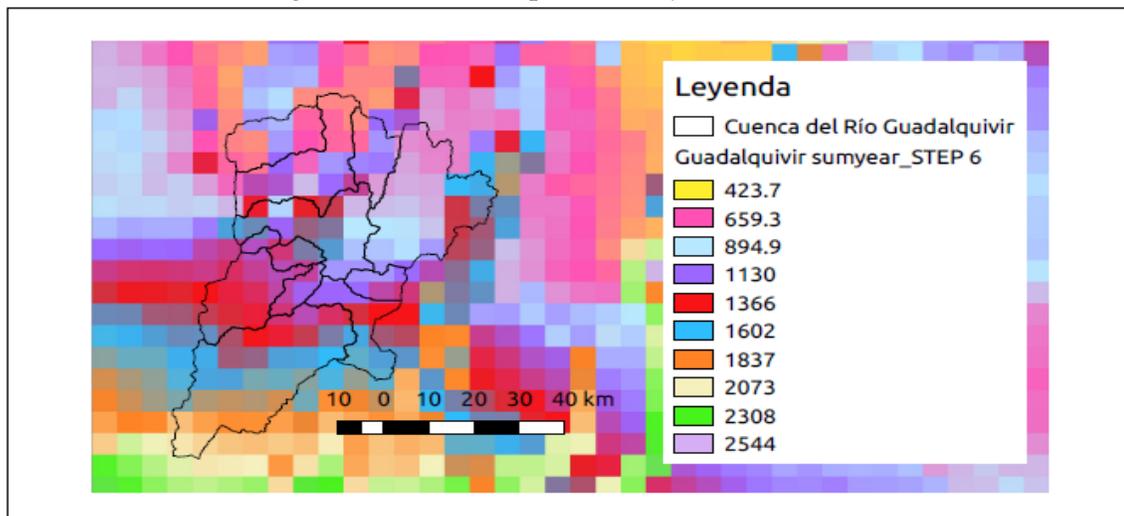
Figura N°52 Guadalquivir_Sumyear_STEP 1



Sumyear			
STEP 1, 1980	Desde	Hasta	
DEM 90, resolución espacial 0.05	320	1574	

Fuente: Elaboración Propia

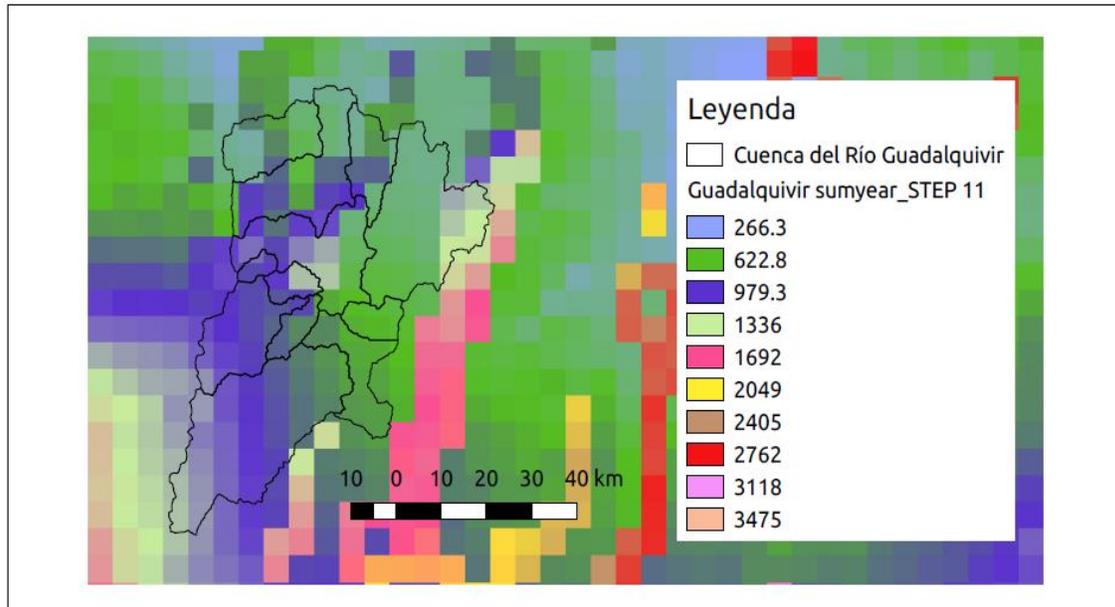
Figura N°53 Guadalquivir_Sumyear_STEP 6



Sumyear			
STEP 6, 1985	Desde	Hasta	
DEM 90, resolución espacial 0.05	423.7	2073	

Fuente: Elaboración Propia

Figura N°54 Guadalquivir_Sumyear_STEP 11



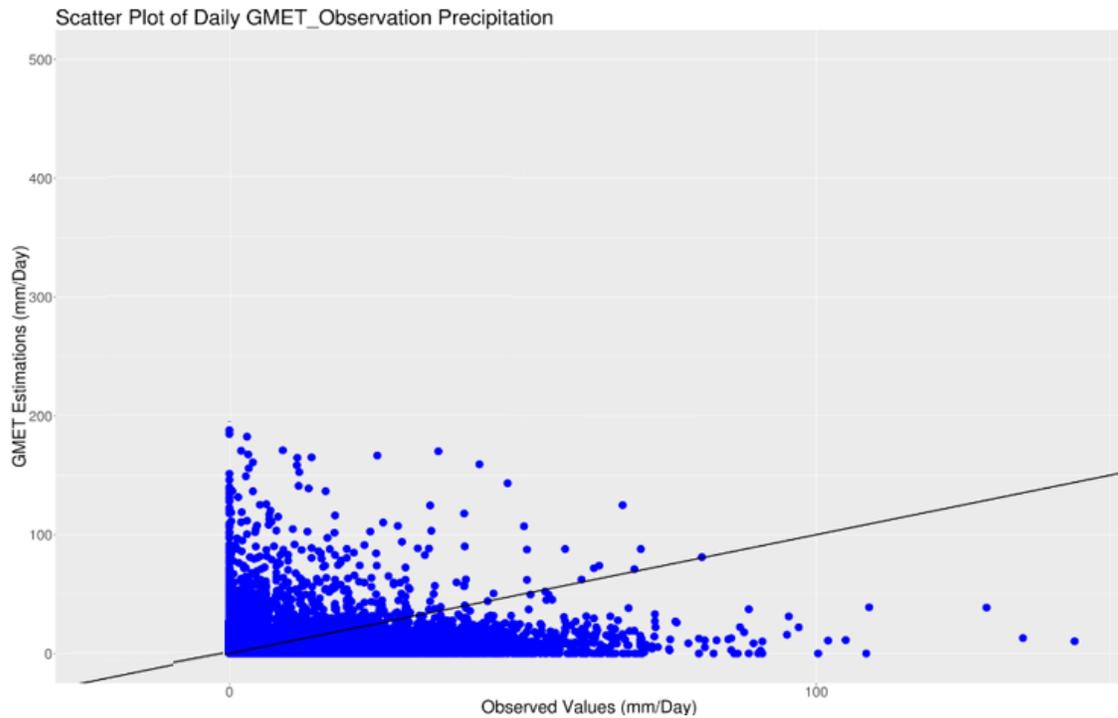
Sumyear			
STEP 11, 1990	Desde	Hasta	
DEM 90, resolución espacial 0.05	266.3	1336	

Fuente: Elaboración Propia

3.3.2.3. Validación de la grilla meteorológica GMET

La validación de la grilla se realiza de acuerdo a la evaluación del desempeño del modelo, usando un enfoque estadístico y de descomposición del error, la Figura N°55 muestra el diagrama de dispersión de los valores estimados por GMET respecto a los valores observados, el mismo sería perfectamente correcto cuando la línea de tendencia de los puntos estimados se alinea (estimados=observados).

Figura N°55 Diagrama de dispersión de GMET vs valores observados.



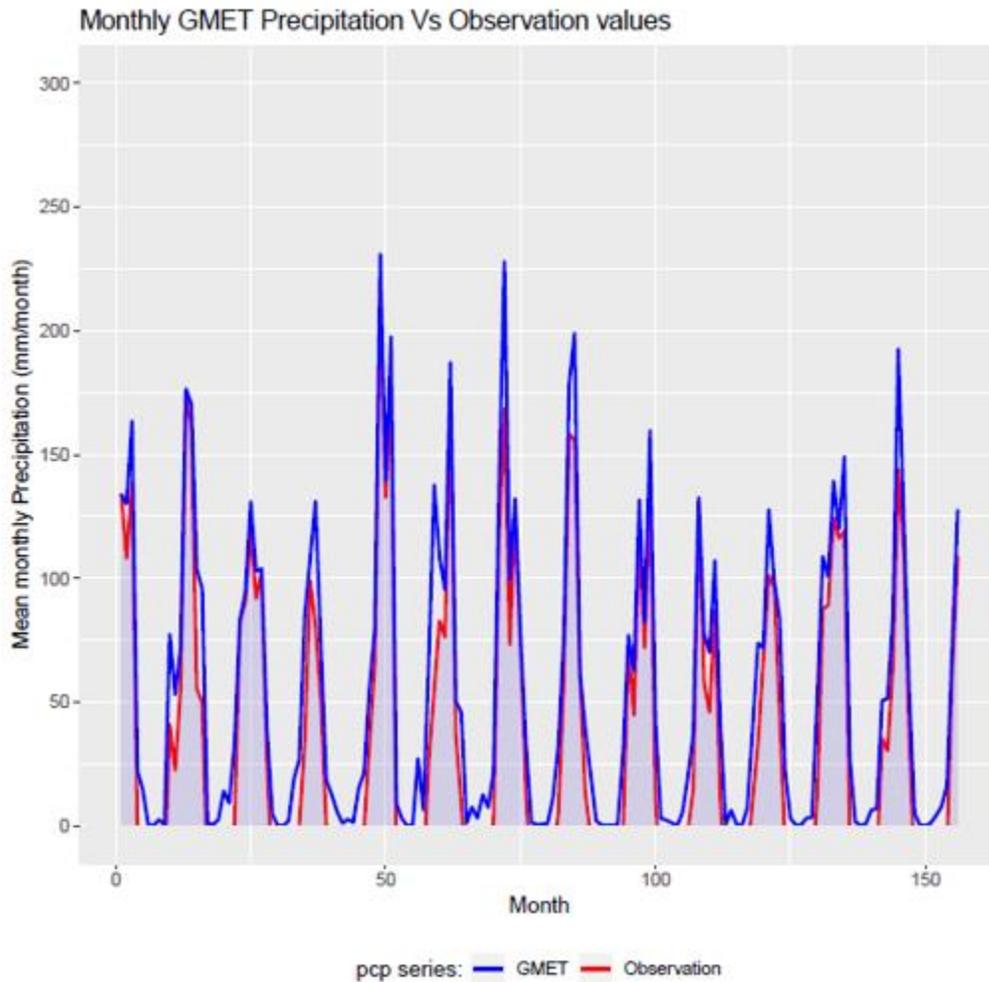
Fuente: Elaboración Propia

A continuación, se presentan los diferentes pasos de evaluación de la grilla generada:

3.3.2.3.1. Visualización de los datos

El primer paso en la comparación de los sets de datos es simplemente graficarlos para la escala temporal mensual, se preparó la Figura N°56 en la cual se observa que GMET reproduce los patrones mensuales de variación local de la red de datos con cierta sobreestimación en algunos meses del periodo.

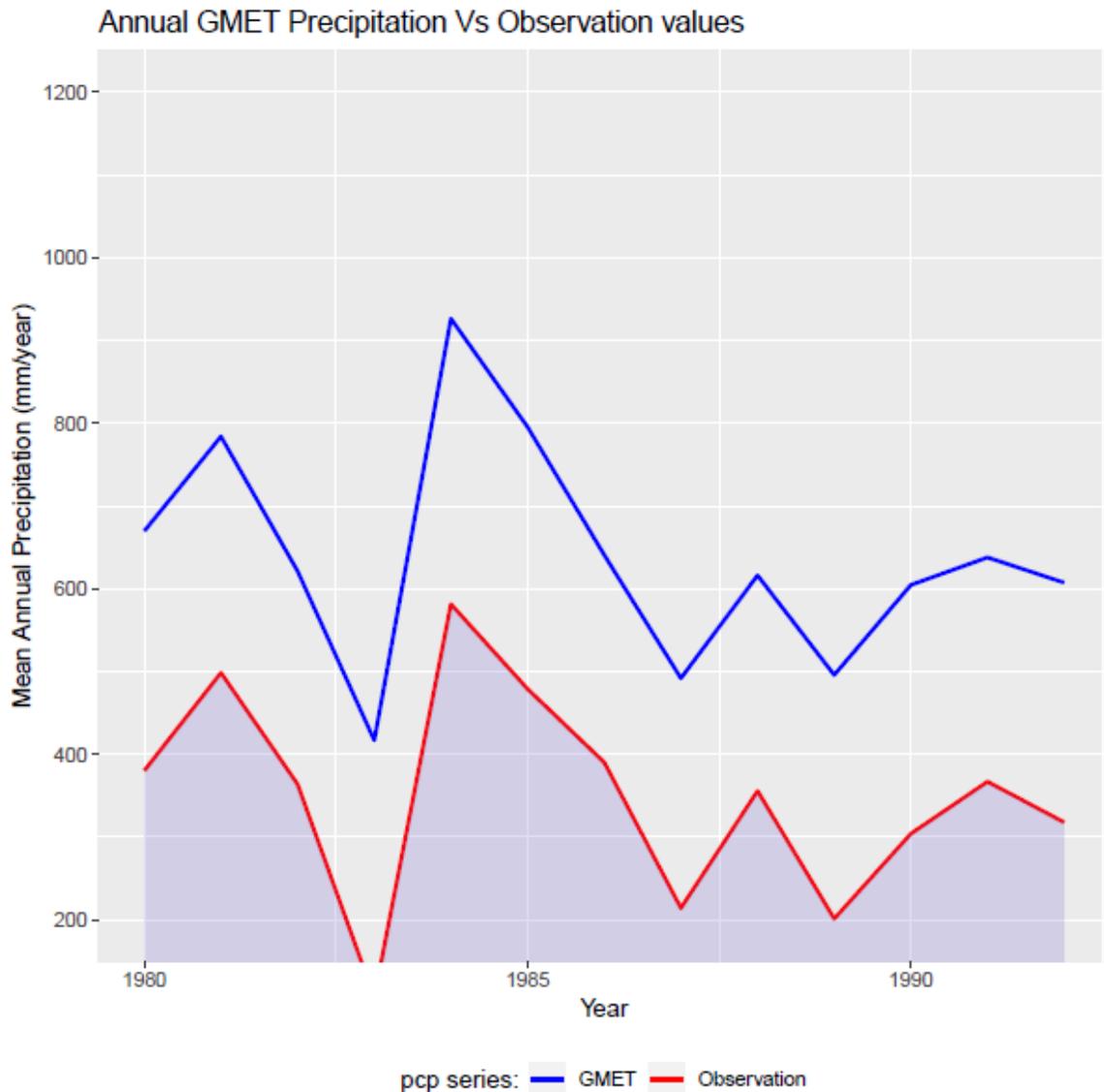
Figura N°56 Series Mensuales de GMET vs estaciones en la Cuenca del Río Guadalquivir



Fuente: Elaboración Propia

También se generó para este periodo la gráfica para datos anuales como se observa en la Figura N°57, GMET reproduce los patrones anuales de variación de los datos observados en la cuenca, pero con una sobrestimación respecto a los mismos en todo el periodo analizado.

Figura N°57 Series Anuales de GMET vs estaciones en la Cuenca del Río Guadalquivir



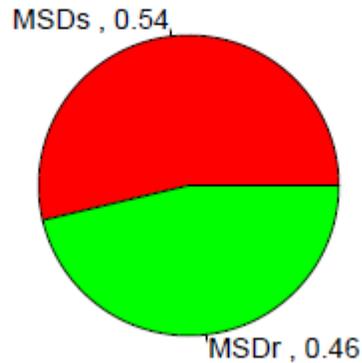
Fuente: Elaboración Propia

3.3.2.3.2. Descomposición del error

En la Figura N°58 observamos que, para este periodo analizado, el 54% del error asociado con GMET es causado por situaciones del sistema de estimación y que el 46% es causado por errores aleatorios. Calculando las porciones del error sistemático y aleatorio es posible identificar que el error sistemático es mayor que el componente del error aleatorio, esto significa que GMET debería ser ajustado basado en los patrones hidrológicos de la cuenca.

Figura N°58 Componentes del error sistemático y aleatorio

Systematic and random error components for the daily time series



Fuente: Elaboración Propia

3.3.3. Tercer Criterio de Selección

La selección de estaciones a emplearse en esta nueva generación de la Grilla meteorológica GMET para el periodo 1980-1992 se basó en la depuración de estaciones que no cuenten con información homogénea. Después de este análisis de homogeneidad realizado con el método de las Alternancias y de Helmert, se encontró que, de las 40 estaciones analizadas en su serie anual, solamente 30 pasaron dicho control. Las mismas se muestran en la Tabla N°16 a continuación:

Tabla N°16 Estaciones que pasaron el Análisis de Homogeneidad

N°	COD	Estación
1	GU001	Alto Cajas
2	GU002	Campanario
3	GU003	Cañas
4	GU004	Ce.Na. Vit
5	GU005	Ciudad-Tja
6	GU006	Coimata
7	GU010	Pinos Sud
8	GU012	San Andrés
9	GU013	San Jacinto Sud
10	GU015	Sella quebrada
11	GU016	Tarija Aeropuerto

12	GU017	Tomayapo Pueblo
13	GU018	Trancas
14	GU019	Tucumillas
15	GU025	Calderillas
16	GU023	Alizos
17	GU027	Canchasmayu
18	GU028	Chocloca
19	GU031	Copacabana Taxara
20	GU032	El Molino Tomayapo
21	GU034	Iscayachi San Antonio
22	GU035	Junacas
23	GU036	Ladera Centro
24	GU037	León Cancha
25	GU038	Padcaya
26	GU039	Pampa Redonda_h
27	GU040	Rejara
28	GU041	Rincón Cañas_h
29	GU044	San Agustín_tja
30	GU051	Tolomosa Grande

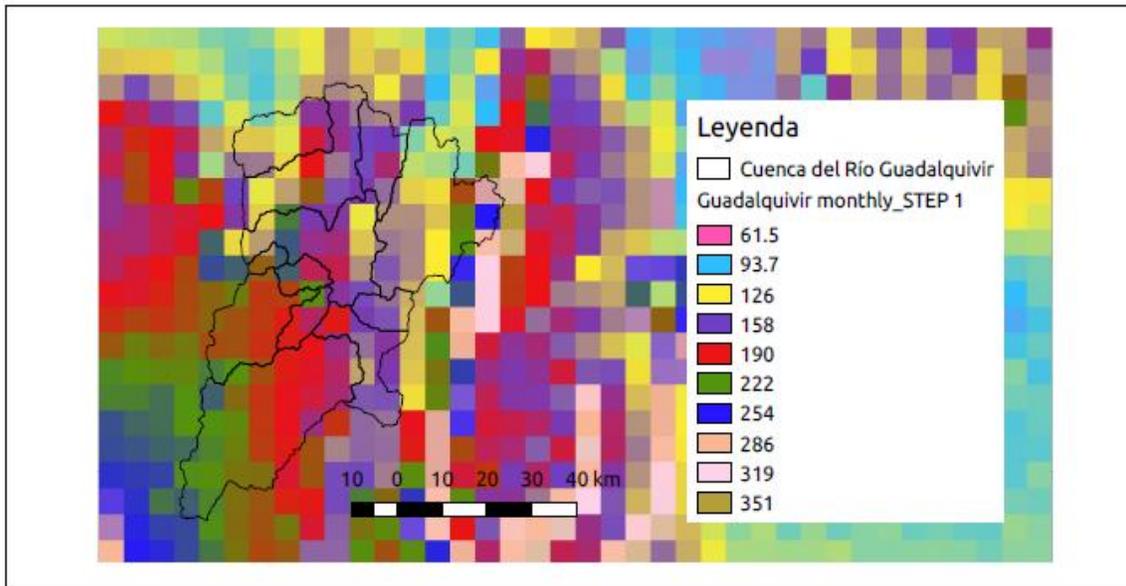
Fuente: Elaboración Propia

3.3.3.1. Aplicación de la Grilla Meteorológica GMET

Con la selección de las 30 estaciones mostradas en el punto 3.3.3 y realizado el control de calidad de los registros de las mismas, se corrió GMET siguiendo la metodología aplicada en el punto 3.3.1.2.

Posteriormente, se realizan los promedios correspondientes de los ensambles generados, en el caso de la acumulación mensual se tiene 156 pasos de tiempo para el periodo estudiado. En las figuras a continuación, se muestran como ejemplo los resultados en algunos pasos de tiempo:

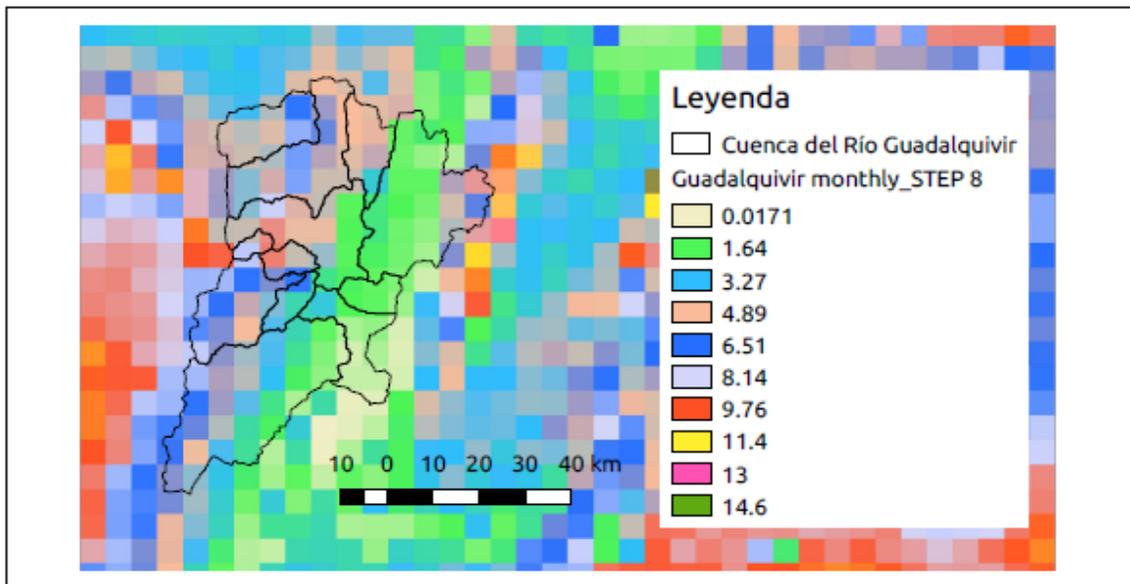
Figura N°59 Guadalquivir monthly_STEP 1



Monsum		
STEP 1, Enero 1980	Desde	Hasta
DEM 90, resolución espacial 0.05	93.7	286

Fuente: Elaboración Propia

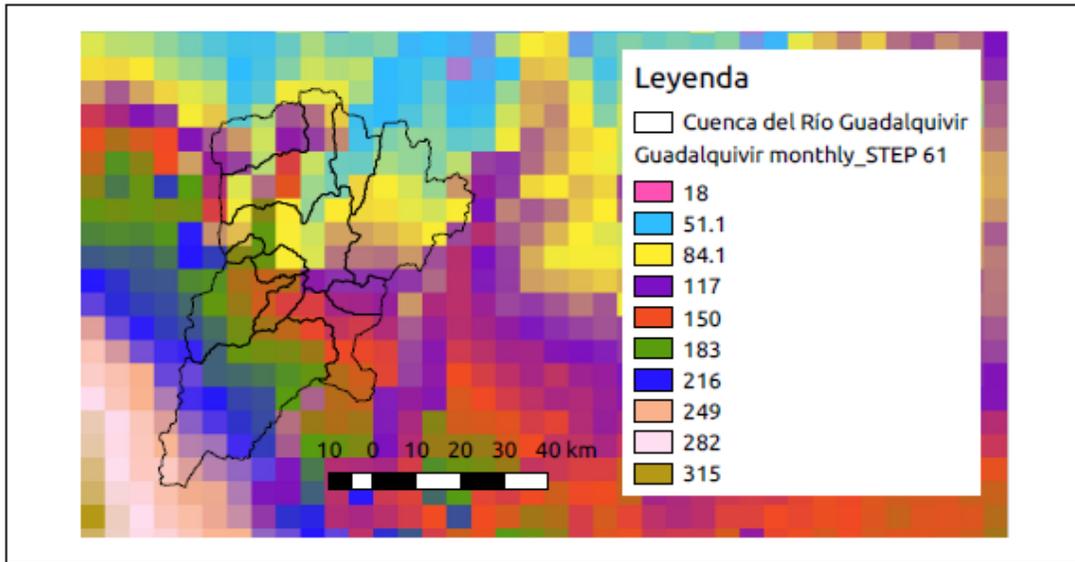
Figura N°60 Guadalquivir monthly_STEP 8



Monsum		
STEP 8, Agosto 1980	Desde	Hasta
DEM 90, resolución espacial 0.05	0.0	9.76

Fuente: Elaboración Propia

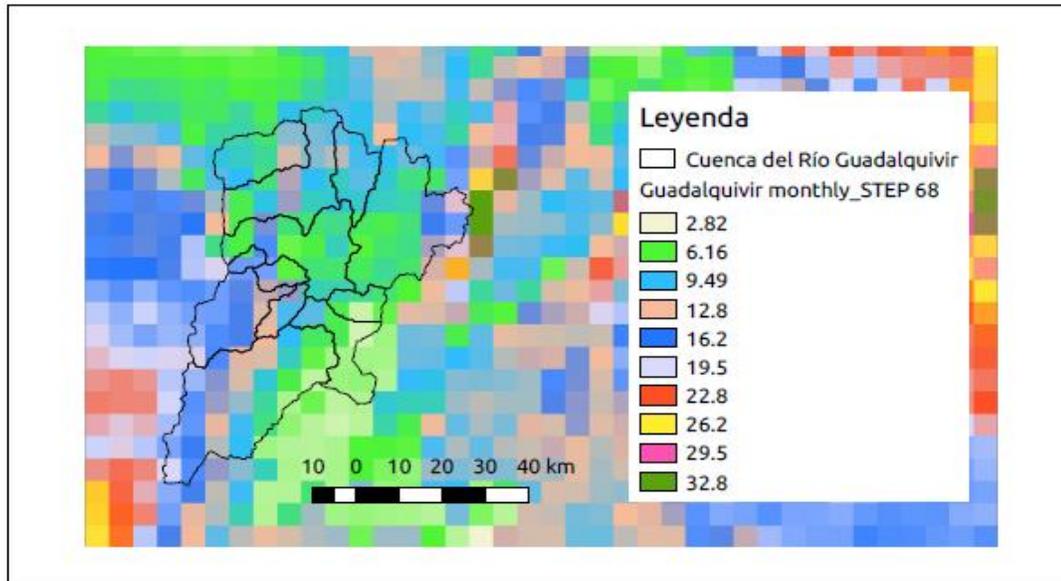
Figura N°61 Guadalquivir monthly_STEP 61



Monsum		
STEP 61, Enero 1985	Desde	Hasta
DEM 90, resolución espacial 0.05	51.1	249

Fuente: Elaboración Propia

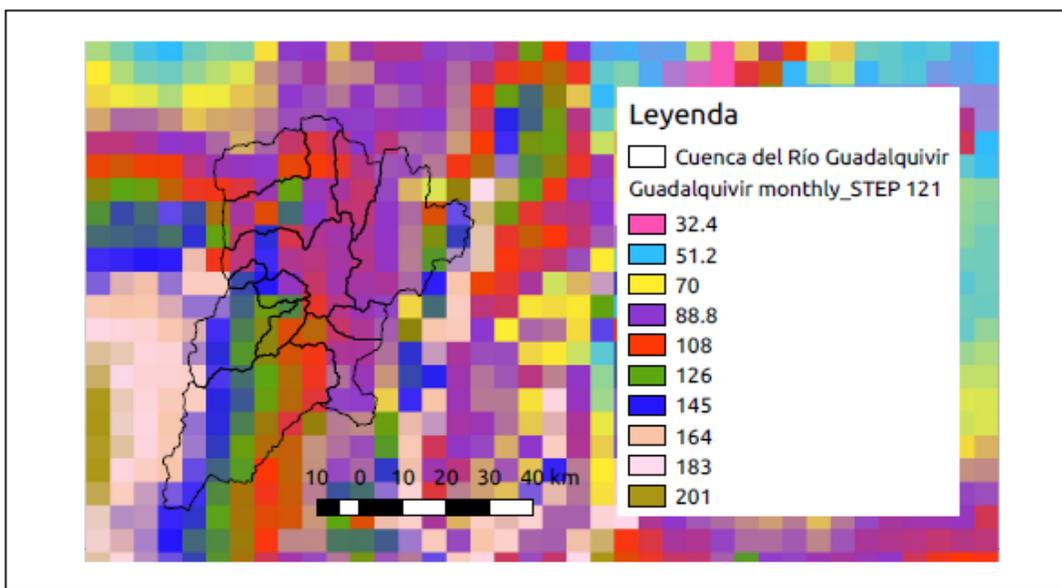
Figura N°62 Guadalquivir monthly_STEP 68



Monsum		
STEP 68, Agosto 1985	Desde	Hasta
DEM 90, resolución espacial 0.05	2.82	22.8

Fuente: Elaboración Propia

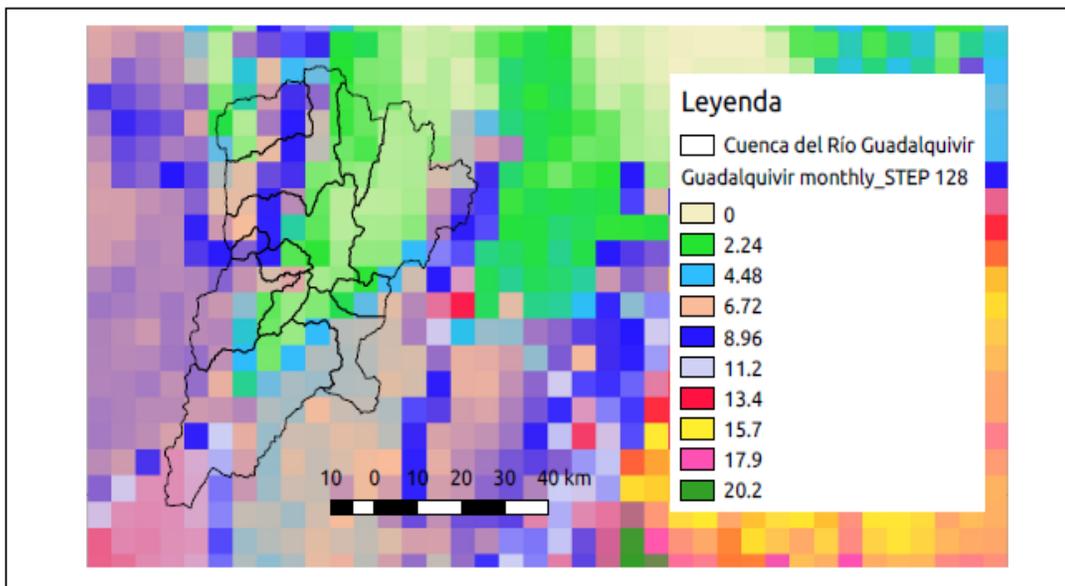
Figura N°63 Guadalquivir monthly_STEP 121



Monsum		
STEP 121, Enero 1990	Desde	Hasta
DEM 90, resolución espacial 0.05	51.2	164

Fuente: Elaboración Propia

Figura N°64 Guadalquivir monthly_STEP 128

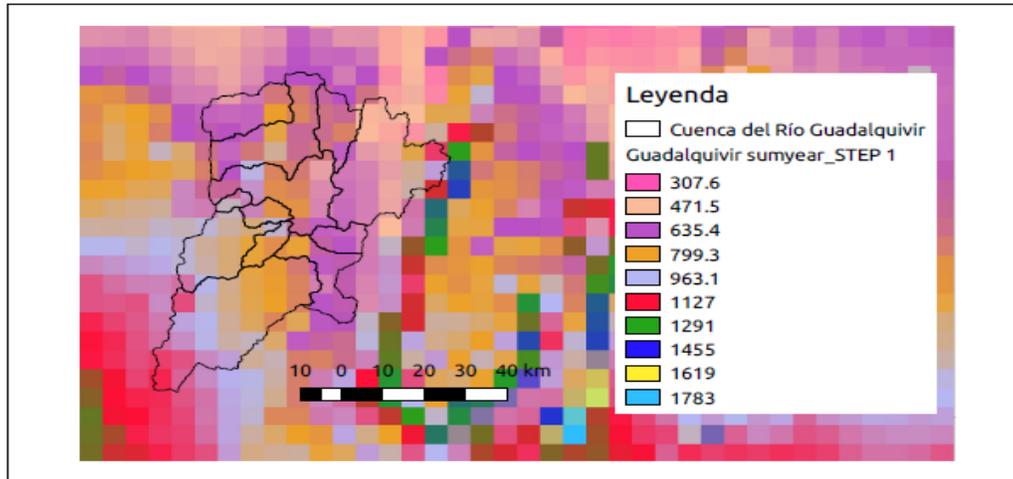


Monsum		
STEP 128, Agosto 1990	Desde	Hasta
DEM 90, resolución espacial 0.05	0.0	11.2

Fuente: Elaboración Propia

Para el promedio anual de los ensambles generados se tiene 13 pasos de tiempo. A continuación, se muestran como ejemplo los resultados en algunos pasos de tiempo:

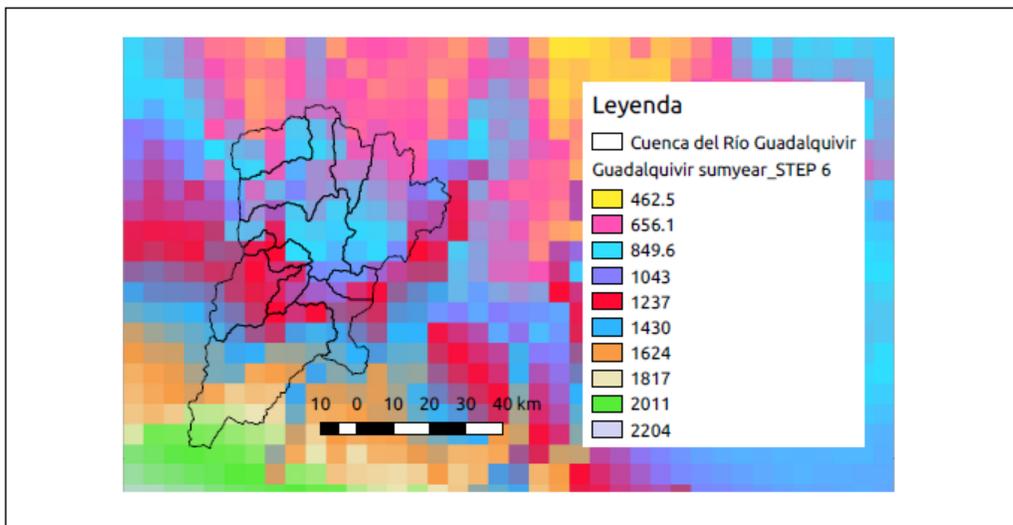
Figura N°65 Guadalquivir sumyear_STEP 1



Sumyear			
STEP 1, 1980		Desde	Hasta
DEM 90, resolución espacial 0.05		307.6	1127

Fuente: Elaboración Propia

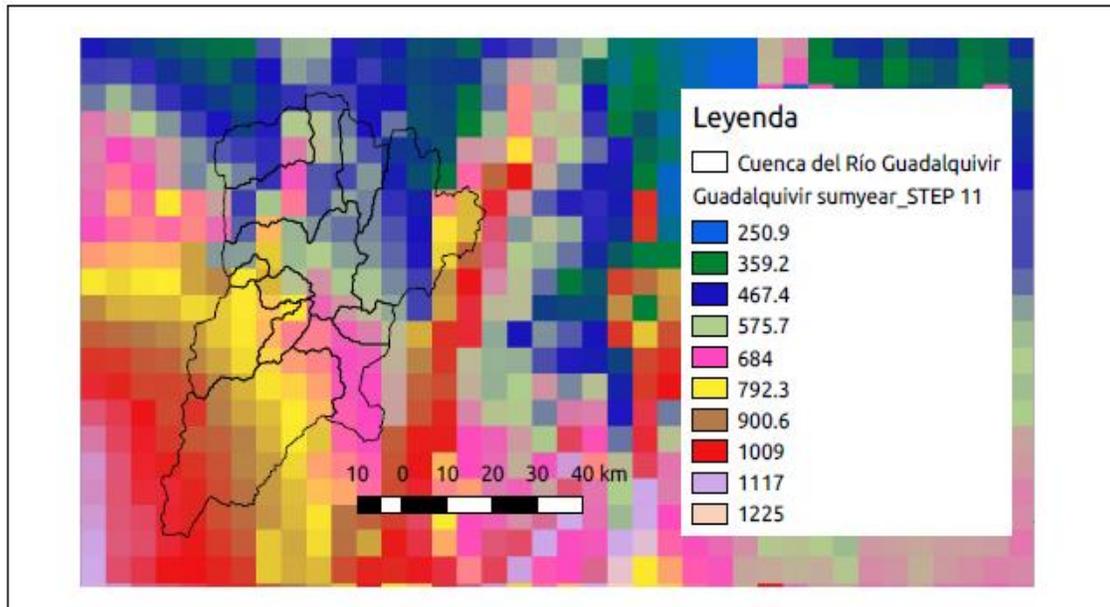
Figura N°66 Guadalquivir sumyear_STEP 6



Sumyear			
STEP 6, 1985		Desde	Hasta
DEM 90, resolución espacial 0.05		462.5	2011

Fuente: Elaboración Propia

Figura N°67 Guadalquivir sumyear_STEP 11



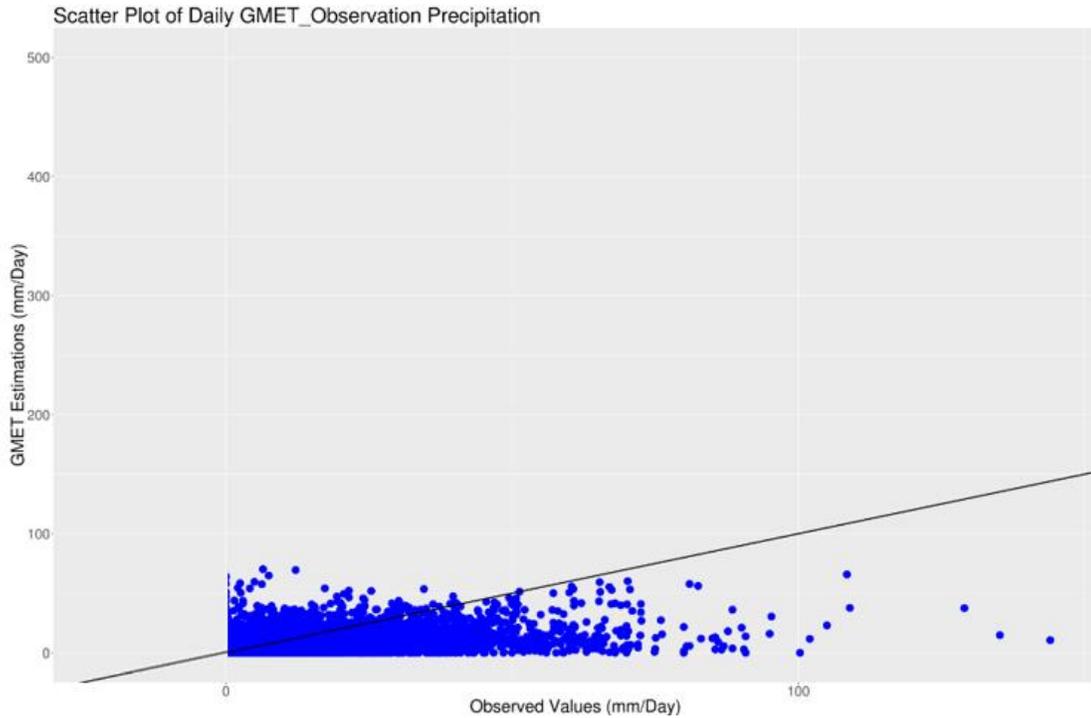
Sumyear			
STEP 11, 1990		Desde	Hasta
DEM 90, resolución espacial 0.05		359.2	1009

Fuente: Elaboración Propia

3.3.3.2. Validación de la grilla meteorológica GMET

La validación de la grilla se realiza de acuerdo a la evaluación del desempeño del modelo, usando un enfoque estadístico y de descomposición del error, la Figura N°68 muestra el diagrama de dispersión de los valores estimados por GMET respecto a los valores observados, el mismo sería perfectamente correcto cuando la línea de tendencia de los puntos estimados se alinea (estimados=observados).

Figura N°68 Diagrama de dispersión de GMET vs valores observados.



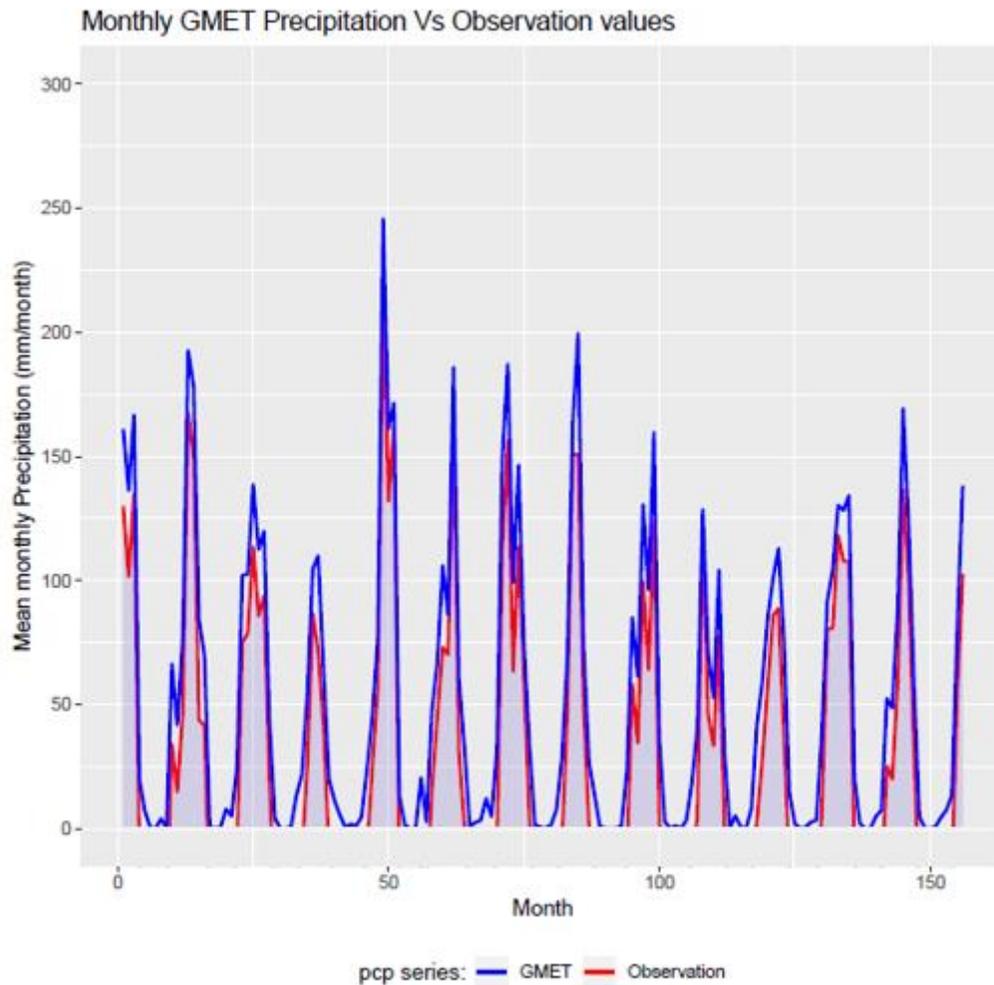
Fuente: Elaboración Propia

A continuación, se presentan los diferentes pasos de evaluación de la grilla generada:

3.3.3.2.1. Visualización de los datos

El primer paso en la comparación de los sets de datos es simplemente graficarlos para la escala temporal mensual, se preparó la Figura N°69 en la cual se observa que GMET reproduce los patrones mensuales de variación local de la red de datos con cierta sobreestimación en algunos meses del periodo.

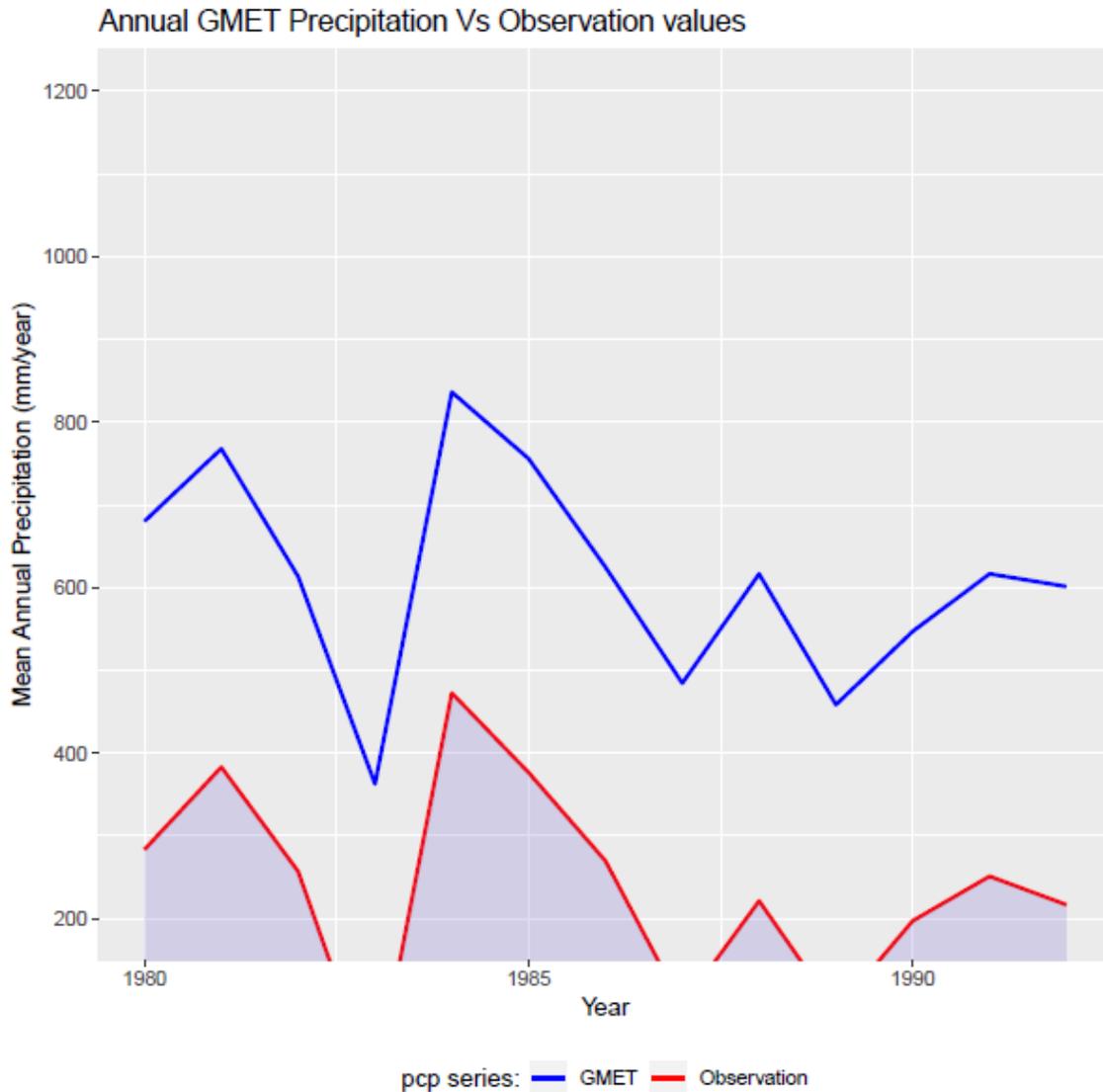
Figura N°69 Series Mensuales de GMET vs estaciones en la Cuenca del Río Guadalquivir



Fuente: Elaboración Propia

Además, se generó la gráfica para datos anuales del periodo analizado como se observa en la Figura N°70, GMET reproduce los patrones anuales de variación de los datos observados en la cuenca, pero presenta una sobrestimación respecto a los datos observados en el periodo estudiado.

Figura N°70 Series Anuales de GMET vs estaciones en la Cuenca del Río Guadalquivir



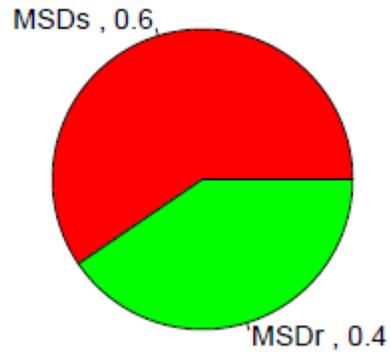
Fuente: Elaboración Propia

3.3.3.2.2. Descomposición del error

En la Figura N°71 observamos que, para el periodo 1980-1992 con las 30 estaciones empleadas, el 60% del error asociado con GMET es causado por situaciones del sistema de estimación y que el 40% es causado por errores aleatorios, analizando las porciones del error sistemático y aleatorio es posible identificar que el error sistemático es mayor que el componente del error aleatorio esto significa que GMET debería ser ajustado basado en los patrones hidrológicos de la cuenca.

Figura N°71 Componentes del error sistemático y aleatorio

Systematic and random error components for the daily time series



Fuente: Elaboración Propia

CAPÍTULO IV

ANÁLISIS DE RESULTADOS

En la Cuenca del Río Guadalquivir las precipitaciones tienen un régimen monomodal, es decir que tiene solamente un máximo bien definido en el año. En esta zona las lluvias se inician de diciembre a marzo, con precipitaciones máximas en enero y febrero; la estación seca es en invierno, como mínimo de mayo a julio, y dos períodos de transición separan estas dos épocas, uno en abril y otro de septiembre a noviembre.

Después de analizar diferentes criterios para seleccionar las estaciones empleadas para cada una de las corridas; GMET obtiene un producto final con una resolución espacial de 0.05° (celda) a escala de tiempo diario para el periodo estudiado.

Para la generación del ensamble (rango de resultados producidos por GMET), se usaron campos aleatorios correlacionados espacialmente (SCRFs por sus siglas en inglés – Spatially Correlated Random Fields). El método para generar los SCRF genera números progresivamente para cada punto de la grilla, condicionando los números aleatorios de acuerdo con los números generados previamente. El SCRF se generó para cada día de forma independiente para la variable precipitación, un resultado importante de uso de SCRF para generar los valores puntuales es que permite que sucedan eventos más extremos que las observaciones en secciones de la grilla donde no hay observaciones. Este es un punto clave con respecto a cualquier esquema basado en interpolaciones en que cada punto de la grilla es limitado por el valor máximo de las observaciones, lo cual no siempre es correcto y peor aún en terrenos con bastante influencia orográfica.

Con el promedio de los 30 miembros de ensamble obtenido a paso de tiempo diario, se genera las grillas a escala de tiempo mensual y anual para el periodo estudiado también con una resolución espacial de 0.05° .

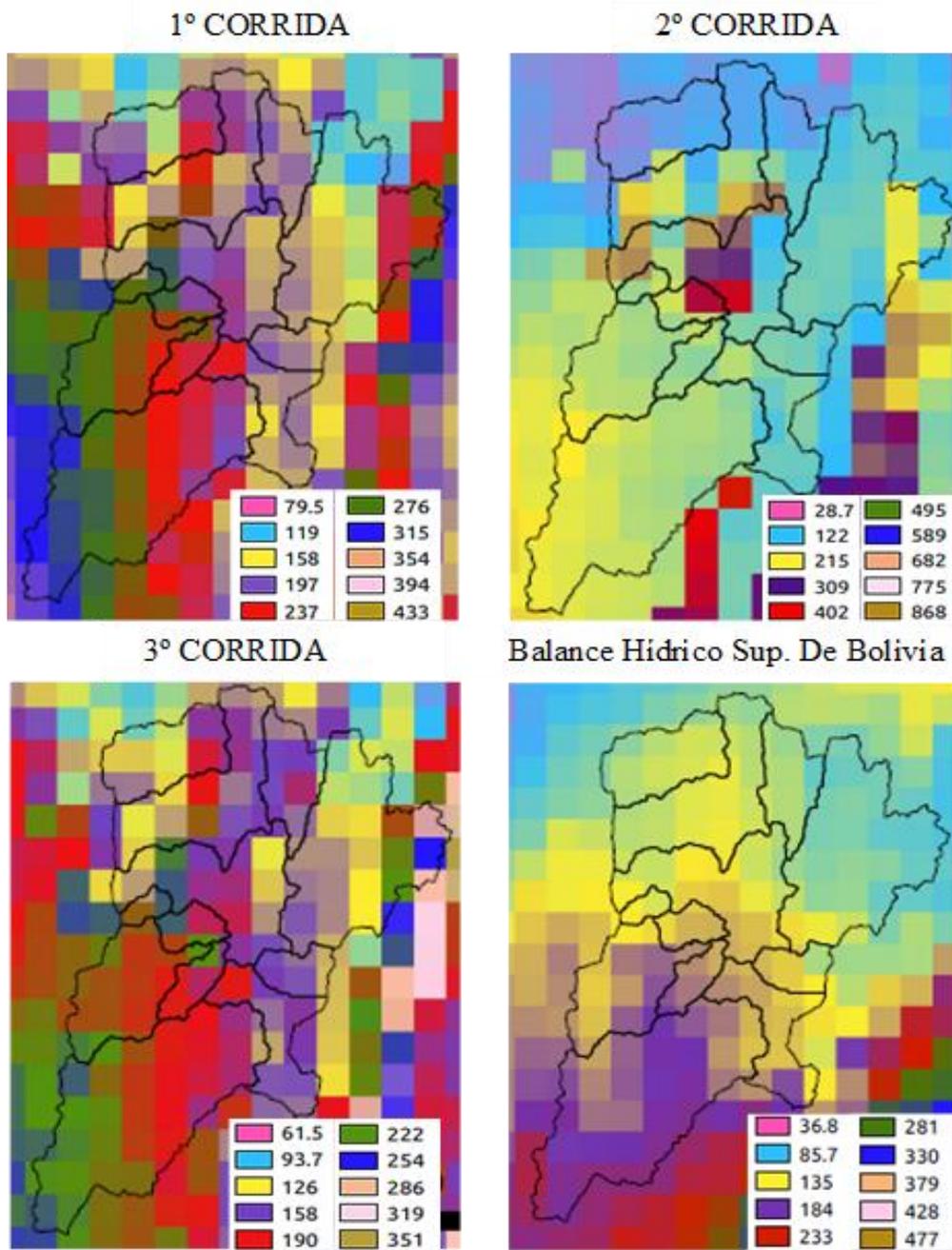
A continuación, se presenta un análisis del resultado de la grilla obtenida con GMET en cada una de las diferentes corridas realizadas para algunos pasos de tiempo como ser enero y agosto del periodo en estudio:

- Enero 1980

Los resultados producidos por GMET para las tres corridas con diferente criterio de selección de estaciones para enero de 1980 (Figura N°72) muestran una variación en los mismos ya que si se observan las precipitaciones obtenidas en la primera corrida con relación a la tercera, se puede percibir que la variación es menor si se comparara los mismos con los resultados de la segunda corrida, entonces, se observa que la variación es mayor ya que se tiene precipitaciones más elevadas.

Podemos observar también que en la primera corrida existe una variabilidad creciente en dirección de norte a sur, siendo más húmedo en el sur de la cuenca con precipitaciones que varían desde 197 hasta 315mm., mientras que al norte varía desde 119 hasta 237mm. En la tercera corrida se puede apreciar de igual manera variabilidad creciente de norte a sur, pero con precipitaciones de menor magnitud ya que varían desde 93.7 hasta 190mm en el norte y desde 190 hasta 254mm en el sur. Por otro lado, se puede observar que en la segunda corrida los resultados difieren ya que si bien se observa una variación creciente de norte a sur con magnitudes que oscilan entre 122 y 215mm., se aprecia que la parte central de la cuenca sería la más húmeda con precipitaciones que alcanzan los 402mm.

Figura N°72 Resultados Enero 1980



	Desde	Hasta
1° Corrida	119	315
2° Corrida	122	402
3° Corrida	93.7	254
B.H.S.B	85.7	233

Fuente: Elaboración Propia

Si se realiza un análisis visual de los resultados obtenidos en las tres corridas con los del Balance Hídrico Superficial de Bolivia, se podría decir que no hay similitud entre estos, ya que la variabilidad de las precipitaciones en la cuenca es distinta. Sin embargo, si se realiza un análisis de las magnitudes de las precipitaciones podemos observar que la tercera corrida presenta menor variación con los resultados del B.H.S.B., ya que el mismo tiene precipitaciones que varían desde 85.7 hasta 184mm. en el norte y desde 184 hasta 233mm., en el sur.

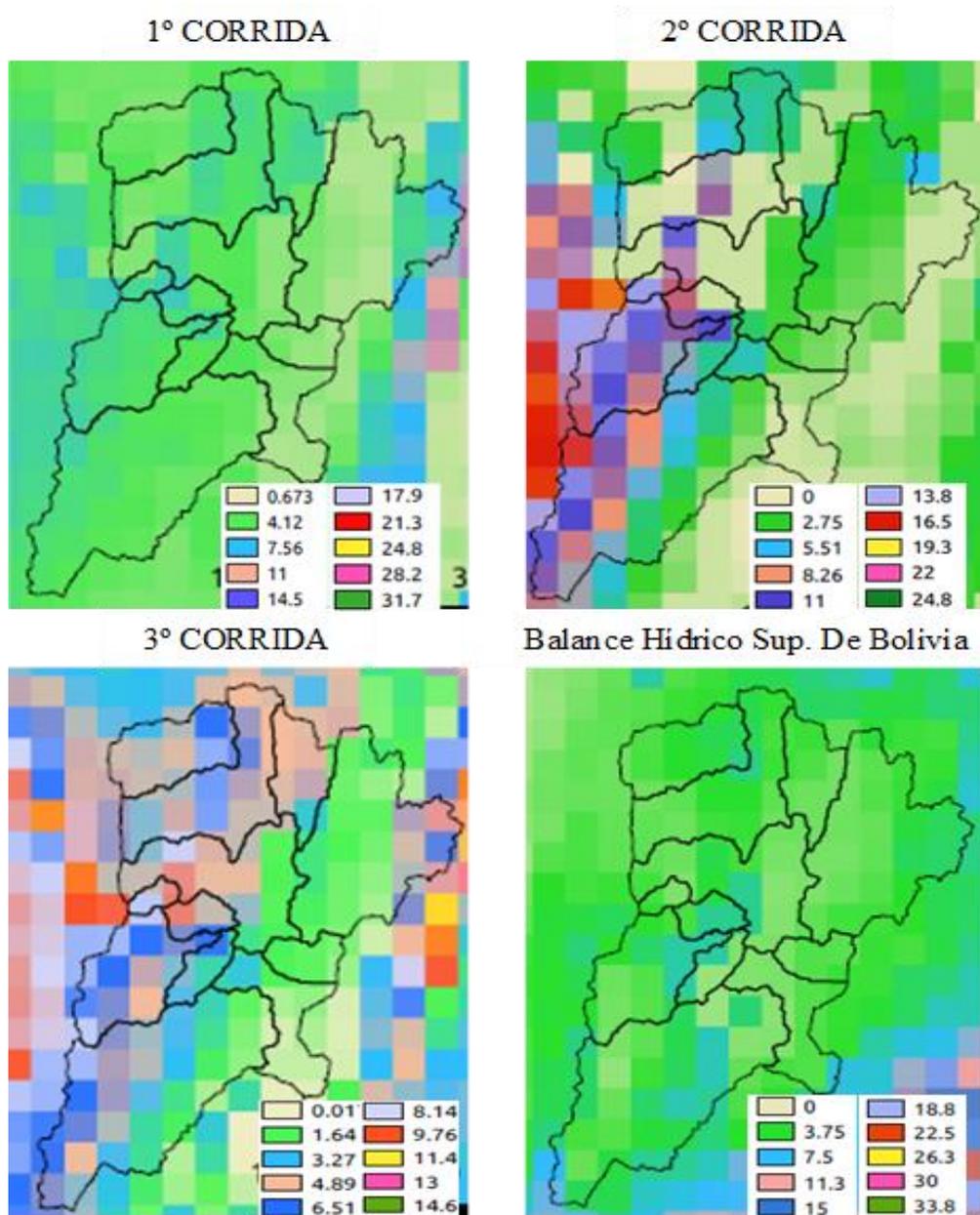
- Agosto 1980

Los resultados producidos por GMET en las tres corridas para agosto de 1980 (Figura N°73) muestran una variación en los mismos, ya que, si se observa las precipitaciones obtenidas en la primera corrida con relación a la tercera, se puede percibir que la variación es menor si se comparara los mismos con los resultados de la segunda corrida. Se observa así que la variación es mayor, ya que se tiene precipitaciones más elevadas al oeste de la cuenca.

Podemos observar también, que en la segunda corrida existe una variabilidad creciente de este a oeste, siendo más húmedo en esta época del año el SW de la cuenca con precipitaciones que varían desde 8.26 hasta 16.5mm., mientras que al este, varían desde 0.0 hasta 5.51mm. Por otro lado, se puede observar que en la tercera corrida la variación es de igual manera creciente de este a oeste, pero con precipitaciones menores a las presentadas anteriormente, ya que varían entre 4.89 y 9.76mm., siendo más húmedo el oeste de la cuenca en este mes del año mientras que al este varían desde 0.0 hasta 3.27mm.

Al comparar estos resultados obtenidos en las tres corridas con los del Balance Hídrico Superficial de Bolivia, después de realizar un análisis visual, se podría decir que se encuentra similitud con la primera corrida ya que la variabilidad de las precipitaciones es parecida; pero, si se analiza las magnitudes de las mismas podemos decir que existe menor variación con los resultados obtenidos en la tercera corrida.

Figura N°73 Resultados Agosto 1980



	Desde	Hasta
1° Corrida	0.67	11.0
2° Corrida	0.00	16.5
3° Corrida	0.00	9.76
B.H.S.B	0.00	7.5

Fuente: Elaboración Propia

- Enero 1985

Los resultados producidos por GMET para las tres corridas para el mes analizado (Figura N°74) muestran una variabilidad similar en la primera y tercera corrida donde se puede observar que el sur de la cuenca sería más húmedo, mientras que en la segunda las precipitaciones más elevadas se presentan en la parte central de la cuenca.

Podemos observar también que en la primera corrida existe una variabilidad creciente en dirección de norte a sur, siendo más húmedo en el sur de la cuenca con precipitaciones que varían desde 226 hasta 309mm., mientras que en el norte varía desde 59.5 hasta 184 mm, similar variabilidad, pero con precipitaciones menores se pueden observar en la tercera corrida, puesto que el sur sería más húmedo con precipitaciones que varían desde 150 hasta 249mm. a diferencia del norte en donde varía desde 51.1 hasta 117mm. Por otro lado, en la segunda corrida los resultados difieren ya que si bien se observa una variación creciente de norte a sur con magnitudes que oscilan entre 41.0 y 213mm., la parte central de la cuenca sería la más húmeda presentando precipitaciones que alcanzan los 300mm.

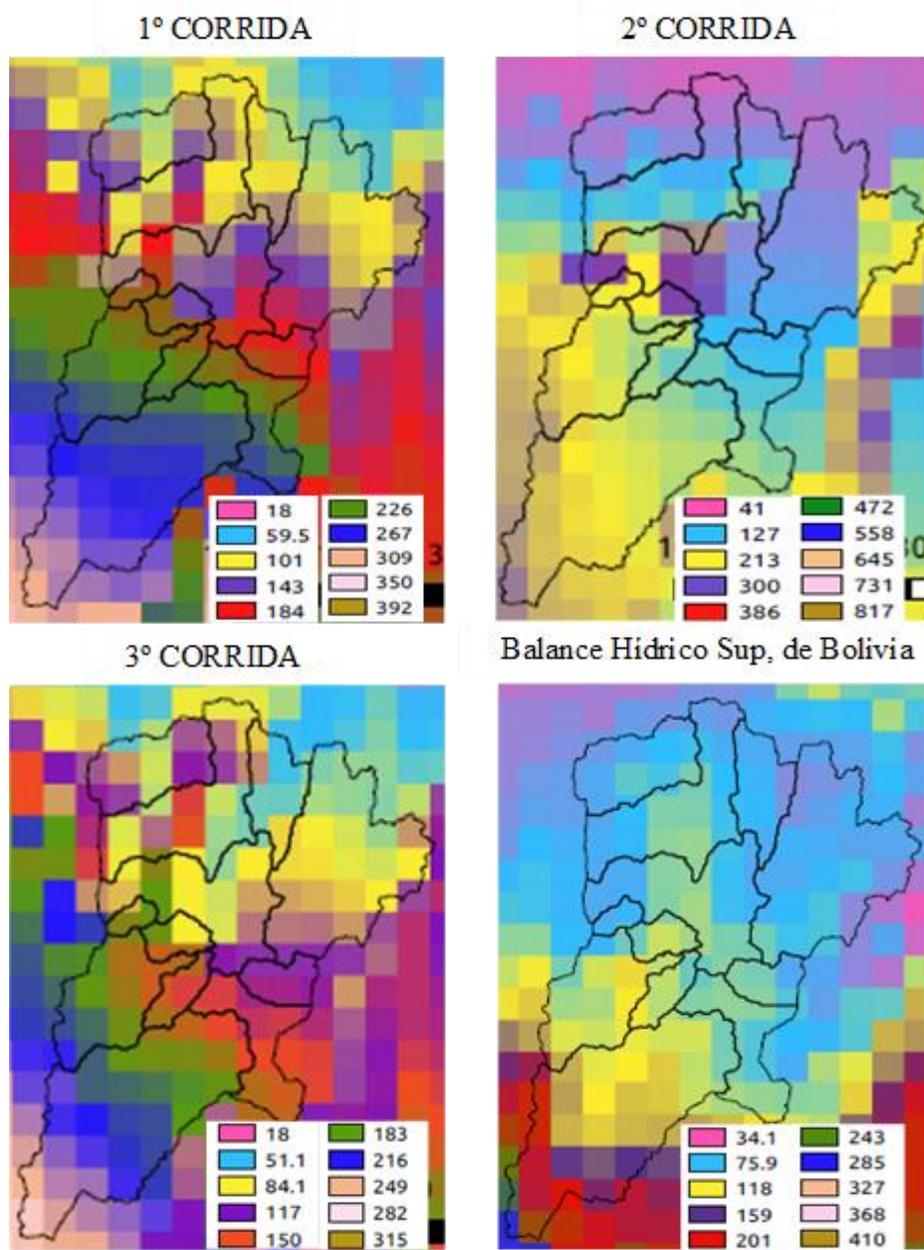
Comparando estos resultados obtenidos en las tres corridas realizadas con los del Balance Hídrico Superficial de Bolivia y realizando un análisis visual de los mismos podríamos decir que no hay similitud entre ellos puesto que como se puede observar en la figura, la variabilidad de las precipitaciones en la cuenca es distinta. Sin embargo, realizando el análisis de sus magnitudes se podría decir que los resultados más cercanos son los obtenidos en la tercera corrida ya que en el B.H.S.B se tiene precipitaciones que varían desde 31.1 hasta 285mm.

- Agosto 1985

Los resultados producidos por GMET para las tres corridas para este mes en análisis (Figura N°75) muestran diferencias en las que se puede observar como la variabilidad de las precipitaciones en la cuenca son distintas para cada una de las corridas.

Se observa que, en la primera corrida, existe una variabilidad creciente de este a oeste, siendo más húmedo el oeste de la cuenca con precipitaciones que varían desde 21.7 hasta 35.9mm., mientras que al este solo alcanzan los 14.6mm. En la segunda corrida, se puede observar que los resultados difieren ya que las precipitaciones varían desde 0.0 hasta 39.5mm., siendo

Figura N°74 Resultados Enero 1985



	Desde	Hasta
1° Corrida	59.5	309
2° Corrida	41.0	300
3° Corrida	51.1	249
B.H.S.B	34.1	285

Fuente: Elaboración Propia

más húmeda la parte central de la cuenca donde las precipitaciones alcanzan magnitudes que varían desde 29.6 hasta 39.5mm. Por otro lado, se puede observar que en la tercera corrida las precipitaciones son variables a lo largo de la cuenca y van desde 2.82 hasta 22.8mm., siendo más húmedo el suroeste de la cuenca.

Si se realiza un análisis visual de los resultados obtenidos en las tres corridas realizadas con los del Balance Hídrico Superficial de Bolivia se podría decir que no hay similitud en los mismos puesto que la variabilidad de las precipitaciones en la cuenca es distinta. El B.H.S.B presenta precipitaciones que varían desde 0.0 hasta 15.3mm., en el norte de la cuenca y alcanza los 30.5mm., en el sur siendo esta la zona más húmeda en este mes.

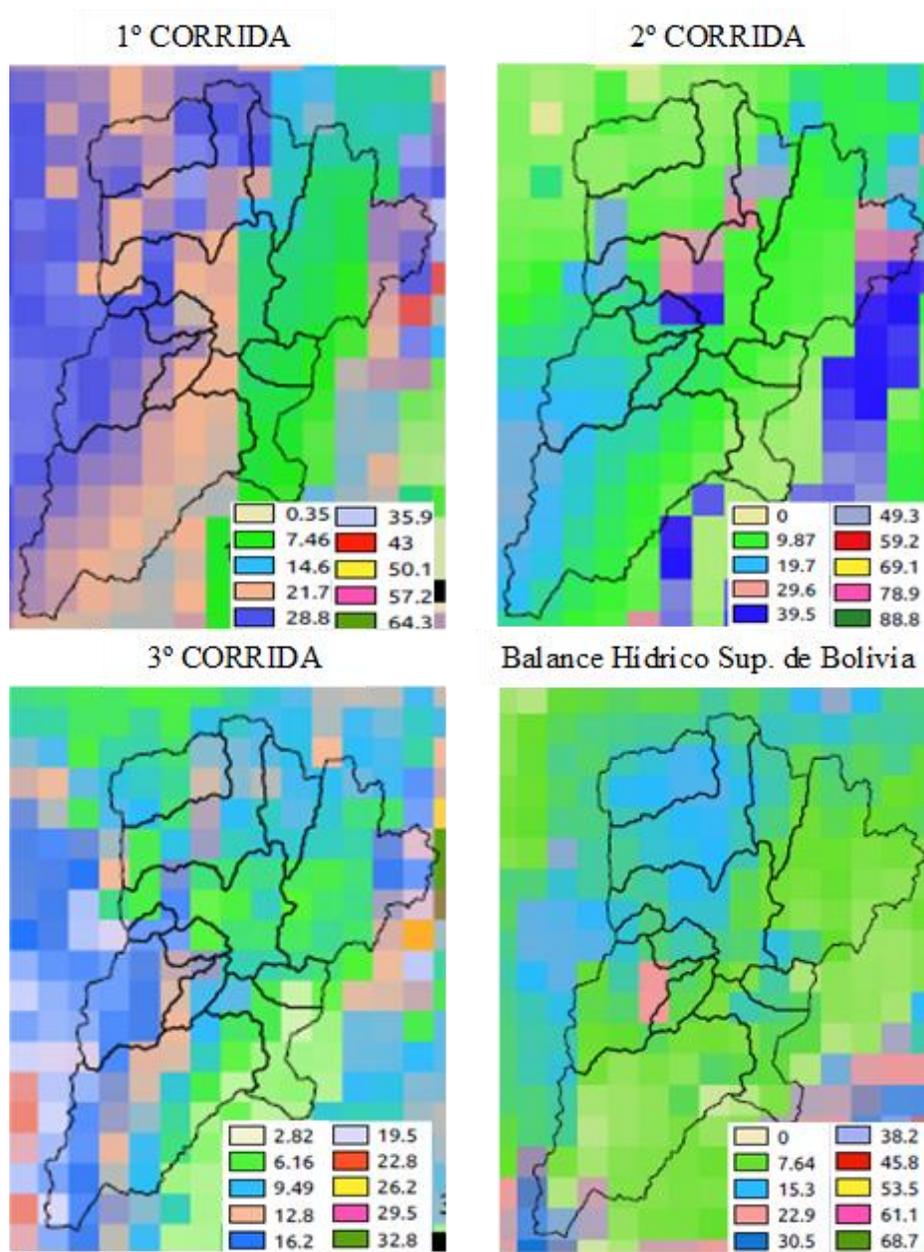
- Enero 1990

Los resultados producidos por GMET para las tres corridas con diferente criterio de selección de estaciones para enero de 1990 (Figura N°76) muestran diferencias ya que si de un análisis visual se observa similar variabilidad de las precipitaciones en la primer y tercer corrida en cuanto a magnitudes presentan diferencias.

Se observa que en la primera existe una variabilidad creciente de norte a sur, siendo más húmedo en el sur de la cuenca con precipitaciones que varían desde 154 hasta 239mm., mientras que al norte varía desde 68.6 hasta 182mm. En la tercera corrida, se puede observar igual variabilidad, pero con precipitaciones de menor magnitud, las precipitaciones más elevadas para esta corrida se presentan al sur de la cuenca con variaciones desde 108 hasta 164mm. Por otro lado, se ve que en la segunda corrida los resultados difieren porque si bien se observa una variación creciente de norte a sur con magnitudes que oscilan entre 33.5 y 163mm. Así, se concluye que la parte central de la cuenca sería la más húmeda porque presenta precipitaciones que alcanzan los 292.5mm.

Comparando los resultados obtenidos en las tres corridas realizadas con los del Balance Hídrico Nacional de Bolivia y después de realizar un análisis visual de los mismos se podría decir que no tienen similitud, basado en las observaciones hechas en la figura porque la variabilidad de las precipitaciones en la cuenca es distinta, observando que en el B.H.S.B se tiene que la variación de precipitación es creciente de este a oeste con magnitudes que oscilan

Figura N°75 Resultados Agosto 1985



	Desde	Hasta
1° Corrida	0.36	35.9
2° Corrida	0.00	39.5
3° Corrida	2.82	22.8
B.H.S.B	0.00	30.5

Fuente: Elaboración Propia

entre 32.6 y 211mm., siendo más húmedo el SW de la cuenca donde las precipitaciones alcanzan magnitudes que varían desde 122 hasta 211mm.

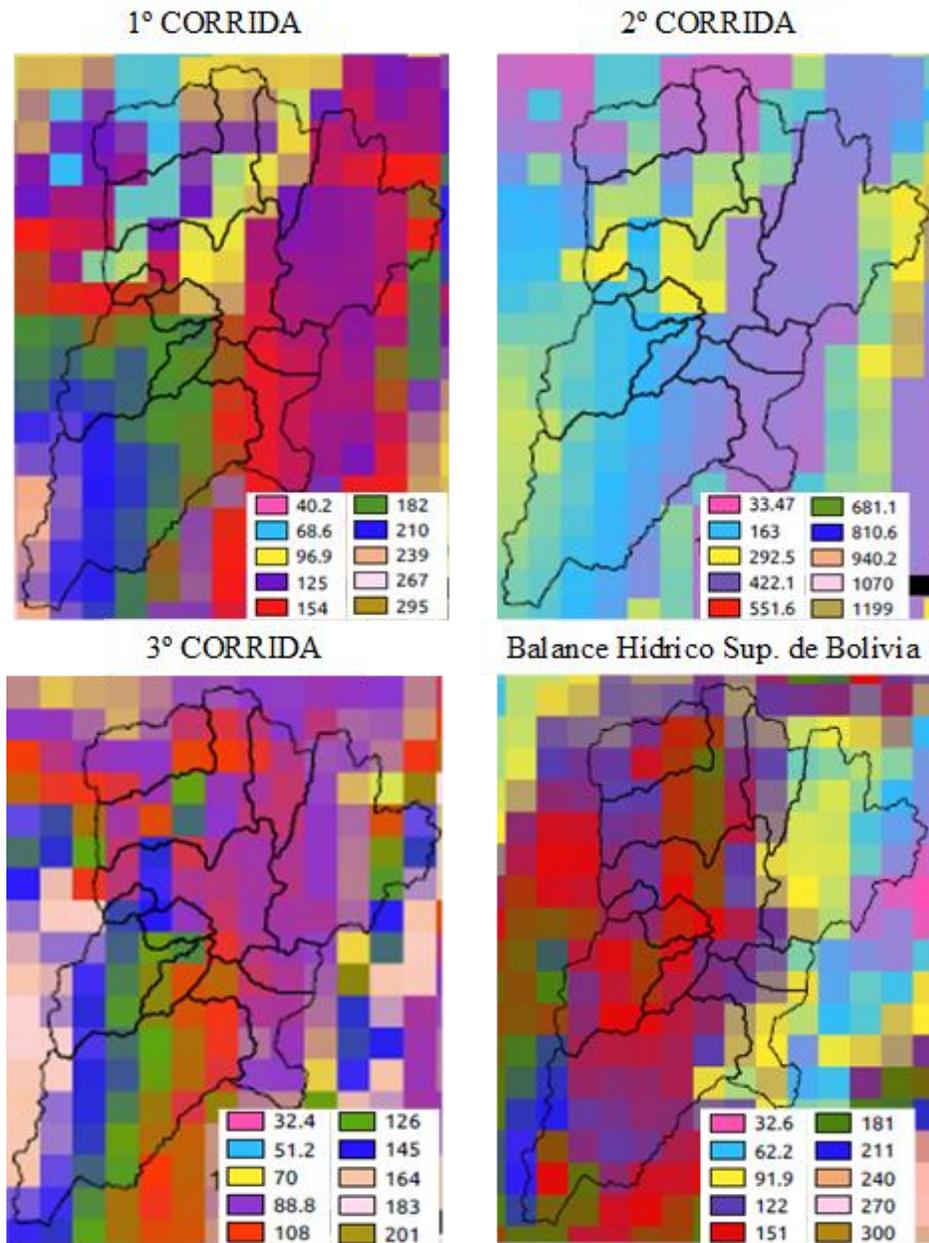
- Agosto 1990

Los resultados producidos por GMET en las tres corridas para agosto de 1990 (Figura N°77) muestran diferencias en la variabilidad de las precipitaciones en la cuenca que es distinta para cada una de las corridas.

Podemos observar también que en la primera corrida existe una variabilidad creciente de norte a sur, siendo más húmedo en el sur de la cuenca con precipitaciones que varían desde 5.75 hasta 15.3mm., y al norte varía desde 0.0 hasta 5.75mm. En la segunda corrida, se observa también variabilidad creciente de norte a sur por lo que las precipitaciones más elevadas se presentan al sur de la cuenca con magnitudes que oscilan entre 7.32 hasta 14.6mm., y al norte precipitaciones que varían desde 0.0 hasta 7.32mm. Por otro lado, se puede observar que, en la tercera corrida, los resultados difieren y se presentan precipitaciones menores a las obtenidas en las dos corridas previas ya que solo alcanzan magnitudes que varían desde 0.0 hasta 11.2mm., siendo más húmedo el sur de la cuenca.

Comparando estos resultados obtenidos en las tres corridas realizadas con los del Balance Hídrico Superficial de Bolivia y después de realizar un análisis visual se podría decir que no hay similitud en los mismos porque la variabilidad de las precipitaciones en la cuenca es distinta. Sin embargo, si se hace un análisis de las magnitudes, se observa que la tercera corrida presentaría resultados más cercanos ya que el B.H.S.B presenta precipitaciones con magnitudes que varían desde 0.0 hasta 12.7mm., en el sur de la cuenca siendo esta la zona más húmeda para este mes.

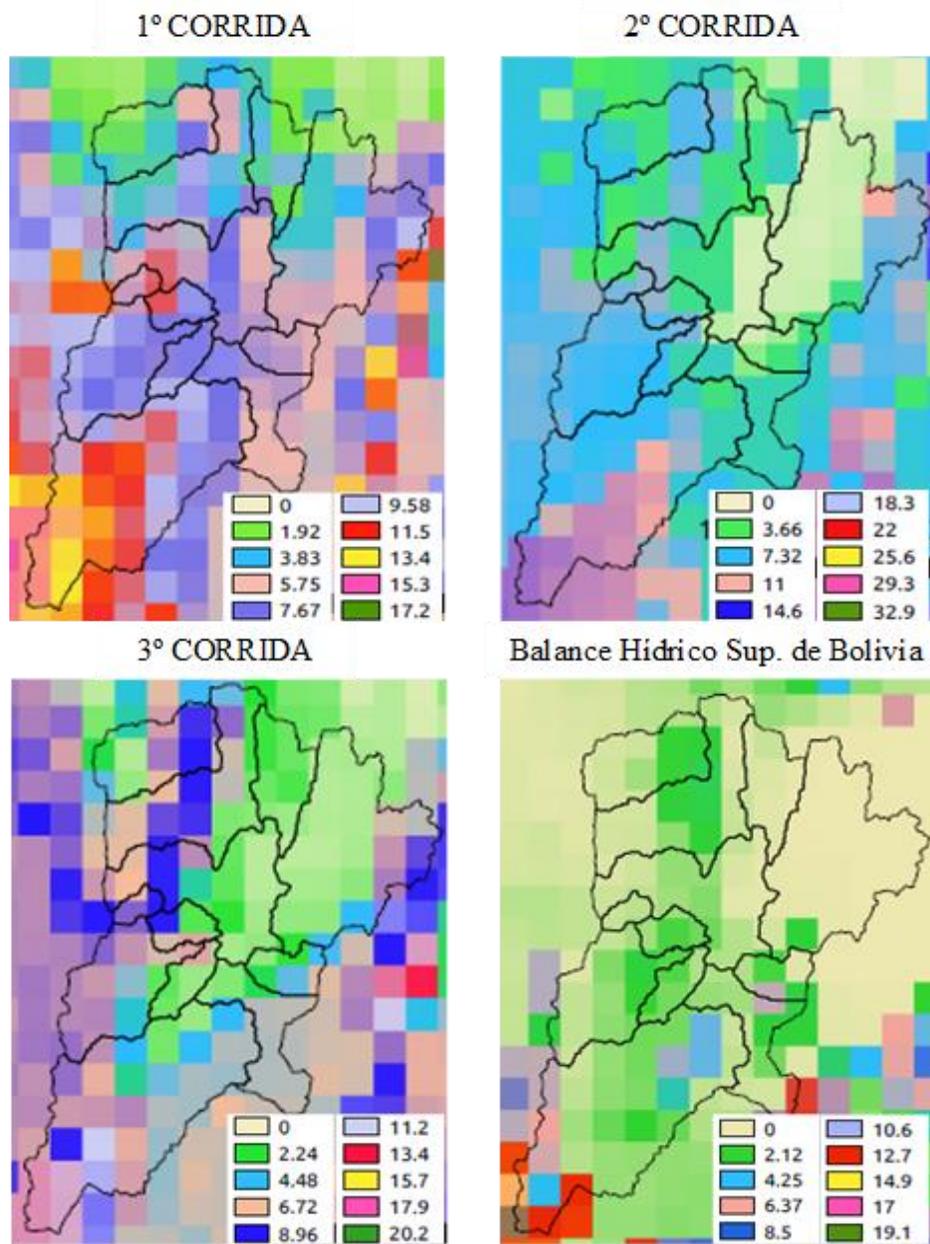
Figura N°76 Resultados Enero 1990



	Desde	Hasta
1° Corrida	68.6	239
2° Corrida	33.5	292.5
3° Corrida	51.2	164
B.H.S.B	32.6	211

Fuente: Elaboración Propia

Figura N°77 Resultados Agosto 1990



	Desde	Hasta
1° Corrida	0.0	15.3
2° Corrida	0.0	14.6
3° Corrida	0.0	11.2
B.H.S.B	0.0	12.7

Fuente: Elaboración Propia

A continuación, se tiene el análisis de las precipitaciones anuales obtenidas con GMET para algunos pasos de tiempo del periodo en común de las tres corridas realizadas con diferente criterio de selección de estaciones:

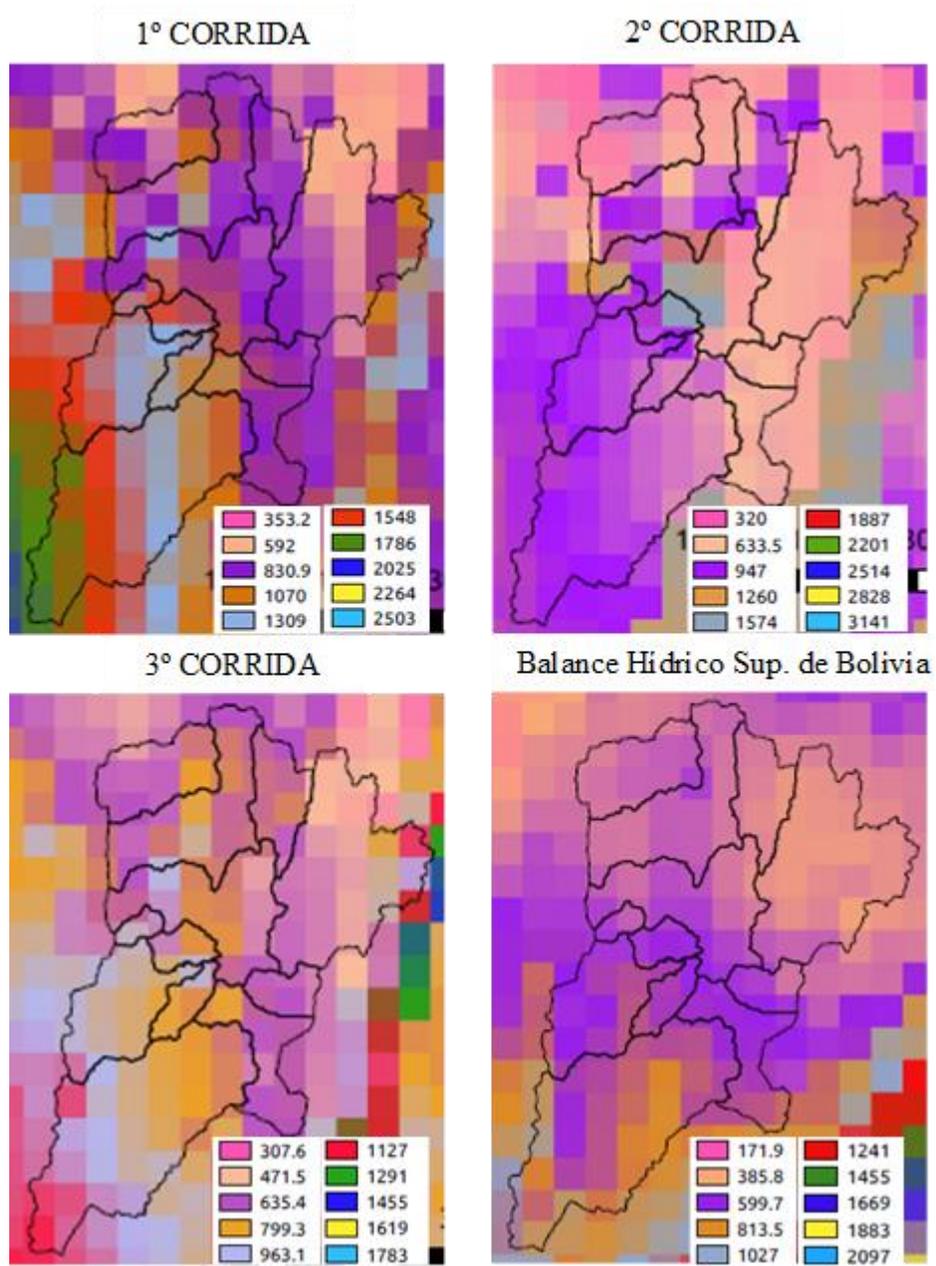
- Año 1980

Los resultados obtenidos con GMET para la precipitación anual en las tres corridas con diferente criterio de selección de estaciones para el año 1980 (Figura N°78) muestran diferencias en los mismos, esto se evidencia a través del análisis visual que muestra que no hay similitud en la variabilidad de las precipitaciones en la cuenca.

Se puede ver que en la primera corrida existe una variabilidad creciente de norte a sur, siendo más húmedo en el sur de la cuenca con precipitaciones que varían desde 1070 hasta 1786mm., mientras que al norte varía desde 353.2 hasta 830.9mm. En la tercera corrida, existe de igual manera variabilidad creciente de norte a sur, pero con precipitaciones de menor magnitud, ya que en el norte varían desde 307.6 hasta 635.4mm., y en el sur desde 635.4 hasta 1127mm. Por otro lado, se observa que en la segunda corrida los resultados difieren ya que si bien hay una variación creciente de norte a sur con magnitudes que oscilan entre 320 y 947mm., se concluye que la parte central de la cuenca sería la más húmeda debido a las precipitaciones presentadas con magnitudes que alcanzan los 1574mm.

Comparando estos resultados obtenidos en las tres corridas realizadas con los del Balance Hídrico Superficial de Bolivia podemos apreciar a través del análisis visual que no hay similitud en la variabilidad de las precipitaciones en la cuenca pero si se hace un análisis de las magnitudes de las precipitaciones podemos ver que la tercera corrida presentaría resultados más cercanos a los presentados en el B.H.S.B ya que en el mismo se tiene que la variación de precipitación es de igual manera, creciente de norte a sur con magnitudes que oscilan entre 385.8 y 599.7mm., y en el sur desde 599.7 hasta 1027mm.

Figura N°78 Resultados 1980



	Desde	Hasta
1° Corrida	353.2	1786
2° Corrida	320.0	1574
3° Corrida	307.6	1127
B.H.S.B	385.8	1027

Fuente: Elaboración Propia

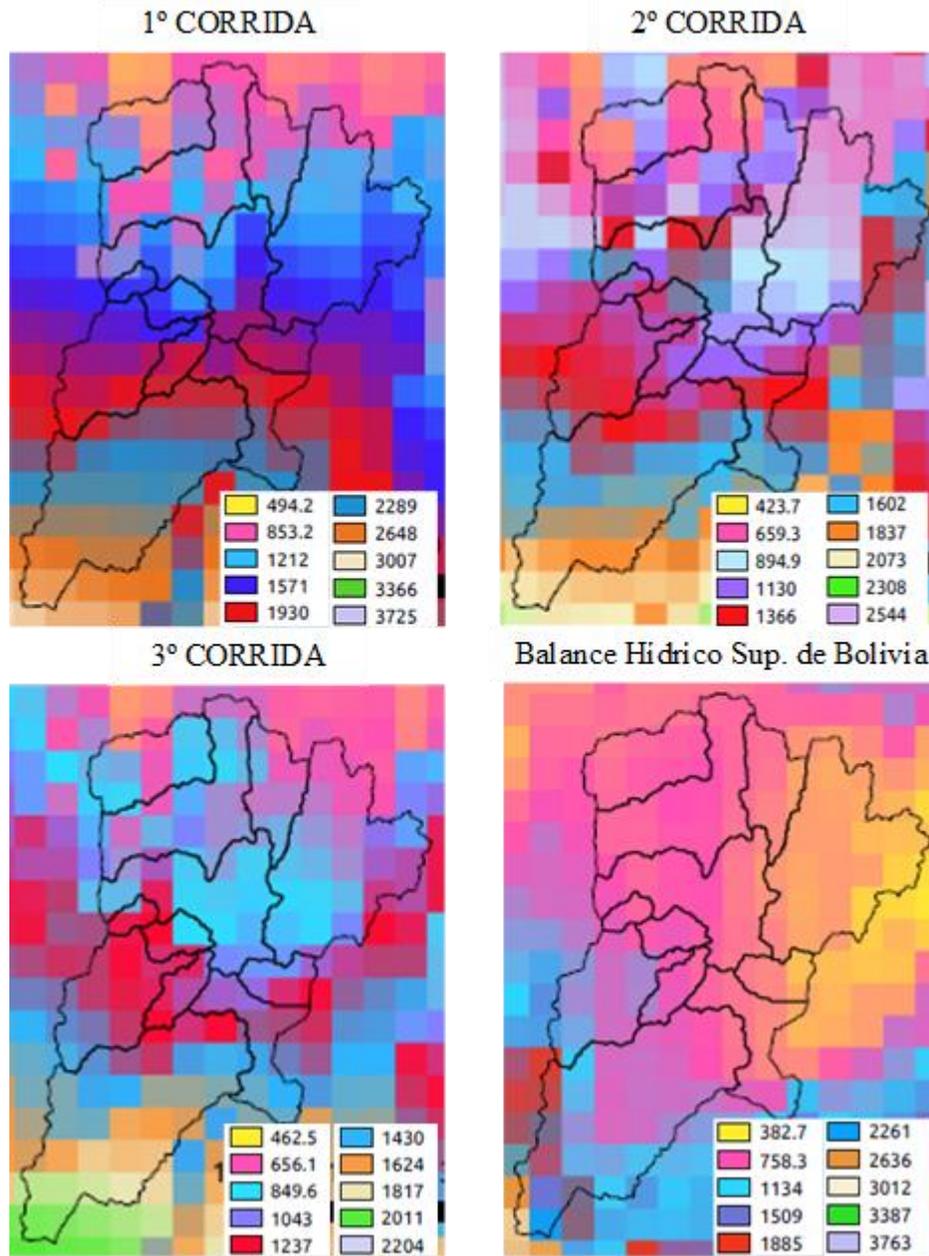
- Año 1985

Los resultados obtenidos con GMET para la precipitación anual en las tres corridas para el año 1985 (Figura N°79) muestran diferencias en los mismos, esto evidenciado a través del análisis visual que muestra que no hay similitud en la variabilidad de las precipitaciones en la cuenca.

Se observa también que en las tres corridas existe una variabilidad creciente de norte a sur siendo más húmedo el sur de la cuenca. En la primera corrida se tienen precipitaciones que varían desde 1571 hasta 2648mm. en el sur mientras que al norte varía desde 494.2 hasta 1571mm. En la segunda corrida se observa que las precipitaciones son menores a las anteriores con magnitudes que varían desde 423.7 hasta 894.9mm., en el norte y en el sur desde 1130 hasta 2073mm. En la tercera corrida se tienen precipitaciones con magnitudes que oscilan entre 462.5 y 849.6mm., en el norte, mientras que en el sur varían desde 1043 hasta 2011mm.

Comparando estos resultados obtenidos de las tres corridas realizadas con los del Balance Hídrico Superficial de Bolivia podemos apreciar que no hay similitud en la variabilidad de las precipitaciones en la cuenca pero con un análisis de las magnitudes de las precipitaciones obtenidas podemos decir que en la primera corrida se presentan resultados más cercanos a los presentados en el B.H.S.B ya que en el mismo se tiene que las precipitaciones varían desde 382.7 hasta 1134mm., en el norte de la cuenca y en el sur oscilan entre 1134 y 2636mm.

Figura N°79 Resultados 1985



	Desde	Hasta
1° Corrida	494.2	2648
2° Corrida	423.7	2073
3° Corrida	462.5	2011
B.H.S.B	382.7	2636

Fuente: Elaboración Propia

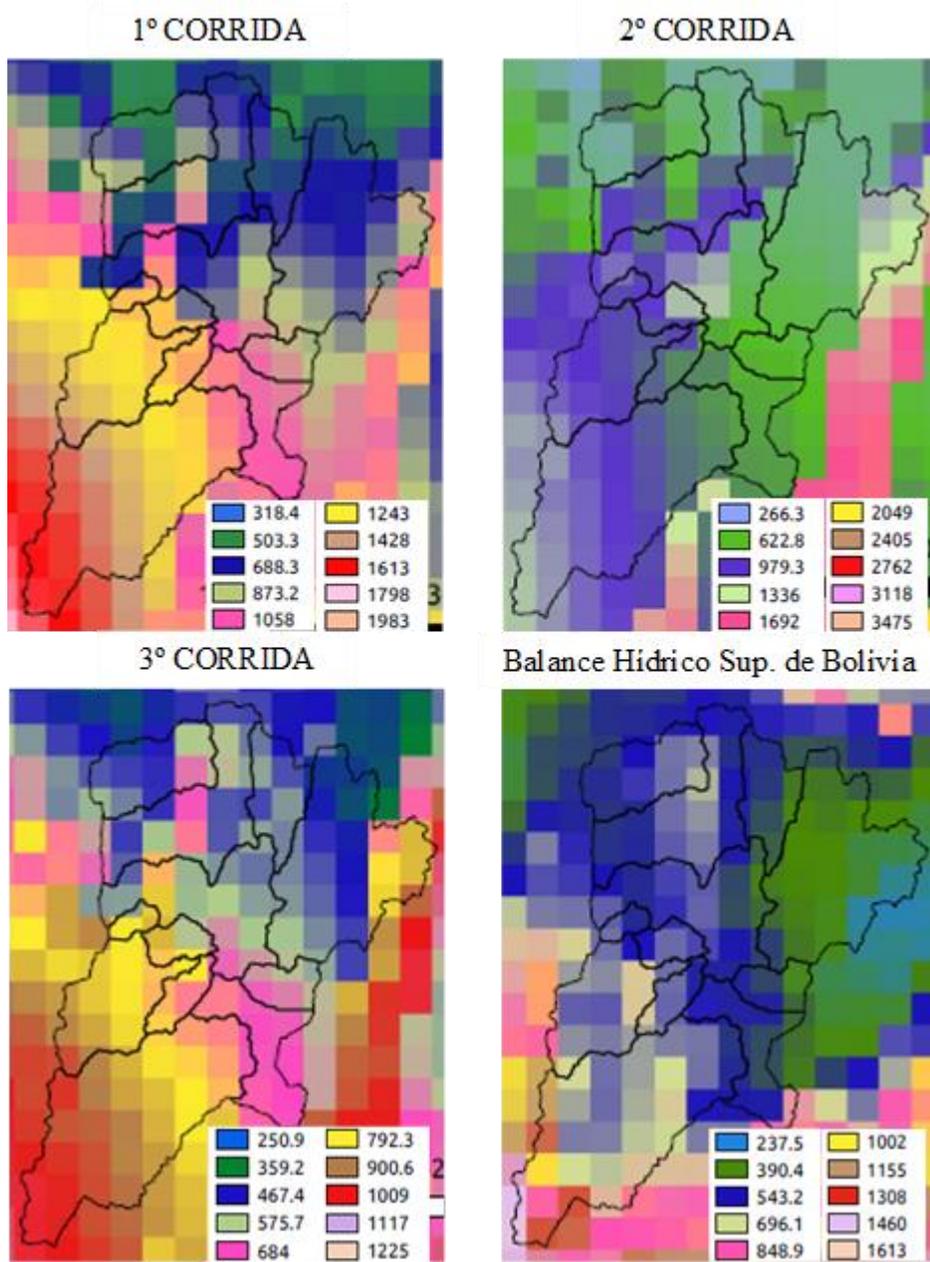
- Año 1990

Los resultados obtenidos con GMET para la precipitación anual en las tres corridas para el año 1990 (Figura N°80) muestran diferencias en los mismos ya que no hay similitud en la variabilidad de las precipitaciones en la cuenca.

Se observa que en la primera corrida existe una variabilidad creciente de norte a sur, siendo más húmedo en el sur de la cuenca con precipitaciones que varían desde 1058 hasta 1613mm., mientras que al norte varían desde 318.4 hasta 1058mm. En la tercera corrida se puede observar de igual manera variabilidad creciente de norte a sur, pero con precipitaciones de menor magnitud en el sur ya que varían desde 684 hasta 1009mm. y en el norte desde 359.2 hasta 684mm. Finalmente, en la segunda corrida los resultados difieren ya que si bien se observa una variación creciente de norte a sur con magnitudes que oscilan entre 266.3 y 979.3mm., la parte central de la cuenca sería la más húmeda debido a que presenta precipitaciones con magnitudes que alcanzan los 1336mm.

Comparando estos resultados obtenidos en las tres corridas realizadas con los del Balance Hídrico Superficial de Bolivia podemos apreciar que no hay similitud en la variabilidad de las precipitaciones en la cuenca, pero si se hace un análisis de las magnitudes de las precipitaciones podemos observar que la tercera corrida presentaría resultados más cercanos a los presentados en el B.H.S.B cuyas precipitaciones varían desde 237.5 hasta 696.1mm en el norte de la cuenca y en el sur oscilan entre 696.1 y 1002mm.

Figura N°80 Resultados 1990



	Desde	Hasta
1° Corrida	318.4	1613
2° Corrida	266.3	1336
3° Corrida	359.2	1009
B.H.S.B	237.5	1002

Fuente: Elaboración Propia

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

Para concluir con este trabajo de tesis, el presente capítulo se dedicará a mostrar las conclusiones y recomendaciones obtenidas a lo largo del estudio, esto con el fin de que se le pueda dar continuidad al proyecto, así como mostrar los resultados obtenidos.

La investigación realizada trabajó en la recopilación de información climatológica de las estaciones presentes dentro de la Cuenca del Río Guadalquivir y en zonas cercanas a la misma. Se procedió a hacer un análisis de los datos correspondientes a la precipitación y temperatura a nivel diario, con diferentes herramientas como ser el RCLimindex y DECADE que nos proporcionaron un control de calidad de los datos para verificar límites y detectar errores graves; también se realizaron los test de homogeneización con el método de las Alternancias y el método de Helmert para verificar que las series de precipitación no presenten inhomogeneidades. Asimismo, en cada caso seleccionar con diferentes criterios las estaciones que serían utilizadas en la generación de la grilla meteorológica GMET.

Las conclusiones que se pudieron extraer y surgieron tanto de las informaciones como de los análisis presentados en los capítulos precedentes, son las que se exponen a continuación.

5.1. Conclusiones

- Después de realizar la recopilación de la información se identificó que dentro de la Cuenca del Río Guadalquivir se cuenta con 16 estaciones en funcionamiento, estas tienen periodos de registro variados y 45 estaciones cerradas de las cuales algunas presentan un periodo de registro de mediciones mayor a 20 años a pesar de que actualmente no se encuentren en funcionamiento.
- Las estaciones ubicadas en cercanía al área de estudio son seis activas y diez cerradas de las cuales algunas de ellas al contar con registros en los periodos que se deseaba analizar nos brindaron información relevante para la continuidad del trabajo de generación de la grilla meteorológica GMET.
- Un riguroso control de calidad y la homogeneización de las series de precipitación es un proceso previo, ineludible y clave al momento de usar estos datos, ya que los mismos nos garantizan que las series de tiempo de las distintas estaciones

representen efectivamente señales asociadas al sistema climático en la zona y no a otros factores (errores del sensor, humanos, desplazamiento de la estación, cambio del medio circundante a la estación, etc.). Esto, nos asegura que los resultados alcanzados en estudios posteriores sean lo suficientemente robustos como para poder confiar en ellos.

- En cada una de las corridas realizadas con diferente criterio de selección de estaciones, se muestran resultados que deberían ser sometidos a mayor análisis y a una evaluación mucho más profunda con mayores cambios y mejoras en la información pluviométrica de entrada para así obtener resultados óptimos.
- En la primera corrida se generó GMET dentro del área en estudio para el periodo 1980-2016 con 26 estaciones que cumplieron el análisis de control de calidad realizado con RCLimindex y DECADE, además de tener el 60% o más de completitud en su registro como primer criterio para seleccionar las estaciones. Así, los resultados presentados en la grilla meteorológica nos muestran que en algunos pasos de tiempo tanto en escala mensual como anual existe similitud con los resultados presentados en el Balance Hídrico Superficial de Bolivia por lo que podemos decir que se necesita aún mayor análisis en la información de entrada.
- En la segunda corrida GMET fue generado dentro del área de estudio para el periodo 1980-1992 ya que después de un análisis del periodo de registro de las estaciones, se determinó que en éste se tenía un registro más completo y continuo que fue el segundo criterio con el que se seleccionó 37 estaciones que cumplieron el análisis de control de calidad para la generación de la grilla meteorológica, siendo en la misma donde observamos que la inclusión de estaciones en la parte central de la cuenca influyó en el desempeño de GMET generando resultados para esta zona con gran variación respecto al B.H.S.B. puesto que nos muestra precipitaciones con magnitudes elevadas.
- En la tercera corrida la selección de estaciones a emplearse en la generación de la grilla meteorológica se basó en la depuración de las estaciones que no cumplieron el análisis de homogeneidad realizado con los métodos de Helmert y método de las Alternancias en las series de precipitación anual, por lo que GMET fue generado para el periodo 1980-1992 con 30 estaciones que pasaron dicho control

mostrándonos resultados más óptimos y con menor variación respecto al B.H.S.B. Por lo tanto, se podría decir que al aumentar la cantidad de estaciones de entrada y trabajar con datos completos sin rellenar periodos largos de las series, sería una mejor alternativa para el trabajo realizado, comprobando lo expuesto en la bibliografía donde hace énfasis en que el proceso de relleno de datos faltantes introduce error en las series de datos históricos de precipitación.

5.2.Recomendaciones

- Definir criterios y procedimientos que permitan mejorar la base de datos climática histórica. Esta recomendación se basa en el hecho de que debido a la escasez de estaciones de observación existen vacíos de datos de entrada de precipitación y temperatura, lo que conlleva a obtener cierta incertidumbre en los resultados.
- Hacer una exploración de variables adicionales que puedan ser correlacionadas con el clima y que puedan introducirse como factores dentro del algoritmo de cálculo de la precipitación, ya que estas deberían ayudar a explicar los efectos de la topografía en la organización espacial de la precipitación.
- Realizar un estudio que evalúe las zonas que requieren la implementación de estaciones meteorológicas para tener una mejor distribución y así poder optimizar lo más que se pueda los procesos de obtención de datos, evitando así la presentación de vacíos en la información que puedan generar incertidumbre.
- Aplicar métodos que ayuden a establecer el grado de correlación entre las estaciones y averiguar si sus ubicaciones son las suficientes para mostrar la confiabilidad de la información tomada en los distintos puntos de observación.
- Aumentar la cantidad de datos de entrada y/o incluir datos de productos de estimación satelital de precipitación con datos satelitales, como podría ser la extracción de puntos de muestreo de CHIRPS u otros productos satelitales en regiones en las que no se cuente con estaciones pluviométricas para mejorar la distribución espacial y tener menos vacíos en la información.