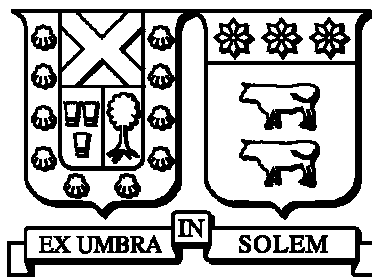


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE OBRAS CIVILES
VALPARAÍSO – CHILE



DESAGREGACIÓN DE LA PRECIPITACIÓN
PRONOSTICADA A CORTO PLAZO UTILIZANDO
VARIABLES DE CFSv2 EN LA ZONA CENTRO DE CHILE

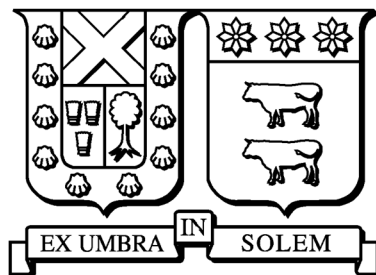
LUIS SEBASTIÁN CÓRDOBA ROMERO

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL

PROFESORA GUÍA
LINA MABEL CASTRO HEREDIA

Enero 2022

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE OBRAS CIVILES
VALPARAÍSO – CHILE



DESAGREGACIÓN DE LA PRECIPITACIÓN
PRONOSTICADA A CORTO PLAZO UTILIZANDO
VARIABLES DE CFSv2 EN LA ZONA CENTRO DE CHILE

MEMORIA DE TITULACIÓN PRESENTADA POR
LUIS SEBASTIÁN CÓRDOBA ROMERO

COMO REQUISITO PARCIAL PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL

PROFESORA GUÍA
LINA MABEL CASTRO HEREDIA

Enero 2022

AGRADECIMIENTOS

Agradezco a todas las personas que me apoyaron, que creyeron en mí, a quienes me entregaron consejos, guía, a quienes siempre me desearon lo mejor. Agradezco a mi familia, mis compañeras, compañeros, amigas, amigos, profesoras y profesores.

A mis hermanas, Yasna y Katerin, a mis sobrinos Nicolas, Alejandro y Javier, a mi sobrina Agustina, a mi padre Sebastián, a mi tío abuelo Luis, y a mi tío Hernán. En especial a mi madre, Laura Romero, quien me apoyó desde pequeño con la idea de ir a la universidad. Este desafío fue posible gracias a su apoyo incondicional y su ejemplo diario de esfuerzo y perseverancia.

A mis profesores del Liceo Abate Molina, Nilda Paris, Francisco Orellana, Orlando Figueroa y Eduardo Neira. Gracias por creer en mí.

También a mis amigas y amigos de Revolución Animal UTFSM por su contención y sus esperanzas, en especial a Daniel Cáceres, Gonzalo Peña, Javiera Rosas, Constanza Matthews, Gregori Casanova, Camila González, Gabriel Sanllehi, Natali Muñoz y Valentina Astete. Gracias por estar siempre.

A mis compañeros y amigos de la carrera, Juan Pablo Morales, Norman Gallardo, Cristian Jara y Ricardo Zúñiga, gracias por su apoyo en este proceso. De manera especial, a Fernanda Barrientos y Sergio Carrasco, por su amistad, guía y consejos. A quienes me apoyaron desde la universidad; gracias Eliana Ponce, Paulino Gallardo y Felipe Salgado. A mi compañero y amigo de Enseña Chile, Felipe Vergara.

Agradezco a las personas que hicieron posible este trabajo. A mis compañeras y compañeros de la salita de memoristas que me facilitaron generosamente equipos para el procesamiento de datos. A la profesora Diana Pozo y al profesor Omar Cuevas, docentes de meteorología de la Universidad de Valparaíso, quienes me orientaron de forma sabia en el estudio fenómeno físico. Finalmente, quiero reconocer a mi profesora guía, Lina Castro, por darme la posibilidad de trabajar en este tema, por acompañarme, por su paciencia y profunda dedicación.

RESUMEN

Existe una necesidad de predecir la precipitación a corto plazo, pero los modelos numéricos de predicción a escala global no incorporan las condiciones a escala local, puesto que las escalas espaciales globales son muy gruesas. Este estudio se centra en la influencia de distintas variables atmosféricas utilizadas para la corregir la desagregación estadística (*statistical downscaling*) de la precipitación en la zona centro de Chile, desde los modelos globales a una escala local. Por un lado, para la precipitación observada se utilizan los datos de 29 estaciones meteorológicas en la zona de estudio en el periodo 1995-2017. Por otra parte, las variables atmosféricas son obtenidas desde el sistema de pronóstico del clima CFSv2, las cuales son: precipitación, humedad, temperatura, presión y altura geopotencial. Se evalúa la correlación entre las variables atmosféricas y la precipitación observada. Además, con estos conjuntos de datos se generan modelos de regresión lineal multivariada y se obtienen precipitaciones pronosticadas, para dos periodos de datos: calibración y validación. Los resultados indican que la desagregación desde la variable atmosférica de precipitación de CFSv2 a una estación meteorológica que mide la precipitación observada, puede ser corregida mediante la incorporación de tres potenciales variables atmosféricas de acuerdo al siguiente orden: altura geopotencial a 500 hPa, altura geopotencial a 850 hPa y la humedad específica a 700 hPa.

ABSTRACT

There is a need for short-term precipitation prediction, but global-scale numerical prediction models do not incorporate local-scale conditions since global spatial scales are very coarse. This study focuses on the influence of different atmospheric variables used to correct the statistical downscaling of precipitation in central Chile, from global models to a local scale. First of all, data from 29 meteorological stations in the study area from 1995-2017 are used for the observed precipitation. On the other hand, the atmospheric variables are obtained from the climate forecast system CFSv2: precipitation, humidity, temperature, pressure, and geopotential height. The correlation between atmospheric variables and observed precipitation is evaluated, multivariate linear regression models are generated with these data sets, and predicted precipitation is obtained for two data periods: calibration and validation. The results shows that the disaggregation from the atmospheric variable of precipitation from CFSv2 to a meteorological station that measures observed precipitation can be corrected by incorporating three potential atmospheric variables according to the following order: geopotential height at 500 hPa, geopotential height at 850 hPa, and specific humidity at 700 hPa.

GLOSARIO

CFS	<i>Climate Forecast System</i> – Sistema de Pronóstico del Clima
CFSv2	<i>Climate Forecast System version 2</i> – Sistema de Pronóstico del Clima versión 2
Cluster	Grupo, conglomerado
Downscaling	Desagregación – Reducción de escala
DW	Test estadístico Durbin Watson, proporciona el grado de independencia de los residuos.
FIV	Factor de incremento de la varianza, test estadístico ocupado para estudiar la no-colinealidad
Grilla	Malla o cuadrícula del modelo NWP
MAE	<i>Mean Absolute Error</i> – Error Absoluto Medio, diferencia promedio entre los datos del predictando y los valores pronosticados.
NCEP	<i>National Centers for Enviromental Prediction</i> – Centros Nacionales de Predicción Ambiental
NWP	<i>Numerical Weather Prediction</i> – Modelo Numéricos de Predicción Meteorológica
Operational	Operacional, periodo de validación de los modelos (2011 – 2017)
P_{obs}	Precipitación observada en la estación meteorológica
P_{pron}	Precipitación pronosticada por un modelo
PRATE	Precipitación medida por CFSv2, en unidades de kg/(m ² s)
Predictando	<i>Predictand</i> - Variable que se predice a partir de los predictores
Predictor	Variable independiente que en conjunto o no de otros predictores, predice al predictando
PRMSL	Presión superficial a nivel del mar, en Pascales (Pa)
Pronostico	Valor pronosticado por medio de un modelo a partir de los predictores
Q700	Humedad específica a 700 hPa, en kg/kg
r	Coeficiente de Correlación de Pearson
RCM	<i>Regional Climate Model</i> – Modelo Climático Regional
Reforecast	Retrospectiva, periodo de calibración de los modelos (1995 – 2010)
T850	Temperatura a 850 hPa, en grados Kelvin (k)

TMPSFC	Temperatura superficial, en grados Kelvin (k)
VA	Variable Atmosférica a escala global del modelo NWP CFSv2
Z500	Altura geopotencial a 500 hPa, en metros geopotenciales (mgp)
Z850	Altura geopotencial a 850 hPa, en metros geopotenciales (mgp)

ÍNDICE

Agradecimientos	I
Resumen.....	II
Abstract	II
Glosario	III
Índice	IV
1. Introducción	8
2. Objetivos	9
2.1. Objetivo principal	9
2.2. Objetivos específicos.....	9
3. Revisión del estado del arte	10
3.1. Modelos numéricos de predicción.....	10
3.1.1. Sistema de Pronóstico del Clima	11
3.2. Métodos de desagregación	12
3.2.1. Desagregación dinámica	13
3.2.2. Desagregación estadística.....	14
3.2.3. Comparación de técnicas de desagregación	17
3.3. Desagregación estadística basada en la regresión lineal	19
3.3.1. Regresión lineal	19
3.3.2. Selección de predictores	21
3.3.3. Supuestos del modelo de regresión lineal	22
3.4. Descripción de las variables atmosféricas.....	24
3.4.1. Humedad específica	25

3.4.2.	Temperatura superficial y atmosférica	25
3.4.3.	Presión atmosférica.....	26
3.4.4.	Altura geopotencial	27
3.4.5.	Estudios relacionados.....	27
3.5.	Análisis cluster.....	31
3.5.1.	Método jerárquico	31
3.5.2.	Método no jerárquico	33
4.	Área de estudio y conjuntos de datos.....	33
4.1.	Área de estudio y clima	33
4.2.	Estaciones meteorológicas.....	34
4.3.	Variables atmosféricas	35
5.	Metodología	37
5.1.	Análisis cluster.....	38
5.2.	Proceso de datos CFSv2	39
5.3.	Análisis de correlación.....	40
5.3.1.	Correlación espacial	40
5.3.2.	Correlación mensual.....	41
5.3.3.	Correlación entre variables	41
5.4.	Calibración de modelos.....	42
	Regresión lineal multivariada.....	42
5.5.	Verificación de supuestos	42
5.6.	Validación de modelos	43
6.	Resultados	44
6.1.	Clusters de estaciones meteorológicas.....	44
6.1.1.	Método <i>Single</i>	44
6.1.2.	Método Ward	45
6.2.	Correlaciones.....	47
6.2.1.	Correlación espacial	47
6.2.2.	Correlación mensual.....	51
6.2.3.	Correlación entre variables	55
6.3.	Calibración de los modelos	56
6.4.	Verificación de supuestos	60
6.4.1.	Linealidad	60

6.4.2.	Independencia.....	60
6.4.3.	Homocedasticidad.....	61
6.4.4.	Normalidad.....	63
6.4.5.	No colinealidad.....	66
6.5.	Validación de los modelos.....	66
6.5.1.	Correlación de los modelos.....	66
6.5.2.	Error absoluto medio de los modelos	68
7.	Discusión de resultados.....	71
7.1.	Análisis cluster.....	71
7.2.	Correlaciones.....	71
7.2.1.	Correlación espacial	71
7.2.2.	Correlación mensual.....	72
7.2.3.	Correlación entre variables	73
7.3.	Calibración.....	73
7.4.	Verificación de supuestos	74
7.4.1.	Linealidad	74
7.4.2.	Independencia.....	74
7.4.3.	Homocedasticidad.....	74
7.4.4.	Normalidad.....	74
7.4.5.	No colinealidad.....	75
7.5.	Validación	75
7.5.1.	Correlación de los modelos.....	75
7.5.2.	Error absoluto medio de los modelos	76
8.	Conclusiones.....	78
9.	Referencias.....	80
	Índice de Figuras	82
	Índice de tablas	84
	Anexos.....	85
A.	Regresión lineal.....	85
1.	Regresión lineal simple	85
2.	Regresión lineal multivariada.....	86
B.	RUTINAS EN SOFTWARE.....	87
1.	<i>Cluster.m</i>	87

2.	Proceso de datos de las variables atmosféricas.....	87
3.	<i>Correlacion_x_e.m</i>	88
4.	<i>Modelos_regresion_mensual.R</i>	89
5.	<i>Test_supuestos.R</i>	89
6.	<i>Validacion_correlacion.m</i>	89
C.	Puntos de grilla de máxima correlación	90
D.	Verificación de supuestos de modelos lineales	91
1.	Independencia.....	91
2.	Homocedasticidad.....	92
3.	Normalidad.....	96
4.	No-colinealidad	99
E.	Relación entre variables.....	100

1. INTRODUCCIÓN

Durante el desarrollo de la sociedad, ha existido una importante necesidad de predecir el tiempo atmosférico, y por sobre todo en la estimación de la precipitación. El análisis de esta variable es parte de importantes decisiones político-económicas, afectando principalmente a la actividad agrícola (Kilsby, 1998). Por otro lado, la predicción de la precipitación en sistemas de alerta temprana permite emitir información frente a los efectos de crecidas y áreas de inundación.

Para el pronóstico meteorológico a corto plazo, se han desarrollado modelos numéricos de predicción meteorológica (NWP, por sus siglas en inglés) a una escala global, destacándose el Sistema de Pronóstico del Clima (CFS, siglas en inglés). Este modelo incorpora la interacción global entre los océanos, la superficie y la atmósfera terrestre. No obstante, los modelos a escala global no incorporan las condiciones a escala local, debido a que las escalas espaciales globales son muy gruesas (grillas mayores a 50 km), en comparación con las cuencas usualmente estudiadas. Por lo que se hace necesario la reducción de escala (desagregación - *downscaling*).

Las técnicas de desagregación se clasifican en dos grandes grupos: desagregación dinámica y desagregación estadística. Las técnicas de desagregación estadística a su vez se subdividen en tres grupos: modelos de regresión, clasificación del clima y generadores de clima (Trzaska & Schnarr, 2014). En el grupo de los modelos de regresión, se encuentra la técnica de desagregación basada en relaciones numéricas, entre las variables atmosféricas entregadas por el NWP y la precipitación observada en estaciones meteorológicas (Anuchaivong, 2017). A partir de estas relaciones se pueden generar modelos de desagregación estadística, siempre y cuando, las variables atmosféricas estén relacionadas estrechamente con la precipitación observada (Wilby R. , 2002).

En Chile se han realizado estudios de desagregación estadística (Figueroa 2008, Souvignet 2010, Vega 2018, Araya 2020, entre otros), sin embargo, no se tiene la suficiente claridad de las variables a escala global que se debieran usar para corregir los modelos en una desagregación de precipitación. El presente estudio busca realizar un análisis más profundo sobre las variables que afectan en mayor medida en la desagregación de la precipitación en la zona centro de Chile.

Este trabajo presenta 4 secciones que se describen a continuación:

- Revisión del estado del arte: se presenta la teoría de variados métodos de desagregación, en especial la regresión lineal multivariada. Se realiza una revisión bibliográfica respecto a las variables utilizadas en otros estudios, analizando el comportamiento físico y los resultados.
- Área de estudio y conjuntos de datos: se presentan las estaciones meteorológicas de la precipitación observada y los puntos de grilla de CFSv2 para las variables atmosféricas globales, con sus respectivos periodos de análisis.

- Metodología: se indican los métodos de selección del punto de grilla, selección de las variables atmosféricas, aplicación de los modelos de regresión, verificación de supuestos de los modelos y método de análisis cluster.
- Resultados y discusión: se presentan los resultados de los clusters entre estaciones, las variables seleccionadas, las constantes de los modelos, las verificaciones de los supuestos de linealidad, los coeficientes de correlación, los errores absolutos medios y la respectiva discusión de los resultados.

2. OBJETIVOS

2.1. OBJETIVO PRINCIPAL

Estudiar el pronóstico de precipitación en la zona centro de Chile, mediante la generación de modelos de desagregación, utilizando pronósticos globales de CFS.

2.2. OBJETIVOS ESPECÍFICOS

1. Determinar las variables atmosféricas de CFS, necesarias para generar el modelo de desagregación.
2. Establecer la correlación entre las variables atmosféricas de CFS determinadas, con la precipitación observada en las estaciones meteorológicas.
3. Calibrar y validar modelos que contemplen la desagregación espacial y corrección del pronóstico de precipitación diaria con CFSv2.

3. REVISIÓN DEL ESTADO DEL ARTE

3.1. MODELOS NUMÉRICOS DE PREDICCIÓN

Los modelos numéricos de predicción meteorológica (NWP, siglas en inglés) generan una simulación del clima por medio de ecuaciones matemáticas consistentes con la física del problema; describiendo procesos, interacciones y estableciendo una retroalimentación entre la atmósfera, los océanos y los factores bióticos (ver cuadro de la Figura 1). Los modelos NWP son la principal herramienta que proporciona la información climática a escala mundial, de una manera razonablemente precisa (Trzaska & Schnarr, 2014).

Un NWP está constituido por celdas de cuadrículas (grilla) que representan áreas horizontales y verticales en la superficie de la tierra (Figura 1). En cada una de las celdas los modelos NWP calculan entre otros factores: vapor de agua y nubes, interacciones atmosféricas, cambios de la capa de nieve y el hielo marino, almacenamiento y transporte de calor, etc. (Wilby R. , 2009)

La resolución espacial de los modelos NWP oscila aproximadamente entre los 100 a 500 Km, donde cada celda de cuadrícula es homogénea, es decir, dentro de cada celda hay un valor para cada variable (Trzaska & Schnarr, 2014).

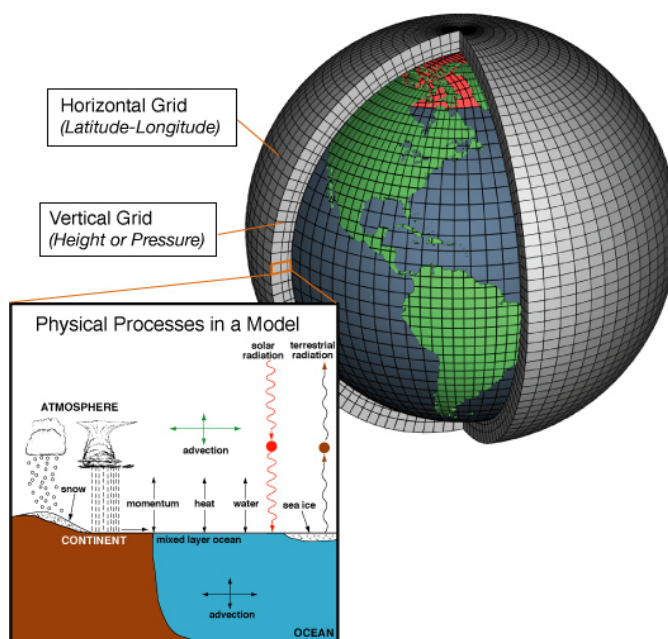


Figura 1: Modelo Numérico de Predicción, Fuente: (NOAA, NOAA Celebrates 200 Years, s.f.)

Existen una gran cantidad de modelos de numéricos de predicción, dentro de los cuales encontramos el modelo entregado por los *National Centers for Enviromental Prediction* (NCEP) de los Estados Unidos, que brindan los pronósticos del clima espacial a nivel nacional y mundial a sus socios y comunidades de usuarios externos. Actualmente, se encuentran disponibles los modelos de CFSv2.

3.1.1. SISTEMA DE PRONÓSTICO DEL CLIMA

El Sistema de Pronóstico del Clima (CFS, siglas en inglés) es un modelo de interacción global entre los océanos, la tierra y la atmósfera, el cual, se encuentra disponible desde el año 2004. Este modelo ha sido generado por docenas de científicos bajo la guía de *los National Center for Enviromental Prediction*, ofreciendo datos por hora con una resolución horizontal de hasta la mitad de un grado (aproximadamente 56 km), alrededor de la Tierra para muchas variables.

CFS utiliza los últimos enfoques científicos para tomar o asimilar observaciones de fuentes de datos, incluidas observaciones de superficie, de globo aerostático superior, de aeronaves y satelitales (National Centers for Enviromental Information). El modelo está dividido en dos etapas temporales, CFSv2Reforecast y CFSv2Operational:

I. CFSv2 Reforecast

Según los desarrolladores de este modelo numérico de predicción: *“Los pronósticos posteriores de 9 meses comienzan a partir de cada quinto día y se ejecutan los 4 ciclos de ese día (00, 06, 12 y 18 UTC), comenzando desde el 1 de enero de cada año, durante un período de 29 años, desde el año 1982 hasta el año 2010; estos pronósticos son necesario para calibrar las predicciones estacionales a más largo plazo. También, es posible generar pronósticos estacionales (123 días), iniciada desde cada ciclo UTC entre estos cinco días (00 UTC), durante el período de 12 años comprendido entre 1999 y 2010; necesario para calibrar las predicciones operativas de la primera temporada para pronósticos hidrológicos (precipitación, evaporación, escorrentía, flujo de corriente, etc.). Otra forma es ejecutar tres pronósticos posteriores de 45 días con 3 pronósticos diarios (6, 12 y 18 UTC), durante el período de 12 años comprendido entre 1999 y 2010; esto es necesario para las predicciones operacionales de circulaciones tropicales”.* Fuente: (NOAA & NCEP, CFS version 2 Information, 2010).

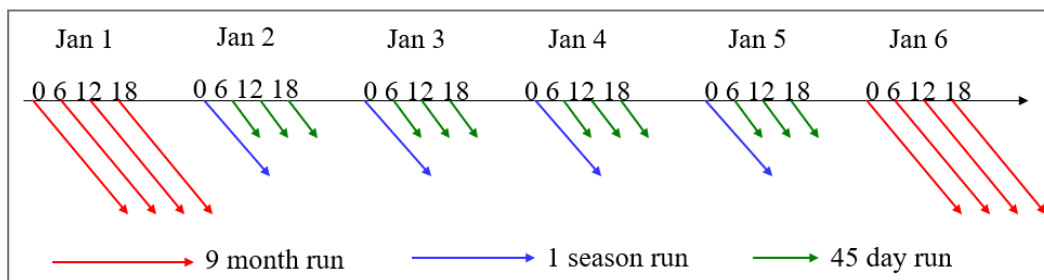


Figura 2: Tipos de pronósticos CFSv2 Reforecast. Fuente: (NOAA & NCEP, CFS version 2 Information, 2010)

II. CFSv2 Operational

Para esta etapa del modelo, existen cuatro ejecuciones de control por día a las 00, 06, 12 y 18 UTC del sistema de datos en tiempo real CFS, en un periodo de 9 meses. También, se encuentra la ejecución de control en una temporada (123 días) a las 00 UTC. Además, existe el pronóstico de 45 días que entrega 3 datos diarios (06, 12 y 12 UTC). Hay un total de 16 carreras de CFS por día, de las

cuales 4 saldrán a 9 meses, 3 saldrán a 1 temporada y 9 saldrán a 45 días. Fuente: (NOAA & NCEP, CFS version 2 Information, 2010).

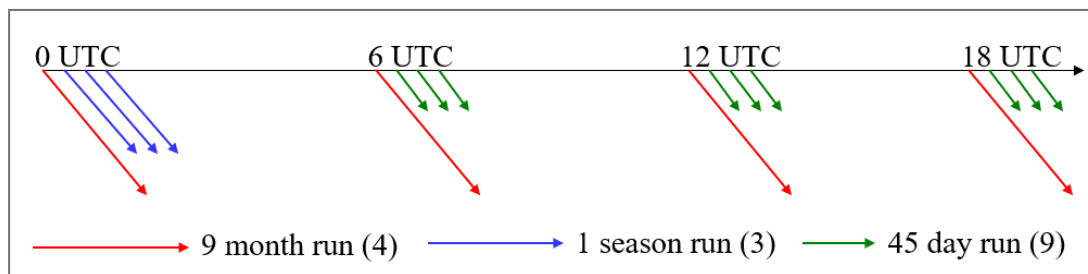


Figura 3: Tipos de pronósticos CFSv2 Operational. Fuente: (NOAA & NCEP, CFS version 2 Information, 2010)

3.2. MÉTODOS DE DESAGREGACIÓN

Los modelos numéricos de predicción del sistema climático nos entregan datos atmosféricos en una grilla espacial gruesa (Figura 4). Quedando restringidos en su utilidad para muchas aplicaciones de menor escala (sub-grilla), debido a su resolución espacial aproximada (Wilby & Wigley, 1997). Por lo anterior, se hace imprescindible un escalamiento de los datos, desde un nivel global a un escenario regional o local. Para conseguir la disminución de esta brecha de resolución se utilizan las llamadas técnicas de desagregación (*downscaling*) (Wilby R. , 2002).

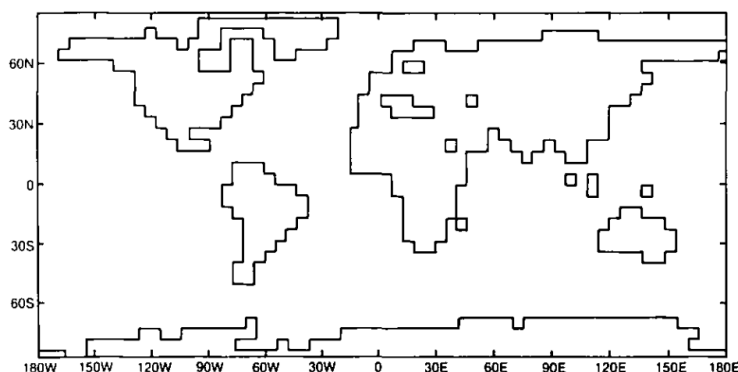


Figura 4: Grilla de un Modelo Numérico de Predicción típico. Fuente: (IPPC, 1990)

La desagregación o reducción de escala se basa en el supuesto de que el clima local es una combinación de características climáticas-atmosféricas a gran escala (global, continental, regional) y condiciones locales (topografía, cuerpos de agua, propiedades de la superficie terrestre) (Trzaska & Schnarr, 2014). Por ejemplo, los modelos hidrológicos requieren con frecuencia de procesos a menor escala (subcuenca), que ocurren en escalas espaciales mucho más pequeñas que las resueltas en los modelos NWP (Wilby & Wigley, 1997) (Figura 5).

Actualmente, existen dos grandes conjuntos de técnicas de desagregación como un medio para derivar el clima a una menor escala, se pueden clasificar ampliamente como una reducción de escala "dinámica" o "estadística" (Wu, 2012).

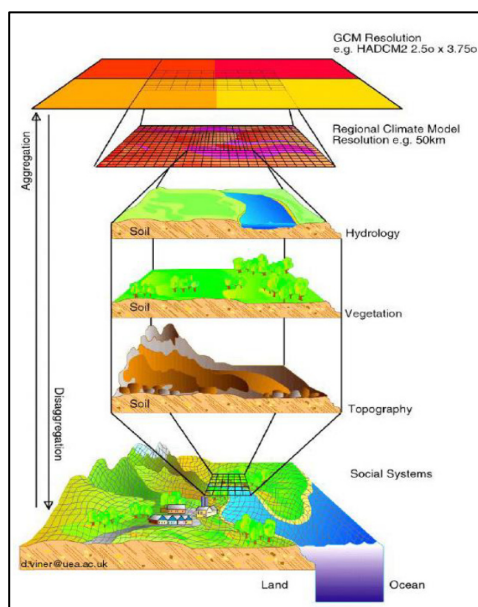


Figura 5: Conceptualización de agregación y desagregación entre modelos hidrológicos y atmosféricos. Fuente: (Trzaska & Schnarr, 2014)

3.2.1. DESAGREGACIÓN DINÁMICA

Las técnicas básicas de un Modelo Climático Regional (RCM, siglas en inglés) provienen esencialmente de la predicción numérica del clima desarrollada por Dickinson et al. (1989). Utilizar un método dinámico consta en la anidación de un RCM con mayor resolución dentro de una grilla con resolución NWP más gruesa. El RCM usa el modelo NWP para definir condiciones de contorno atmosféricas variables en el tiempo, alrededor de un dominio finito (Wilby R. , 2002), ver Figura 6. Así, en una cuadrícula más fina se calculan las propiedades atmosféricas mediante la resolución de las ecuaciones de movimiento y termodinámica (Spak, 2007). Los RCM simulan características climáticas a resoluciones de 10 – 50 km según las condiciones atmosféricas variables en el tiempo, donde los campos atmosféricos simulados por un modelo NWP se introducen en el límite del RCM a diferentes niveles verticales y horizontales. Esta información es procesada por el RCM, de modo que la física y la dinámica del modelo interno pueden generar la desagregación buscada mediante el ingreso de características detalladas (topográficas, costeras y dinámicas), por lo que este modelo presenta un gran costo de recursos computacionales, el que aumenta exponencialmente al aumentar la resolución (Wilby R. , 2009).

Si bien, los RCM han avanzado en la capacidad técnica, la gran cantidad de información de superficie conlleva un número de cálculos que podría ser tan grande o mayor que en un NWP, necesitando también procesos de calibración complejos. Además, diferentes RCM contienen diferentes parámetros físicos y esquemas climáticos, lo que significa que dos RCM con el mismo NWP de *input* pueden arrojar resultados diferentes. Por último, las cuadrículas suelen ser mayor a 10 km, lo que aún es demasiado grueso para estudios de impacto hidrológico y agrícola, por lo cual, en estos casos se utilizan métodos estadísticos en vez de un RCM, o se reduce la salida de un RCM por métodos estadísticos (Trzaska & Schnarr, 2014)

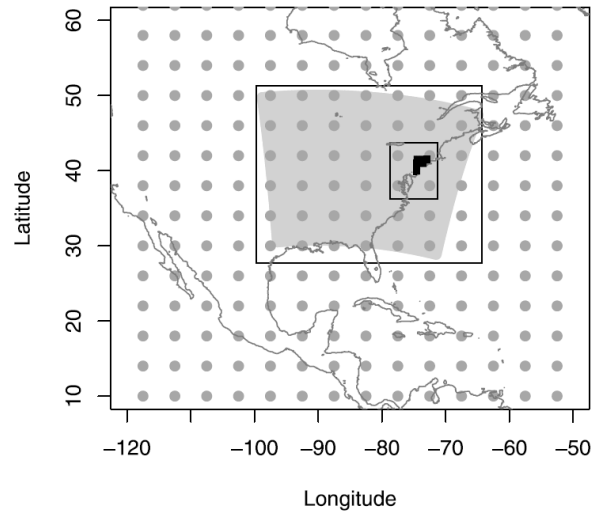


Figura 6 : Ejemplo de un campo de predicción y áreas de estudio. El dominio del modelo regional se encuentra en gris. Fuente: (Spak, 2007)

Uno de los proyectos más importantes que involucra un RCM es “*The Europe-South America Network for Climate Change Assessment and Impact Studies*” (CLARIS). Éste, fue un proyecto interdisciplinario de tres años para construir una red integrada entre Europa y Sudamérica, dedicándose a la obtención de estrategias de investigación común, para observar y predecir los cambios climáticos, junto con observar impactos en América del Sur y proporcionar información. Durante el proyecto, se utilizaron cinco RCM (MM5, PROMES, RCA3, REMO y WRF) y un modelo global de rejilla estirada (LMDZ), con el objetivo de determinar cuál RCM tenía un mejor funcionamiento en América del Sur. Los resultados arrojaron que los rendimientos de los RCM dependían del clima simulado, la región y la variable de interés (Menéndez, 2010).

3.2.2. DESAGREGACIÓN ESTADÍSTICA

La esencia del *statistical downscaling* (desagregación estadística) involucra establecer relaciones empíricas entre las variables atmosféricas a gran escala simuladas por un modelo NWP (tratadas como predictores) y las variables meteorológicas locales observadas (Huth, 1999), acuñada como *predictands* o predictandos para efectos de este trabajo.

Para construir los modelos estadísticos, se utilizan estas relaciones empíricas, los cuales, podrían traducir las anomalías del flujo a gran escala en anomalías de alguna variable local (Zorita & von Storch, 1999). Una vez que se ha determinado y validado una relación empírica, se emplean las futuras condiciones atmosféricas a gran escala, proyectadas por los modelos NWP para predecir las características futuras del clima local. Los métodos de desagregación estadística abarcan un variado grupo de métodos (ver Tabla 2) tanto en sofisticación como en aplicabilidad (Trzaska & Schnarr, 2014).

Los métodos estadísticos se emplean, en la reducción de escala a la salida de un NWP o de un RCM, mediante un análisis rápido y de menor costo computacional que el método dinámico, y además no

es necesario contar con la experiencia técnica que requieren los RCM, más aún, proporcionan información climática a escala de estación (Trzaska & Schnarr, 2014).

Los métodos de desagregación estadísticos, normalmente se han categorizado en tres conjuntos: los modelos de regresión (Wilby R. , 2002), los generadores de clima (Wilby & Wigley, 1997) y los de clasificación del clima (Bardossy, 1992; Storch, 1993). A continuación, se describen cada una de las categorías principales (ver Tabla 1).

➤ MODELOS DE REGRESIÓN

Establecen una relación empírica entre las variables de salida de un NWP o un RCM (predictor/es) y variables locales (predictandos). Es un método ampliamente aplicado, algunos autores son: Souvignet (2010), Anandhi (2011), y Vega (2018), entre otros. Esta metodología se puede aplicar a un predictando con su determinado grupo de predictores o a campos espaciales de predictores-predictando. Sin embargo, una de las limitaciones de los modelos de regresión es que las variables predictoras y predictandos requieren una distribución normal, por esta condición, hay inconvenientes con los datos extremos, puesto que éstos tienden a ubicarse más allá del conjunto de datos de calibración (Wilby R., 2002).

➤ CLASIFICACIÓN DEL CLIMA

La variable local se predice en base a escenarios de escala atmosférica. Estos métodos agrupan los días en un número finito de tipos de clima discretos (o estados) acorde con su similitud sinóptica o sistemas complejos ocultos. El estado atmosférico histórico seleccionado corresponde entonces a un valor o a una clase de valores de la variable local, que luego se replican en el estado atmosférico futuro. Comúnmente, los estados climáticos se definen, aplicando análisis *cluster* a los campos atmosféricos o usando esquemas de clasificación de circulación subjetiva, donde los patrones climáticos se agrupan conforme a su semejanza con los “vecinos más cercanos” o un conjunto de referencia (Wilby R., 2004). Además, estos métodos son más exigentes desde el punto de vista numérico que los modelos de regresión, debido a la gran cantidad de datos diarios analizados y generados.

➤ GENERADOR DE CLIMA

Utilizados regularmente en la desagregación temporal, los modelos generadores de clima replican atributos estadísticos de los registros meteorológicos observados, como la media diaria, varianza y covarianza, frecuencia, extremos, etc., permitiendo completar datos faltantes o producir series meteorológicas sintéticas largas de la variable local, para ser utilizadas más adelante en los modelos de impacto.

La esencia de la mayoría de los generadores de clima es un modelo de cadena Markov simple (de primer orden), para la ocurrencia de precipitación, suponiendo que la probabilidad de lluvia en cualquier día depende solo de que, si el día anterior fue lluvioso o seco, definido así, por las probabilidades de transición (Wilks & Wilby, 1999). Las variables secundarias como la temperatura máxima y mínima, la radiación y la velocidad del viento se agrupan en conjuntos de días húmedos y secos (Wilby R., 2009). Los generadores meteorológicos son intensivos en datos, requieren largas

secuencias de datos diarios y son sensibles a la falta de datos o a datos erróneos en el conjunto de calibraciones.

Tabla 1: Desagregación espacial estadística: categoría, método, variables predictoras y de predicción, ventajas y desventajas. Fuente: (Trzaska & Schnarr, 2014)

CATEGORÍA Y MÉTODO		PREDICTOR Y PREDICTANDO	VENTAJAS	DESVENTAJAS
Métodos lineales (Espacial)	Método Delta	El mismo tipo de variable (p.e., tanto la temperatura mensual, como la precipitación mensual)	- Relativamente fácil de aplicar - Emplea toda la gama de variables de predicción disponibles	- Requiere la normalidad de los datos (p.e., la temperatura mensual, la temperatura media a largo plazo) - No se puede aplicar a distribuciones No-normales (p.e., la precipitación diaria) - No es adecuado para eventos extremos
	Regresión lineal simple y múltiple	Las variables pueden ser del mismo tipo o diferentes (p.e., ambas temperaturas mensuales, o un viento mensual con una precipitación mensual)		
	CCA & SVD			
Clasificación del Clima (Espacial y Temporal)	Método análogo	Las variables pueden ser del mismo tipo o diferentes (p.e., tanto la temperatura mensual, un campo de presión atmosférica a gran escala y el otro las precipitaciones diarias)	- Produce vínculos físicamente interpretables con el clima de la superficie - Versátil, es decir, puede aplicarse tanto a los datos distribuidos normalmente como a los No distribuidos normalmente.	- Requiere un paso adicional de clasificación del tipo de clima - Requiere una gran cantidad de datos y algunos recursos computacionales - Incapaz de predecir nuevos valores que estén fuera del rango de los datos históricos.
	Análisis Cluster			
	ANN			
	SOM			
Generador del Clima (Espacial y Temporal)	LARS-WG	El mismo tipo de variable, diferentes escalas temporales (p.e., el predictor es la precipitación mensual y	- Capaz de simular la duración de los períodos húmedos y secos	- Requiere una gran cantidad de datos - Sensible a la falta de datos o a datos erróneos en el

	MarkSim GCM	la predicción es la precipitación diaria)	- Produce un gran número de series, lo que es valioso para el análisis de la incertidumbre	periodo de calibración. - Sólo algunos generadores meteorológicos pueden comprobar la coherencia entre múltiples variables (p.e., no se debe predecir la alta insolación en un día con precipitación). - Requiere la generación de múltiples series temporales y el post-procesamiento estadístico de los resultados.
	NHMM	Las variables pueden ser del mismo tipo o diferentes (p.e., tanto la temperatura mensual, una presión atmosférica a gran escala y la otra las precipitaciones diarias)	- Produce escenarios novedosos	

3.2.3. COMPARACIÓN DE TÉCNICAS DE DESAGREGACIÓN

Como se indica en la sección anterior, existen diversas técnicas de desagregación con una variedad general de enfoques. Esto es reflejo, de los distintos objetivos y recursos que requiere cada método. Es así, como existen variados métodos, los que obedecen a la resolución espacial y temporal buscada para las variables climáticas de mayor impacto o interés (Trzaska & Schnarr, 2014). A continuación, se presenta la Tabla 2, en la que se resumen y comparan diferentes aspectos de los dos grandes grupos de métodos de desagregación: dinámicos y estadísticos.

*Tabla 2: Fortalezas y debilidades de desagregación estadística y dinámica.
Fuente: (Wilby R. , 2002) & (Trzaska & Schnarr, 2014)*

Desagregación dinámica	
Proporciona	* Información de celda de la grilla de 10-50 km * Información en sitios sin datos de observación * Series temporales diarias * Series temporales mensuales * Escenarios para eventos extremos
Requiere	* Altos recursos computacionales y experiencia * Gran volumen de información y datos introducidos * Simulaciones fiables de los modelos NWP
Fortalezas	* Basado en un mecanismo físico consistente * Resuelve los procesos atmosféricos y superficiales que ocurren a escala sub-grilla del NWP

Debilidades	* No está limitado por los registros históricos, de modo que se pueden simular nuevos escenarios * Debido a las exigencias de los cálculos, los RCMs son típicamente dirigidos por sólo una o dos simulaciones de escenarios de NWP/emisiones. * Existe un limitado número de RCMs disponibles y pocos resultados para distintos escenarios globales * Requiere recursos informáticos significativos * Los resultados dependen de los supuestos del RCM; diferentes RCM darán resultados diferentes * Los resultados del RCM pueden requerir una mayor desagregación y corrección de sesgo. * No es transferible fácilmente a nuevas regiones
Desagregación estadística	
Proporciona	* Información a cualquier escala donde se posean datos de mediciones, por ejemplo, a escala de una estación meteorológica * Series temporales diarias (algunos métodos) * Series temporales mensuales * Escenarios para eventos extremos (solo algunos métodos) * Escenario para cualquier variable observada de manera consistente
Requiere	* Recursos computacionales medio/bajo * Volumen de entrada de datos medio/bajo * Suficiente cantidad de datos de observación de buena calidad * Simulaciones fiables del modelo NWP
Fortalezas	* Entrega información climática a escala-estación * Barato, poco exigente computacionalmente y fácilmente transferible a otras regiones * Los conjuntos de escenarios climáticos permiten análisis de riesgo/incertidumbre * Los métodos van de simples a elaborados y son lo suficientemente flexibles como para adaptarse a fines específicos
Debilidades	* Es posible que no se disponga de datos observados de alta calidad para muchas áreas o variables * La elección del tamaño del dominio y la ubicación afecta a los resultados * Requiere datos de alta calidad para la calibración del modelo * Las relaciones predictor-predictando a menudo son no-estacionarias * La elección de las variables predictoras afecta a los resultados * Los métodos más sencillos sólo pueden proporcionar proyecciones con una resolución mensual * Baja frecuencia de variabilidad climática

3.3. DESAGREGACIÓN ESTADÍSTICA BASADA EN LA REGRESIÓN LINEAL

La desagregación estadística basada en la regresión lineal es una de las formas más comunes de desagregación de proyecciones futuras, desde un modelo NWP a un lugar de interés. Se debe contar con tres tipos diferentes de datos:

1. Datos observados de la variable de interés (predictando).
2. Pronóstico del pasado de las variables atmosféricas (predictores) para estimar modelos estadísticos de predicción; periodo *reforecast* de calibración del modelo.
3. Pronósticos del pasado de las variables atmosféricas usados para validar el modelo en un periodo posterior; periodo *operational* de validación del modelo.

En la desagregación estadística de regresión se construyen modelos por medio de una regresión lineal, simple o multivariada. Primero, se utilizan los datos observados junto con los datos de las variables atmosféricas del periodo *reforecast* para generar modelos de regresión, en segundo lugar, se utilizan los datos de las variables atmosféricas del periodo *operational* para la validación de los modelos de regresión lineal obtenidos en el paso anterior.

A continuación, se describe los elementos que componen una regresión lineal simple o multivariada, con sus limitaciones correspondientes.

3.3.1. REGRESIÓN LINEAL

Para un caso determinista la regresión lineal simple o multivariada, busca relacionar bajo ciertos criterios y supuestos, dos o más variables como lo muestra la ecuación (1), donde el predictando proyectado ($\hat{Y}$) está en función de los predictores ($X_1, \dots, X_n$) con los parámetros $\beta_0, \beta_1, \dots, \beta_n$.

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + \epsilon \quad (1)$$

La salida del modelo incluye un error aleatorio (ϵ) normalmente distribuido ($E(\epsilon) = 0$; $Var(\epsilon) = \sigma_\epsilon^2$). Para más detalle de sobre la regresión lineal ver Anexo A.

➤ CORRELACIONES Y ERROR.

Los coeficientes de correlación permiten obtener información sobre el nivel de relación que poseen el predictando con los predictores. Esto puede darse entre dos series de datos directamente, coeficiente de correlación Pearso, o el efecto de una tercera variable sobre un par definido, coeficiente correlación parcial.

Coeficiente de correlación de Pearson:

El coeficiente de correlación de Pearson $r_{(Y,X)}$ para dos series de datos Y y X, se determina mediante la siguiente ecuación:

$$r_{(Y,X)} = \frac{1}{N-1} \sum_{i=1}^N \left(\frac{y_i - \mu_Y}{\sigma_Y} \right) \left(\frac{x_i - \mu_X}{\sigma_X} \right) = \frac{cov(Y, X)}{\sigma_Y \sigma_X} \quad (2)$$

Donde N es la cantidad de observaciones, y_i y x_i son los valores de las series temporales Y y X ; μ es la media y σ la varianza de las series (MathWorks, 2019).

Coefficiente de correlación parcial:

La correlación parcial $r_{(Y,X_1 | X_2)}$ entre las variables Y y X_1 dada una tercera variable X_2 se calcula de acuerdo a la ecuación (3) (Kim, 2015):

$$r_{(Y,X_1 | X_2)} = \frac{r_{(Y,X_1)} - r_{(Y,X_2)} r_{(X_1,X_2)}}{\sqrt{1 - r_{(Y,X_2)}^2} \sqrt{1 - r_{(X_1,X_2)}^2}} \quad (3)$$

Error absoluto medio (MAE)

El error absoluto medio (MAE, siglas en ingles), se determina como el promedio de los errores aleatorios ϵ , entre una serie de datos Y y los valores evaluados del modelo de regresión con resultado $\hat{Y}$.

$$MAE = \frac{1}{N} * \sum_{i=1}^N |y_i - \hat{y}_i| = \frac{1}{N} * \sum_{i=1}^N |\epsilon| \quad (4)$$

Donde N corresponde a la cantidad de datos, y_i al valor de la serie Y , del mismo modo $\hat{y}_i$ al valor de la serie $\hat{Y}$.

➤ TEST DE SIGNIFICANCIA

Los test de significancia se utilizan para comparar variables entre distintas series de datos. Se ha demostrado que la siguiente variable estandarizada T tiene una distribución t con $n - 2$ grados de libertad:

$$T = \frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}} \quad (5)$$

Esta variable se utiliza bajo la hipótesis que $\beta_1 = 0$. Comparando el valor critico de la distribución t para un nivel de significancia $\alpha(t_{\alpha/2, n-2})$ con la variable T ,

- Hipótesis nula $H_0: \beta_1 = 0$
- Valor estadístico de prueba $T = \frac{\hat{\beta}_1 - 0}{s_{\hat{\beta}_1}}$
- Hipótesis alternativa $H_0: \beta_1 \neq 0$
- Región de rechazo para el nivel de significancia: $T \geq t_{\alpha/2, n-2}$

Donde α es la probabilidad de ocurrencia. Cuando el valor estadístico estimado de la prueba T es mayor que el valor crítico de la t-distribución $t_{\alpha/2, n-2}$, la hipótesis nula $H_0: \beta_1 = 0$ es rechazada y la hipótesis alternativa que $H_0: \beta_1 \neq 0$ es validada. En otras palabras, la variable predictiva X tiene una fuerte relación con Y .

3.3.2. SELECCIÓN DE PREDICTORES

Definida la regresión lineal y distintos parámetros descriptivos de estas mismas, se procede a describir los criterios y métodos de selección predictores (variables regresoras) para ser agregadas a la regresión lineal.

➤ CRITERIOS DE SELECCIÓN DE PREDICTORES

Estos criterios se basan en la maximización del modelo de regresión incorporando el mínimo número de variables posibles. Algunos de estos criterios se presentan a continuación (Huth, 1999):

A) Criterio de significancia

Este criterio planea la incorporación de una variable al modelo de regresión, considerando solo aquellas variables que contribuyen significativamente al modelo. Esta contribución se obtiene al comparar los coeficientes de correlación parcial para las posibles variables a incorporar en el modelo, bajo la hipótesis de independencia entre variables. Entre estos criterios se destaca:

- *Probabilidad F*: una variable se integra al modelo si el nivel crítico del contraste entre la hipótesis de independencia con su coeficiente de correlación es menor a 0,05 (probabilidad de entrada). Se rechaza si este valor es mayor a 0,10 (probabilidad de salida).
- *Valor F*: la variable se integra si el estadístico F utilizado para el contraste entre la hipótesis de independencia con su coeficiente de correlación es mayor a 3,84 (valor de entrada) y queda fuera si el valor es menor a 2,71 (valor de salida).

B) Criterio de tolerancia

Superado el criterio de significancia, una variable se incorpora al modelo solo si cumple con el nivel de tolerancia establecido por defecto (0,0001).

➤ MÉTODOS DE SELECCIÓN DE PREDICTORES

Los métodos de selección por pasos (*stepwise*) son los más aceptados, pero no quiere decir que sean los mejores (Pardo & Ruiz, 2005). El nombre de estos métodos recae en ir incorporando o eliminando variables paso a paso dependiendo de que éstas cumplan los criterios de selección.

A) Método hacia adelante (*forward*)

La incorporación de las variables se realiza una a una en el modelo de regresión lineal multivariado. En el primer paso se selecciona la variable independiente (esta variable debe cumplir los criterios mencionados) que posea el nivel más alto de correlación con la variable dependiente.

Los siguientes pasos seleccionan la variable mediante el coeficiente de correlación parcial, este procedimiento se detiene cuando no hay variables que superen el criterio de entrada. Se utiliza como criterio de entrada el tamaño, en valor absoluto, del coeficiente de correlación parcial, lo cual es equivalente a seleccionar la variable con menor probabilidad de F o mayor valor de F .

B) Método hacia atrás (*backward*)

Este método comienza incluyendo todas las variables independientes en el modelo de regresión, para luego proceder a la eliminación una a una. La primera variable eliminada es la que no cumpla los criterios de entrada y, además, posea el menor peso en la disminución de R^2 . En los pasos sucesivos se van eliminando las variables con coeficientes de correlación no significativos y siempre en orden inverso a cómo cambia R^2 . Este procedimiento se detiene cuando no quedan variables que eliminar.

C) Método bidireccional (*bidirectional*)

Este método es una combinación de los dos métodos anteriormente descritos. A diferencia de los otros puede iniciar con un set limitado de variables, procediendo a agregar o eliminar variables. Cada vez que se agregue una nueva variable se analiza que las variables independientes del modelo cumplan o no con los criterios de salida. Este proceso se detiene cuando no quedan variables que agregar o eliminar.

3.3.3. SUPUESTOS DEL MODELO DE REGRESIÓN LINEAL

Las regresiones lineales en general, deben cumplir cinco supuestos para garantizar su validez. Los supuestos son: linealidad, independencia, homocedasticidad, normalidad y no-colinealidad (Pardo & Ruiz, 2005).

➤ LINEALIDAD

Este primer supuesto indica que debe existir una relación lineal entre las variables predictoras y la variable a predecir. Ha sido, regularmente estudiado por gráficos de correlación parcial, para eliminar la incidencia de las otras variables regresoras del modelo. Cuando un modelo de regresión lineal no cumple con este supuesto se le llama error de especificación. Este error puede aparecer por omisión de variables predictoras importantes, no linealidad de las variables, incorporación de variables predictoras no relevantes en el modelo, no aditividad, etc.

➤ INDEPENDENCIA

Se habla de independencia cuando los errores (residuos) del modelo sean independientes entre sí, en el caso contrario se dice que los residuos se encuentran autocorrelacionados. El estadístico de *Durbin-Watson* (6) proporciona información sobre el grado de independencia existente:

$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad (6)$$

donde $e_i = Y_i - \hat{Y}_i$. El estadístico DW varía entre 0 y 4, siendo igual a 2 cuando son completamente independientes. Los valores menores a 2 implican una autocorrelación positiva y los valores mayores a 2 implican una autocorrelación negativa.

➤ HOMOCEDASTICIDAD

El supuesto de homocedasticidad consiste en que los errores (residuos) tengan una varianza constante, siendo fundamental para la funcionalidad de los modelos. La homocedasticidad describe una situación en que el error es el mismo en todos los valores de las variables independientes. Por el contrario, la heteroscedasticidad está presente cuando el tamaño del error difiere para distintos valores de las variables independientes.

Para el análisis de este supuesto se utiliza la prueba de Breush-Pagan, si el test-F o el p-valor confirma que las variables independientes son significativas, entonces se puede rechazar la hipótesis nula de homocedasticidad.

➤ NORMALIDAD

El supuesto de normalidad hace referencia a que las variables sigan la ley normal de distribución. Para cada uno de los valores de las variables predictoras (variables independientes) o la combinación de valores de éstas, los residuos (e_i) se distribuyen normalmente un ($E(e) = 0 ; Var(e) = \sigma_e^2$). Existen diversos métodos para este test, como el test Kolmogorov-Smirnov y el método Shapiro-Wilk.

En el test Shapiro-Wilk, la hipótesis nula se rechaza si W es demasiado pequeño. El valor de W puede variar entre 0 y 1.

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (7)$$

Donde $x_{(i)}$ es el número que ocupa la i -ésima posición en la muestra (con la muestra ordenada de mayor a menor), $\bar{x}$ es la media muestral, las variables a_i se calculan mediante:

$$(a_1, \dots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$$

Donde

$$m = (m_1, \dots, m_n)^T$$

Siendo $m_1, \dots, m_n$ los valores medidos del estadístico ordenado, considerados como variables aleatorias independientes e idénticamente distribuidas, y normalmente distribuidas. Finalmente, V denota la matriz de covarianzas del estadístico de orden.

En estadística la hipótesis nula es una afirmación sobre un parámetro que obtenido de la población (como la media o desviación estándar) y se representa con H_0 . Es una hipótesis de partida que no se rechaza a menos que los datos de la muestra parezcan evidenciar que es falsa, es decir, que hay una relación entre los parámetros o fenómenos analizados. Para este caso, se define como hipótesis nula que la población está normalmente distribuida, si el p-valor es menor a α (valor de significancia) entonces la hipótesis nula es rechazada (se concluye que los datos no están distribuidos normalmente). Si el p-valor es mayor a α , se concluye que no se puede rechazar dicha hipótesis. Lo mismo ocurre para los otros supuestos que utilizan p-valor.

➤ NO-COLINEALIDAD

Bajo este supuesto no debe existir una relación lineal exacta entre ninguna de las variables independientes. El no cumplimiento de este supuesto causa colinealidad o multicolinealidad, generando que la influencia de cada predictor en la correlación no pueda ser distinguida al quedar su efecto solapado al de las otras variables.

El factor de incremento de varianza (FIV) es un indicador comúnmente ocupado para estudiar el cumplimiento de este supuesto.

$$FIV_i = \frac{1}{1 - R_i^2} \quad (8)$$

Donde R_i^2 es el coeficiente de determinación del modelo de regresión lineal para cada predictor. Si $FIV(\hat{\beta}_i) > 10$ implica un alto nivel de multicolinealidad.

3.4. DESCRIPCIÓN DE LAS VARIABLES ATMOSFÉRICAS

Los modelos Numéricos de Predicción del Tiempo (NWP) entregan datos de variables atmosféricas, las que son utilizadas como input en la predicción en modelos de desagregación estadística. La elección de las variables del NWP es una de las etapas más desafiantes en el desarrollo de cualquier modelo estadístico. Los NWP entregan una gran cantidad de variables, y la elección de las variables a trabajar determina en gran medida el escenario climático reducido. Es importante que las variables predictoras candidatas sean física y conceptualmente acordes a la variable que se desea predecir, para la precipitación se consideran variables que describan cuatro aristas del fenómeno: la circulación atmosférica, el espesor de las capas, la estabilidad y el contenido de humedad (Wilby R., 2002).

Para la comprensión del fenómeno físico que ocurre en la precipitación, se requiere ahondar en la climatología sinóptica (análisis de los elementos atmosféricos), siendo de un interés particular la influencia de las nubes en la relación física entre la precipitación y las variables atmosféricas del modelo CFS. Las variables atmosféricas relevantes con la precipitación son: la temperatura, la

humedad y la presión (Wu, 2012). Adicionalmente, en el modelo CFS también podemos encontrar la variable altura geopotencial, la cual determina la altura a la que se encuentra un determinado nivel de presión.

A continuación se describiremos distintas variables atmosféricas que se utilizarán como potenciales variables predictoras de la precipitación, se presentan desde un punto de vista físico y son seleccionadas de acuerdo a la disponibilidad de variables de CFSv2. Las variables son: humedad específica, temperatura superficial y atmosférica, presión atmosférica y altura geopotencial.

3.4.1. HUMEDAD ESPECÍFICA

La precipitación se produce cuando existen nubes con suficiente vapor de agua y condiciones climáticas que la propician. La variable humedad indica la cantidad de vapor de agua, ayudando a describir la termodinámica de los procesos ocurridos en la atmósfera (Hertig, 2019). La variable de humedad específica entregada por CFSv2, da a conocer la razón entre la cantidad de kilos de vapor de agua y el total de kilos en un cierto volumen de aire. Puesto que cada variable del modelo CFSv2 la podemos encontrar a diferentes altitudes, el siguiente paso es identificar la altitud más apropiada.

3.4.2. TEMPERATURA SUPERFICIAL Y ATMOSFÉRICA

De acuerdo con Wilby (2012) se debe contar con variables que estén relacionadas con la estabilidad atmosférica. En respuesta a esto, Anandhi (2011) indica que la temperatura atmosférica está asociada con la estabilidad termodinámica local.

Se pueden encontrar variables de temperatura superficiales en el modelo CFSv2, éstas son: temperatura superficial (Tmpsfc), temperatura a 2m (TMP2m) o temperatura a 1000 hPa de altitud (T1000). También existen variables de temperatura atmosférica en CFSv2, las que se encuentran a los 850 hPa, 700 hPa, 500 hPa, etc. (T850, T700, T500 respectivamente). Todas las variables de temperatura superficial y atmosférica entregadas por CFSv2 pueden ser observadas en mayor detalle en la Tabla 9.

La temperatura atmosférica a la altura de la nube cambia su valor en presencia de una nube que contiene agua precipitable, puesto que la temperatura atmosférica está relacionada con la retención de agua en la atmósfera (Trenberth, 2003). Para comprender cómo interactúa la temperatura superficial con la precipitación, se analizará la Figura 7, donde se observa el desplazamiento horizontal de una nube que contiene vapor de agua precipitable.

Debido a la presencia de una nube de temperatura (T_c) y emisividad (ϵ_c), se genera una transferencia de flujo (F) con la superficie, aumentando la temperatura superficial (Liou, 2002). En resumen, una nube con agua precipitable hará cambiar el valor de la temperatura atmosférica, según la altura a la que se encuentre. De acuerdo con la Figura 7, la aparición de una nube con vapor de agua precipitable también generará un cambio de temperatura en la superficie.

Cabe señalar que de acuerdo con Trenberth (2005) existe una relación muy estrecha entre el vapor de agua precipitable y la temperatura de la superficie del mar, esto es de gran relevancia, considerando que la región de análisis se encuentra frente al océano Pacífico.

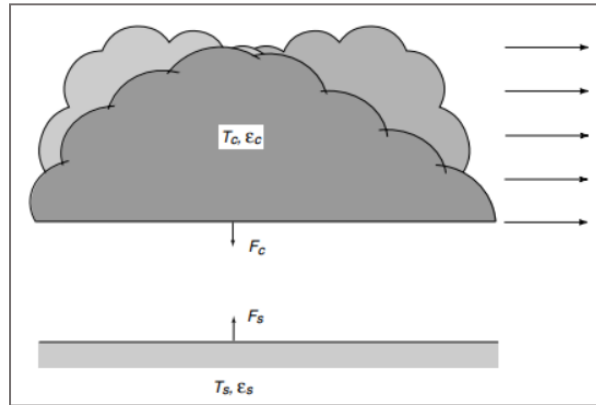


Figura 7: Configuración simple de una nube que se mueve sobre una superficie.
Fuente: (Liou, 2002)

3.4.3. PRESIÓN ATMOSFÉRICA

La formación de nubes ocurre con el ascenso de parcelas de aire, donde la flotabilidad existe debido a una diferencia de presión alrededor de la nube. Si una parcela de aire es menos densa que el aire atmosférico circundante, la presión bajo la parcela será menor que la de sus alrededores (ver Figura 8), generando un gradiente de presión que acelera el aire ambiental hacia la base de la parcela flotante (Houze, 1993). Es así, que se detecta la presencia de una nube, por medio de un cambio de presión.

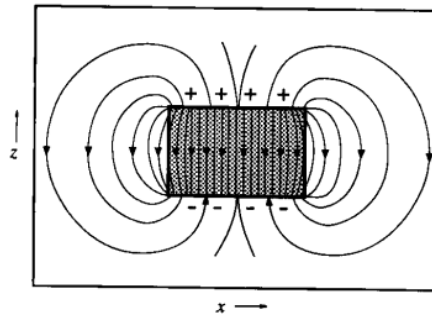


Figura 8: Campo vectorial de fuerza de gradiente de presión de flotabilidad para una parcela flotante en el plano x-z. Fuente: (Houze, 1993)

El modelo CFSv2 utilizado en este estudio cuenta con dos variables de presión: *presión media al nivel del mar* (PRMSL, siglas en ingles) y *presión superficial* (PRESSFC, siglas en ingles). La presión superficial es distinta para terrenos ubicados en distintas altitudes, es por ello que reducirla al nivel del mar (PRMSL) resulta de mayor utilidad para comparar puntos en distintas zonas geográficas (Wallace & Hobbs, 2006).

De acuerdo con Wilby (2002) se debe contar con variables que estén relacionadas, con la circulación atmosférica. La presión reducida al nivel del mar entrega información sobre la direccionalidad de los flujos y forzamientos dinámicos en la atmósfera (Borén, 1995), por lo que será una variable por considerar durante este estudio.

3.4.4. ALTURA GEOPOTENCIAL

Hertig (2019), indica que las variables predictoras de la precipitación normalmente usadas son la presión a nivel del mar y las alturas geopotenciales. La altura geopotencial nos indica la altitud a la cual se encuentra un determinado nivel de presión, por ejemplo, la altitud en metros a la que se encuentra el nivel de presión 850 hPa (Z850 hPa, altura geopotencial a 850 hPa).

La altura geopotencial muestra como un nivel de presión se eleva o desciende, pudiendo deberse a los fenómenos físicos mostrados en la Figura 8, en este caso se relacionaría con el movimiento de las nubes. La altura geopotencial es una variable de presión, y dado que mide distancia, puede determinar el espesor de dos niveles de presión (espesor de capas).

En específico, la variable de altura geopotencial a 850 hPa se encuentra dentro de la nube o por debajo de ésta. En la Figura 8 se observa que en presencia de una nube existe una baja de presión tanto dentro como bajo la nube, por lo que, si disminuye la presión, la altura geopotencial a la que se buscan los 850 hPa será a una menor altitud, es decir, disminuiría el valor de la variable Z850. Por tanto, una disminución en la altura geopotencial indica la presencia de nubes y la posibilidad de precipitación. De forma análoga, en presencia de nubes debido a precipitación, la altura geopotencial tendrá un valor más bajo, existiendo una relación inversa entre la presencia de nubes con posibilidad de precipitación y la variable Z850.

3.4.5. ESTUDIOS RELACIONADOS

A continuación, se presentarán estudios de desagregación estadística que se han realizado en el mundo y en Chile.

En primer lugar, se encuentra la investigación realizada en Toronto, Canadá (Wilby R. , 2002) sobre desagregación estadística, en la cual se muestra la metodología y aplicación del *software* SDSM versión 4.1 para una evaluación de impacto local del cambio climático. En este estudio, se utilizan once variables candidatas a predictores (ver Tabla 3), con datos para la calibración del modelo entre los años 1981-1985, y para su validación en el periodo 1986-1990. En la Tabla 3, se observan los coeficientes de correlación parcial que se obtuvieron para cada una de las variables de los modelos generados.

De acuerdo con los resultados de correlación, se observa, que el mejor de los resultados obtuvo una correlación de 0,40, en la variable atmosférica de viento: Componente Meridional del Flujo de Aire Geostrófico (hPa). Según la Tabla 5, este resultado corresponde a una correlación baja, lo que, sin embargo, se considera un resultado aceptable para la predicción de la precipitación.

*Tabla 3: Variables y correlación parcial de una desagregación estadística en Canadá.
(Wilby R. , 2002).*

Variable predictora	Sigla	Correlación parcial r
Valor del predictando en el día anterior	Lag-1	0,03
Ocurrencia de precipitación ('1'=wet, '0'=dry)	Wet	n/a
Promedio diario de temperatura a 2m (°C)	T _{mean}	-0,04
Humedad especifica cerca de la superficie (gm/kg)	SH	0,06

Humedad relativa cerca de la superficie (%)	RH	-0,01
Presión media a nivel del mar (hPa)	Mslp	-0,08
Componente zonal del flujo de aire geostrófico (hPa)	U	-0,04
Componente meridional del flujo de aire geostrófico (hPa)	V	0,40
Flujo de aire geostrófico (hPa)	F	0,04
Vorticidad (hPa)	Z	0,20
Altura geopotencial a 500 hPa (m)	z500	-0,05

Con mejores correlaciones, se presenta un estudio de desagregación estadística de precipitación en Tailandia (Anuchaivong, 2017). Este estudio tiene como objetivo, obtener pronósticos futuros de la precipitación, considerando la importancia del recurso hídrico en el medio ambiente y la agricultura. Se utilizaron datos de precipitación observada de 30 estaciones desde el año 1979 hasta el 2010. Los datos de las variables atmosféricas (predictores) fueron recopilados desde el sistema de pronóstico del clima NCEP (CFSR y CFSv2). Las correlaciones parciales obtenidas para cada una de las variables se observan en la Tabla 4. En este punto se destaca, que los resultados obtenidos son positivos, puesto se ha utilizado el coeficiente de determinación (R^2), por lo cual no se aprecia si los resultados obtenidos son directa o inversamente proporcionales. Finalmente, se entrega un criterio de niveles de coeficiente de correlación de Pearson (r) según el valor obtenido (Tabla 5).

Tabla 4: Correlación parcial entre variables predictoras y la precipitación en una desagregación estadística en Tailandia (Anuchaivong, 2017).

Variable predictora	Sigla	Correlación (R^2)
Temperatura a 850 hPa	T850	0,79
Humedad específica a 850 hPa	Q850	0,75
Altura geopotencial a 850 hPa	G850	0,80
Presión media al nivel del mar	MSLP	0,79

Tabla 5: Niveles de correlación. Fuente: (Anuchaivong, 2017)

Valor de r	Nivel de correlación
0,90 - 1,00	Muy alta
0,70 - 0,90	Alta
0,50 - 0,70	Media
0,30 - 0,50	Baja
0,00 - 0,30	Muy baja

Por otra parte, en Chile se han realizado estudios utilizando desagregación estadística. Entre ellos se encuentra el estudio enfocado en la cuenca río *Cachapoal*, VI Región (Figueroa, 2008). Este estudio, busca simular la escorrentía superficial y para ello se construyó y calibró un modelo utilizando un periodo de 35 años (1971 al 2005). Para esto, se contó con datos de la precipitación observada recogidos desde la dirección general de aguas (DGA), y datos de 26 variables predictoras obtenidos desde el NCEP.

El proceso de desagregación se realizó mediante el software SDSM 4.2. Luego, tras el análisis se determinaron cuatro variables que obtuvieron una mejor correlación parcial entre los predictores y la precipitación observada (ver Tabla 6). Todos los resultados obtenidos se encuentran en un rango de correlación muy baja (0,00 – 0,30 Tabla 5).

Al comparar estos resultados con los estudios anteriores, se establece que existe una correlación positiva con el viento, por otra parte, se añade una correlación positiva con la humedad, y una correlación negativa con las variables de presión (mslp y p500 de acuerdo con Tabla 6). Cabe señalar, que el software SDSM 4.2 utiliza una función de transformación de los datos, entre las que se encuentran, exponencial, logarítmica, logaritmo natural, etc. En este caso se utilizó una transformación logarítmica para las variables de precipitación, puesto que éstas no tienen un comportamiento normal, además indica que estos procesos de transformación pueden afectar los resultados en la calibración de los modelos.

Tabla 6: Variables y correlación parcial en una desagregación estadística en la VI Región, Chile (Figueroa, 2008)

Variable predictora	Código NCEP	Correlación parcial r
Presión media a nivel del mar	mslp	-0,25
Velocidad zonal superficial	p_u	0,17
Altura geopotencial a 500 hPa	p500	-0,14
Humedad relativa a 500 hPa	r500	0,16

Otro estudio de desagregación estadística en Chile realizado en la subcuenca Alto-Elqui (Souvignet, 2010), en la IV Región del país. El objetivo se centró en evaluar posibles impactos del cambio climático en la zona de estudio, por medio de una desagregación estadística de la precipitación y la temperatura, utilizando el software SDSM versión 4.1, junto con otros análisis.

En este estudio para la desagregación se utilizaron datos de variables atmosféricas de NCEP y para la precipitación observada se consideraron datos de tres estaciones meteorológicas, contemplando 10 años de datos para el periodo de calibración (1980 – 1990), y 10 años para su validación (1991-2000). Uno de los resultados del estudio arrojó valores muy bajos o bajos de correlación parcial (ver Tabla 5) entre las variables atmosféricas y la precipitación observada (ver Tabla 7), para las distintas estaciones meteorológicas y variables en estudio.

Cabe señalar que la mejor correlación parcial en el estudio de Souvignet (2010) se determinó para la estación Huanta con la variable de altura geopotencial a 850 hPa y con un valor de -0,311. Se destaca que la correlación de la altura geopotencial a 850 hPa es negativa al igual como se mostró en el estudio anterior, evidenciando que esta variable es inversamente proporcional a la precipitación observada.

Tabla 7: Variables y correlación parcial en una desagregación estadística en la IV Región, Chile (Souvignet, 2010).

Estación meteorológica	Variable predictora	Código NCEP	Correlación parcial r
Huanta	Altura geopotencial a 850 hPa	p850	-0,311
	Humedad relativa a 500 hPa	r500	0,178
Juntas del Toro	Fuerza del flujo de aire a 850 hPa	p_8f	0,166
	Altura geopotencial a 850 hPa	p850	-0,230
	Humedad relativa a 500 hPa	r500	0,169
La laguna	Velocidad zonal a 850 hPa	p8_u	0,287
	Humedad relativa a 500 hPa	r500	0,231
	Humedad relativa a 850 hPa	r850	-0,080

Un tercer estudio que destacar en Chile consiste en una investigación de desagregación estadística en las Regiones Metropolitana, V y VI (Vega, 2018). Este trabajo tuvo como objetivo proponer un modelo estadístico multivariado para pronosticar y corregir la precipitación en la zona de estudio. Se utilizó una regresión lineal mediante el software SDSM versión 4.1 y regresiones no lineales mediante el software Eureka. Los datos de precipitación observada fueron recopilados de 76 estaciones pluviométricas desde la dirección general de aguas (DGA) y la dirección meteorológica de Chile (DMC), en un periodo 1996 al 2016 para la calibración y validación de los modelos.

Las variables atmosféricas utilizadas se seleccionaron desde el sistema de pronóstico del clima (NCEP-CFSv2), éstas fueron, precipitación (PP), humedad específica a 700 hPa (Q700) y altura geopotencial a 850 hPa (Z850). En general se obtuvieron buenas correlaciones parciales en un análisis mensual, se destacan los resultados en la estación Santo Domingo, obteniendo correlaciones de hasta 0,92 entre Q700 y la precipitación observada para los meses desde abril. En la misma estación para el caso de Z850 se obtuvieron correlaciones negativas de hasta -0,77. Los resultados reafirman la proporcionalidad directa de la humedad y la proporcionalidad inversa de las variables de presión.

Recientemente, se realizó una investigación en todo Chile continental (Araya, 2020), con el objetivo de generar proyecciones de cambio climático. Uno de los análisis fue una desagregación estadística de la precipitación, utilizando un método analógico. Los datos de precipitación observada fueron recogidos de 427 estaciones distribuidas en todo el país, en un periodo 1980-2015 para su calibración y validación. Como predictores se utilizaron doce variables recopiladas de NCEP/NCAR, combinando las variables en nueve formas distintas. Los resultados arrojaron que la mejor combinación para la predicción de la precipitación en el país es la temperatura de aire superficial (Tas), la humedad específica a 700 hPa (Q700) y las alturas geopotenciales a 500 y 850 hPa (Z500 y Z850 respectivamente). Para esta combinación de variables como predictor de la precipitación, se determinaron coeficientes de correlación entre 0,6 a 0,8 en las regiones Metropolitana, V y VI.

3.5. ANÁLISIS CLUSTER

El Análisis Cluster (AC) es un método estándar de análisis multivariado capaz de agrupar elementos con una gran cantidad de información, sin imponer restricciones previas. Formando grupos (clusters) homogéneos de acuerdo con la similitud o similaridad entre los elementos (Castro, 2012). En el estudio de Castro (2012) también se menciona que el AC puede ser usado para la clasificación de estaciones meteorológicas, lo cual es el interés del presente estudio. Se describe también, que las estaciones que se agrupan solas son debido a una condición atípica con respecto a las estaciones vecinas o que presentan deficiencias en sus series de datos.

Seleccionar la cantidad de clusters a formarse no posee un criterio estadístico interno (como los test de significancia), implicando que no exista un procedimiento estándar para determinar dicha cantidad (Hair, 1999). Pero de igual manera, se han desarrollado diversos procesos matemáticos o gráficos para determinar un número óptimo de clusters, por ejemplo, mediante el método de Gap (Oliva, 2015) o simplemente utilizando el criterio del investigador (De la Fuente, 2011). Existen varias técnicas de AC, las que se pueden clasificar en AC jerárquico y no jerárquico.

3.5.1. MÉTODO JERÁRQUICO

El AC jerárquico posee una estructura en forma de árbol, existiendo dos tipos: los aglomerativos y los divisivos (Castro, 2012). Los métodos aglomerativos comienzan con n objetos y cada uno representa a un cluster, los que comienzan a agruparse de manera secuencial hasta que los n elementos son parte de un único cluster, mientras que los métodos divisivos siguen el sentido inverso (Oliva, 2015).

Existen diversos algoritmos jerárquicos aglomerativos que describen distintas características de los datos, difiriendo en la determinación de la medida de distancia o similaridad entre los clusters, dentro de los cuales se tienen (Oliva, 2015):

- (i) *Método Single linkage* (vecino más cercano), en este método la distancia entre los clusters queda definida por la mínima distancia entre dos elementos pertenecientes a distintos clusters, es decir, el par de clusters que comparta los dos elementos más cercanos se unirán en el actual paso.

Este método no exige restricciones en la forma de grupos y sacrifica el poder obtener clusters compactos, a cambio de la capacidad de detectar grupos alargados e irregulares.

- (ii) *Método Complete linkage* (vecino más lejano), define como medida de distancia entre los clusters la distancia entre los elementos más lejanos pertenecientes a distintos clusters. De esta manera, se unirán los clusters que presenten la mínima medida de distancia.

Este método está sesgado hacia la formación de clusters de diámetros similares, pudiendo verse distorsionado por valores atípicos moderados.

- (iii) *Average linkage*, aquí la distancia o similitud entre los clusters está definida como la media aritmética de cada uno de los elementos del primer cluster con cada elemento del segundo cluster.

En este método se tienden a unir grupos que presentan una baja varianza, siendo, levemente sesgado hacia la formación de clusteres de igual varianza.

- (iv) *Método del centroide y de la mediana*, en ambos métodos cada cluster queda determinado por una medida resumen de todos los elementos que conforman el grupo.

En el método del centroide, para un cluster i , que contiene n_i elementos con valor x_l , con $l = 1, \dots, n_i$, así el centroide m_i queda definido por:

$$m_i = \frac{1}{n_i} \sum x_l \quad (9)$$

Luego, se unirán los dos clusters (C_i, C_j) que presenten la menor distancia, y con un nuevo centroide m_t :

$$m_t = \frac{n_i m_i + n_j m_j}{n_i + n_j} \quad (10)$$

El método del centroide tiende a ser más robusto frente a valores atípicos que otros métodos jerárquicos. Sin embargo, la forma de determinar el nuevo centroide puede causar que un gran cluster absorba a uno más pequeño, para ello se creó el método de la mediana, determinando el nuevo centroide bajo el supuesto que los clusters son de igual tamaño.

$$m_t = \frac{m_i + m_j}{2} \quad (11)$$

- (v) *Método de Ward*, tiene por objeto unificar grupos sin que exista un drástico aumento en la varianza. Para ello, se busca minimizar la suma de cuadrados dentro de lo que sería el nuevo cluster $I_{C_i C_j}$, compuesto por los cluster C_i y C_j .

$$I_{C_i C_j} = \frac{n_i n_j}{n_i + n_j} \|m_i - m_j\|^2 \quad (12)$$

Este procedimiento, tiende a unir grupos con centroides cercanos y con pocos elementos, produce clusters esféricos y bien balanceados. También, es sensible a los valores atípicos. Sin embargo, de acuerdo con Hair (1999), está sesgado hacia la formación de clusters con el mismo número de observaciones.

3.5.2. MÉTODO NO JERÁRQUICO

El AC no jerárquico no requiere la construcción de árboles, puesto que está diseñado para la clasificación de individuos en un número predefinido de k clusters, donde frecuentemente son k -medias, k -medianas y k -modas (Castro, 2012). Para este procedimiento, se deben seleccionar puntos “semillas” como centros de los conglomerados, pudiendo ser puntos aleatorios o especificados (no aleatorios) (Hair, 1999).

En el caso de utilizar puntos semilla especificados se obtienen varias ventajas por sobre los métodos jerárquicos, con resultados menos susceptibles a valores atípicos y a la medida de distancia, pero dependen de la capacidad del usuario en la selección de estos puntos, seleccionando de acuerdo a bases prácticas, objetivas y teóricas. Por otro lado, el uso del método no jerárquico con puntos de semilla aleatorio no presenta estas ventajas, por tanto, el uso de esta técnica con puntos de semilla aleatorios es bastante inferior a los métodos jerárquicos (Hair, 1999). De acuerdo con el procedimiento de los métodos no jerárquicos, éstos entregan una convergencia óptima local, pero la solución óptima global no se garantiza (Oliva, 2015).

4. ÁREA DE ESTUDIO Y CONJUNTOS DE DATOS

4.1. ÁREA DE ESTUDIO Y CLIMA

Las estaciones meteorológicas que miden la precipitación observada diaria, se encuentran ubicadas en la parte continental de Chile, limitada al norte y al sur por las latitudes $32,8^\circ$ y $34,3^\circ$ S, mientras que al este y al oeste está limitada por la longitud $71,8^\circ$ y $70,0^\circ$ O. Por otro lado, la grilla de las variables atmosféricas de CFS abarca una mayor área que la zona de las estaciones, bajo la idea de encontrar puntos mejor correlacionados en una zona más amplia. La grilla se encuentra limitada al norte y al sur por las latitudes $30,7^\circ$ y $36,4^\circ$ S; mientras que al este y al oeste está limitada por la longitud $76,0^\circ$ y $68,4^\circ$ O.

Las estaciones meteorológicas se encuentran ubicadas en tres regiones de Chile: la región de Valparaíso (V Región), la región Metropolitana y la región del Libertador Bernardo O’Higgins (VI Región). El interés en estas zonas surge por la gran densidad poblacional, particularmente en la región Metropolitana, donde habitan más de 7 millones de personas, correspondiente al 40% de la población nacional (INE, 2017). A continuación, se profundizará con respecto al clima en cada región de la zona de estudio.

La región de Valparaíso posee un clima templado mediterráneo, con algunas diferencias en cuatro zonas. Un clima de estepa cálida al norte del río Aconcagua, donde se encuentran precipitaciones alrededor de 150 a 200 mm al año. En la costa de la región, un clima mediterráneo templado costero, con una alta humedad relativa (75%) y abundantes precipitaciones, llegando a unos 450 mm al año. Desde el río Aconcagua hacia el sur se está en presencia de un clima mediterráneo templado cálido, donde las precipitaciones aumentan hacia la cordillera desde unos 200 hasta 300 mm. Finalmente, la cordillera de los Andes sobre los 3000 m de altura, cuenta con un clima frío de altura, predominando las precipitaciones sólidas (BCN, 2020).

En la región Metropolitana el clima corresponde principalmente a un tipo mediterráneo, con un corto invierno lluvioso. La precipitación anual es de unos 350 mm, aunque presenta una gran irregularidad, puesto que un año puede ser muy lluvioso mientras que el siguiente puede ser muy seco. La región se puede dividir en dos zonas con diferencias en el clima, en gran parte de la región encontramos un clima templado mediterráneo con una estación seca prolongada y un invierno bien delimitado, mientras que, hacia la cordillera de los Andes sobre los 3000 m de altura, existe un clima frío de altura (BCN, 2020). La variabilidad de la precipitación en la región se debe a la alternancia de los eventos de El Niño y La Niña (Garreaud, 2011).

En la región Del Libertador Bernardo O'Higgins predomina un clima templado mediterráneo, presentando tres distintas variaciones del clima según su topografía. El litoral presenta un clima templado nuboso, caracterizándose por una alta humedad y abundante nubosidad. En la depresión intermedia, encontramos un clima templado mediterráneo cálido. Mientras que el clima frío de altura comienza a los 3500 m (BCN, 2020). La precipitación promedio en la región es de unos 670 mm (Pizarro, 2008).

4.2. ESTACIONES METEOROLÓGICAS

Para describir la precipitación observada diaria se recopilieron datos de 29 estaciones pluviométricas (ver Tabla 8) proporcionados por la dirección general de aguas (DGA) y la dirección meteorológica de Chile, para el periodo 1995-2017. Las estaciones pluviométricas se encuentran en la zona central de Chile como se ilustra en la Figura 9.

Tabla 8: Estaciones pluviométricas en la zona de estudio. Fuente: Elaboración propia

N°	Estación	Latitud	Longitud	Elevación [msnm]	Institución
1	RAPEL	-33,95	-71,73	23	DGA
2	SANTO DOMINGO AD.	-33,66	-71,61	74	DMC
3	ESTERO PUANGUE RUTA 78	-33,66	-71,34	110	DGA
4	LOS GUINDOS	-33,89	-71,23	129	DGA
5	QUILLOTA	-32,90	-71,21	140	DGA
6	BARRERA LONCHA	-34,08	-71,20	150	DGA
7	RODELILLO	-33,05	-71,58	162	DGA
8	LO ROJAS	-32,81	-71,25	192	DGA
9	LOS PANGUILES	-33,44	-71,03	197	DGA
10	LAGUNILLAS	-33,45	-71,46	206	DGA
11	LITUECHE	-34,12	-71,73	245	DGA
12	COLTAUCO	-34,29	-71,08	265	DGA
13	EL VERGEL	-33,70	-70,92	330	DGA
14	LAGUNA ACULEO	-33,89	-70,88	378	DGA
15	PUDAHUEL SANTIAGO	-33,39	-70,80	480	DGA
16	RANCAGUA (CACHAPOAL-DCP)	-34,19	-70,75	524	DGA
17	QUINTA NORMAL SANTIAGO	-33,45	-70,68	529	DGA
18	TERRAZAS OFICINA CENTRAL DGA	-33,45	-70,64	565	DGA

19	HUECHUN EMBALSE	-33,09	-70,80	574	DGA
20	EULOGIO SANCHEZ AD	-33,46	-70,54	654	DGA
21	PIRQUE	-33,67	-70,59	666	DGA
22	QUEBRADA RAMON	-33,43	-70,53	731	DGA
23	LA OBRA RECINTO REMOS	-33,59	-70,49	799	DGA
24	CERRO CALAN	-33,40	-70,54	849	DGA
25	ANTUPIREN	-33,50	-70,52	909	DGA
26	SAN JOSE DE MAIPO RETEN	-33,64	-70,36	963	DGA
27	MAITENES BOCATOMA	-33,53	-70,26	1193	DGA
28	SAN GABRIEL	-33,78	-70,24	1256	DGA
29	EL YESO EMBALSE	-33,68	-70,09	2483	DGA

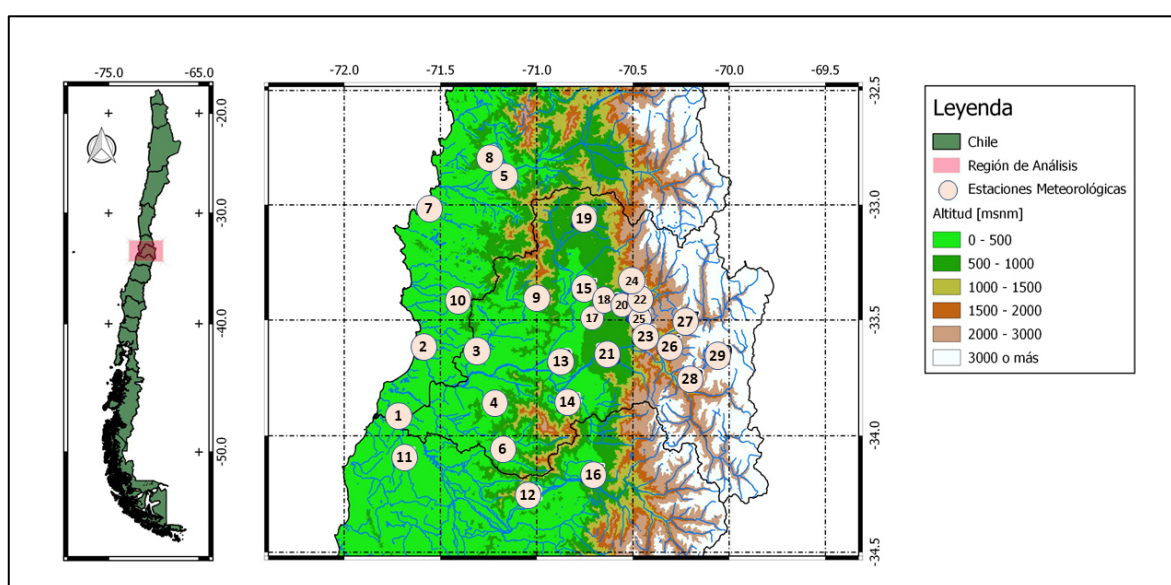


Figura 9: Estaciones pluviométricas. Fuente: Elaboración propia

4.3. VARIABLES ATMOSFÉRICAS

Para una desagregación estadística de regresión lineal con respecto a la precipitación, se requiere en primera instancia de la precipitación observada como predictando, y de la precipitación a nivel atmosférico como predictor, relacionando las variables mediante una regresión lineal simple. Luego, para una corrección del modelo se añaden más variables atmosféricas como predictores, dando lugar a una regresión lineal multivariada.

La precipitación medida en la atmósfera, así como las demás variables atmosféricas, se encuentran medidas en distintos puntos espaciales. Por un lado, están medidas en una grilla en latitud-longitud (ver Figura 10), como también pueden estar medidas a distintas alturas de presión, por ejemplo, a 1000 hPa, 850 hPa, 700 hPa, etc. (ver Tabla 9). Cabe señalar, que en este estudio se denominará variables atmosféricas a las variables recogidas desde NCEP, indistintamente si éstas son variables atmosféricas o de superficie.

Los datos diarios de las variables se obtuvieron de los modelos CFSv2 *Reforecast* y CFSv2 *Operational* (ver sección 3.1.1) desde NCEP. Los pronósticos de estas variables se encuentran en una grilla global de resolución aproximada 100 km (Saha, 2014). En la Tabla 9, se presenta la lista con dos grupos de variables, cada grupo tiene un tipo de archivo de descarga y grilla. A cada grupo de variables le corresponde una grilla: grilla NCEP o grilla gaussiana T126 (NCEP, 2011). Ambas grillas tienen una resolución similar, pero están desfasadas en latitud y longitud entre ellas, como se observa en la Figura 10.

Tabla 9: Variables modelo CFSv2. Fuente: (NCEP, 2011)

Resolución: 1.0° / Grilla NCEP			Resolución: 1.0° / Grilla gaussiana T126		
Descripción	Sigla	Unidad	Descripción	Sigla	Unidad
Altura geopotencial:			Presión:		
200 hPa	Z200	mgp	Media a nivel del mar	PRMSL	Pa
500 hPa	Z500	mgp	Superficie	PRESSFC	Pa
700 hPa	Z700	mgp	Viento zonal y meridional:		
850 hPa	Z850	mgp	10 metros	WND10m	N/m ²
1000 hPa	Z1000	mgp	Temperatura:		
Viento zonal y meridional:			Superficie	TMPSF	K
200 hPa	WND200	N/m ²	2 metros	TMP2m	K
250 hPa	WND250	N/m ²	Temperatura mínima a 2 metros de elevación	Tmin	K
500 hPa	WND500	N/m ²	Temperatura máxima a 2 metros de elevación	Tmax	K
700 hPa	WND700	N/m ²	Temperatura a nivel hibrido 1	TMPHY1	K
850 hPa	WND850	N/m ²	Humedad específica:		
1000 hPa	WND1000	N/m ²	2 metros	Q2m	kg/kg
Temperatura:			Precipitación:		
2 hPa	T2	K	Tasa total	PRATE	kg/m ² /s
50 hPa	T50	K	Tasa convectiva	CPRATE	kg/m ² /s
200 hPa	T200	K	Agua precipitable	PWATE	kg/m ²
250 hPa	T250	K	Radiación en la superficie:		
500 hPa	T500	K	Onda larga	ULWSF	W/m ²
700 hPa	T700	K	Onda corta	USWSF	W/m ²
850 hPa	T850	K	Radiación capa superior de la atmosfera		
1000 hPa	T1000	K	Onda larga	ULWTOA	W/m ²
Humedad específica:			Onda corta	USWTOA	W/m ²
500 hPa	Q500	kg/kg	Flujo de calor:		
700 hPa	Q700	kg/kg	Latente	LHTFL	W/m ²
850 hPa	Q850	kg/kg	Sensible	SHTFL	W/m ²
925 hPa	Q925	kg/kg	Suelo	GFLUX	W/m ²
Velocidad potencial:			Flujo solar descendente:		
200 hPa	CHI200	m ² /s	Visible difuso	VDDSF	W/m ²
850 hPa	CHI850	m ² /s	Infrarrojo cercano	NDDSF	W/m ²

Flujo solar en el cielo:			
Ascendente		CSUSF	W/m ²
Descendente		CSDSF	W/m ²

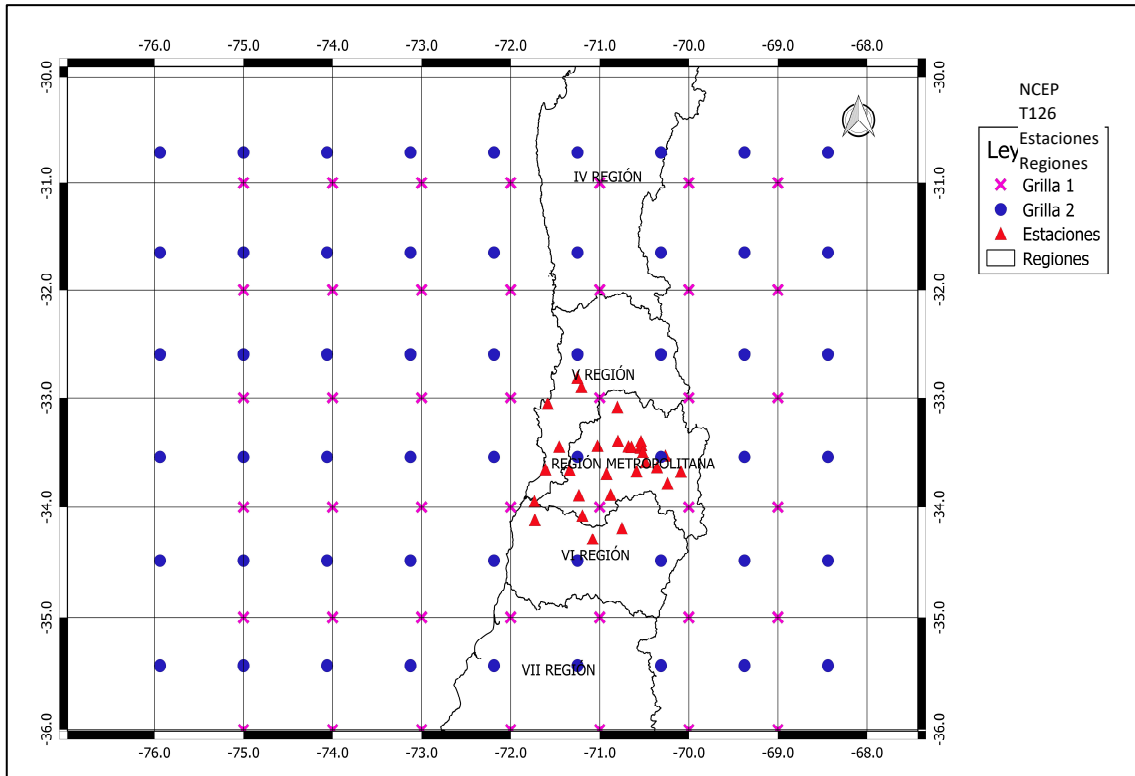


Figura 10: Grillas del modelo CFSv2, NCEP y Gaussiana T126. Fuente: Elaboración propia.

5. METODOLOGÍA

La metodología consta de tres etapas (ver Figura 11). La primera etapa, es un análisis cluster de la precipitación observada (P_{obs}) en las 29 estaciones meteorológicas en estudio (ver Tabla 8), utilizando el periodo completo (1995 – 2010) en los meses de análisis (abril a septiembre). Teniendo como objetivo analizar de forma espacial el comportamiento de los grupos, buscando determinar estaciones atípicas y representativas.

La segunda etapa corresponde a la calibración de los modelos, la cual se separa en cuatro pasos: 1) la obtención de los datos de variables atmosféricas desde NCEP; 2) seleccionar las variables atmosféricas y puntos de grilla para los modelos de regresión lineal multivariada por medio de un análisis de coeficientes de correlación de Pearson (r); 3) generación de modelos lineales multivariados con su respectivo r y error absoluto medio (MAE) y 4) verificación de los supuestos para los modelos generados.

La tercera etapa, tiene como objetivo validar los modelos generados. En esta etapa, se utilizan los modelos lineales multivariados obtenidos en el periodo calibración (1995-2010) y son evaluados con los datos del periodo de validación (2011-2017), determinando el r y MAE para cada periodo

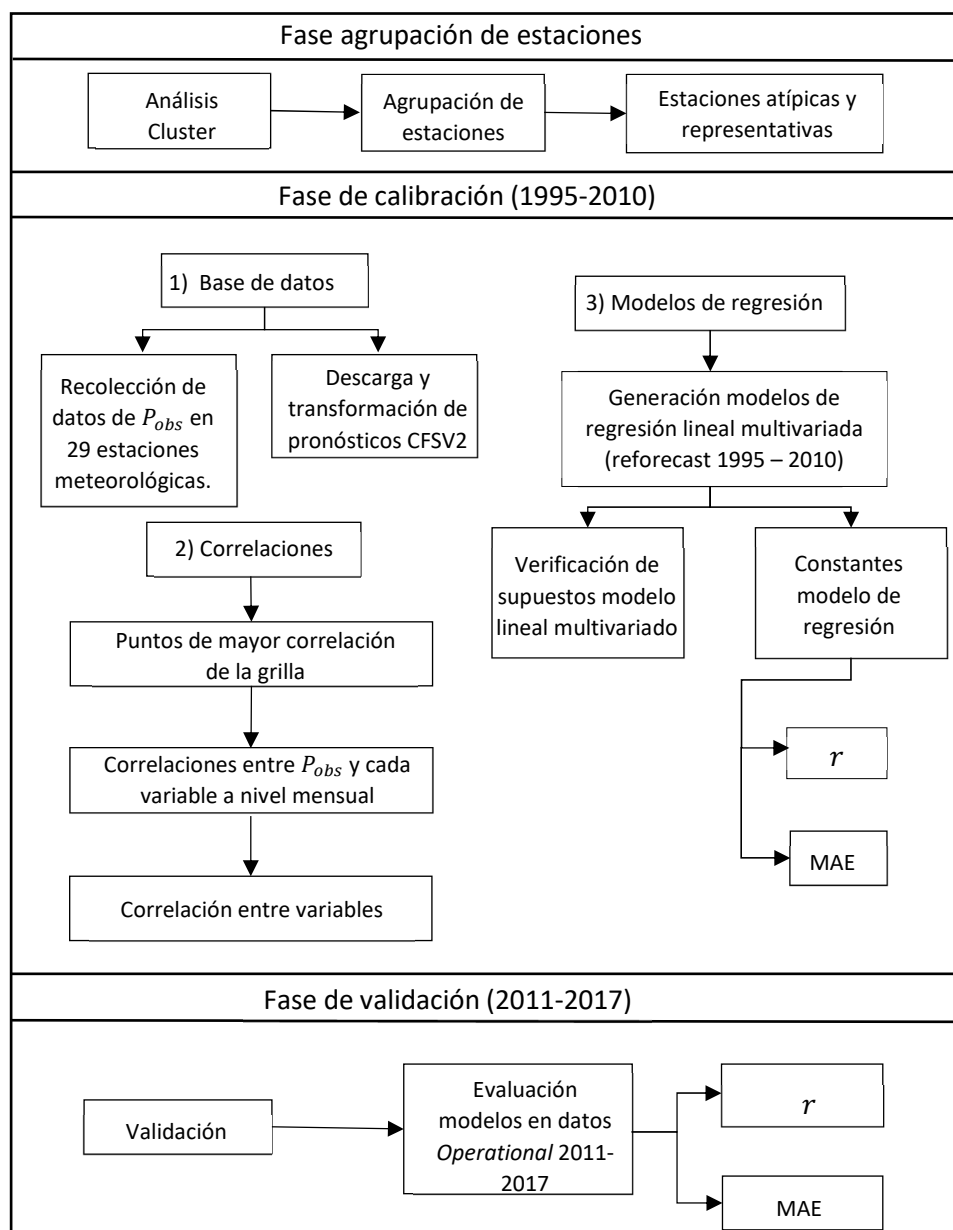


Figura 11: Esquema de la metodología de trabajo.
Fuente: elaboración propia.

5.1. ANÁLISIS CLUSTER

Se utiliza el análisis cluster para determinar grupos de estaciones meteorológicas, de este modo se podrá analizar el comportamiento espacial entre las estaciones. Determinando así, estaciones

atípicas, y estaciones representativas de cada cluster. Una estación representativa tendrá como objetivo mostrar información más detallada en análisis posteriores, puesto que mostrar los resultados de todas las estaciones para algunos análisis tendría una gran extensión.

Para este análisis, se recopilaron los datos de precipitación observada (P_{obs}) de las 29 estaciones meteorológicas (ver Tabla 8) para el periodo completo (1995-2017). Los datos son procesados por medio del *software* Matlab mediante la rutina *cluster.m* de elaboración propia (Anexo B). Los resultados del código *cluster.m* para el método Ward se mostrarán en un mapa de forma espacial utilizando el *software* QGIS. A continuación, se muestra el resumen de pasos del código.

Con los resultados del método Ward se realizará un mapa por medio del *software* QGIS con la ubicación espacial de las estaciones, diferenciando los grupos. Esto se realiza con la finalidad de analizar los grupos (clusters) formados.

5.2. PROCESO DE DATOS CFSV2

En el análisis sobre los estudios de desagregación estadística presentados en la sección 3.4, se plantean variables atmosféricas (VA) que podrían ser potenciales predictores de la precipitación observada. Bajo esta consideración, y junto con la disponibilidad de variables por CFSv2 (Tabla 9), para este estudio, se utilizará la variable de precipitación PRATE para los modelos de desagregación de la precipitación y otras 6 variables climáticas presentadas en la Tabla 10 para una corrección de los modelos.

Tabla 10: Variables atmosféricas de CFSv2 a utilizar. Fuente: (NCEP, 2011)

Variable	Nomenclatura	Unidad		Elevación aproximada [km]
Precipitación	PRATE	kg/(m ² s)		-
Humedad específica a 700 hPa	Q700	Kg/Kg		3.0
Altura geopotencial a 500 hPa	Z500	Metros Geopotenciales	mgp	5.5
Altura geopotencial a 850 hPa	Z850			1.5
Temperatura a 850 hPa	T850	Grados Kelvin	K	1.5
Temperatura superficial	TMPSPFC			Superficial
Presión reducida a nivel medio del mar	PRMSL	Pascales	Pa	Superficial

Para la descarga y transformación de los datos de las variables de CFSv2, se utiliza el *software* Matlab en ambos periodos (reforecast y operational). Los datos son descargados desde NCEP (<https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/climate-forecast-system-version2-cfsv2>), sin embargo, el formato original de los archivos no es leído por el *software* Matlab. Por lo que luego de su descarga, se debe proceder a la transformación del formato. Los datos descargados de cada variable se generan en una grilla con datos recopilados a nivel mundial, considerando, además, que se debe contar con datos diarios para la serie temporal completa (1995-2017).

Esto lleva a reunir una gran cantidad de archivos con muchos datos (alrededor de 600 Gb por variable), por lo que este proceso demora un tiempo considerable. Junto a la transformación de los

archivos, se realiza un filtro de los datos, dejando solo datos que se encuentran en la grilla dentro de la zona de estudio (ver Figura 10).

El procedimiento completo fue realizado mediante el *software* Matlab para procesar los archivos. En este punto, el resultado final son archivos de salida *.mat de pronósticos ordenados por mes para cada una de las variables. Los códigos originales que se presentan fueron desarrollados por el departamento de ingeniería hidráulica y ambiental de la Pontificia Universidad Católica de Chile, a los cuales se les realizó algunas modificaciones según las características particulares de esta investigación (Anexo 0).

5.3. ANÁLISIS DE CORRELACIÓN

Para realizar el análisis de correlación se debe contar con series de datos sobre la precipitación observada y variables atmosféricas con las que se desea predecir la precipitación. En las secciones anteriores, se presentó la metodología para obtener los datos a utilizar en la construcción de los modelos de regresión lineal multivariada.

En esta sección, se emplea el coeficiente de correlación de Pearson (r) para obtener correlaciones de tres formas distintas: espacial, mensual y entre variables. Para esto, se utilizan las series completas de datos (1995-2017), considerando los días pertenecientes a los meses de análisis (abril a septiembre) y con precipitación observada diaria que registren un valor mayor a cero ($P_{obs} > 0$).

- i. Correlación espacial: Tiene como objetivo calcular y ubicar los r para determinar el punto de grilla de mejor correlación. Los r son calculados entre la P_{obs} en cada estación con cada serie de la VA (una serie por cada punto de grilla). Los r se ubican en su respectivo punto de grilla para construir un gráfico espacial.
- ii. Correlación mensual: Se determinan los valores del r a nivel mensual, utilizando la serie del mejor punto de grilla de cada VA. Se presentan los resultados de forma gráfica por estación, variable y mes.
- iii. Correlación entre variables: Se determina el coeficiente de correlación entre variables, utilizando la serie de los puntos de mejor correlación para el periodo completo con todos los meses de análisis juntos en la serie.

5.3.1. CORRELACIÓN ESPACIAL

En esta sección, se realiza el primer acercamiento al comportamiento del coeficiente de correlación de Pearson (r) entre las variables atmosféricas (VA) (ver Tabla 10) y la precipitación observada (P_{obs}). El objetivo es determinar el punto de la grilla de mejor correlación para los análisis posteriores.

Cada VA presenta una grilla con series temporales de la variable, es decir, en cada punto de grilla se encuentra una serie de datos de la variable, un dato para cada día dentro del periodo de análisis (periodo completo 1995-2017). De este modo, existirá un valor de correlación para cada punto de

grilla de acuerdo con una estación y una variable. Se consideran solo los días en los que existe precipitación observada mayor a cero, y que se encuentren dentro de los meses de análisis (abril a septiembre).

El manejo de datos y el cálculo del r se desarrolla en el *software* Matlab, por medio de la rutina *Correlacion_x_e.m* de elaboración propia (Anexo 03.1), con estos datos se elaboran los gráficos espaciales en QGIS. Con los resultados se analiza el comportamiento espacial de la correlación entre la P_{obs} y cada VA para las estaciones representativas escogidas en el análisis cluster, y se busca el punto de mejor correlación en la grilla para cada estación y cada VA, pudiendo ser correlaciones positivas o negativas según el caso.

5.3.2. CORRELACIÓN MENSUAL

De aquí en adelante la serie temporal de cada variable atmosférica corresponderá a la serie del punto de mayor correlación espacial, esto quiere decir que, para una determinada estación las variables atmosféricas pueden estar ubicadas en distintos puntos de la grilla, dicho punto fue determinado en la sección anterior 5.3.1. Se continúa utilizando los datos del periodo completo (1995-2017) para los meses de abril a septiembre y solo en los días con precipitación observada mayor a cero. Pero ahora los datos serán separados a nivel mensual, es decir, se trabajará con una serie de datos para cada mes de análisis, estación y VA. Estas series temporales mensuales son utilizadas como *input* de los modelos de regresión lineal multivariada que se describen en la sección 5.4.

El manejo de datos y cálculos del r es desarrollado por medio del *software* MATLAB, utilizando la rutina *Correlacion_x_e.m* (Anexo 0). Los resultados son exportados a un archivo Excel y se construyen gráficas para cada VA, donde cada gráfica contempla las 29 estaciones meteorológicas en el eje X, y los resultados de correlación para cada mes en el eje Y.

El objetivo de esta sección es analizar los resultados gráficos mensuales de cada variable a través de las 29 estaciones, identificando si tienen una correlación positiva o negativa, observando singularidades por variable, por mes y por estación.

5.3.3. CORRELACIÓN ENTRE VARIABLES

Pueden existir dos o más variables que entreguen la misma información atmosférica, incluso una VA puede ser una combinación lineal de otras VAs. En esta sección se pretende identificar el primer caso, si dos variables entregan la misma información tendrán una alta (0.7 – 0.9) o muy alta (0.9 – 1.0) correlación entre ellas. Primero se separan en dos grupos de VAs, las que sean directamente proporcionales con la precipitación y las que sean inversamente proporcionales, es decir, las que entregarán r positivos o negativos respectivamente.

Posteriormente, se procede a realizar un análisis entre las series temporales de las VAs del mismo signo de correlación, estos signos se han obtenido en la sección anterior 5.3.2. Nuevamente, se utilizan los datos del periodo completo (1995-2017) para los días que se encuentran dentro de los meses de análisis (abril a septiembre) y con P_{obs} mayor a cero. Finalmente, se presenta en forma de grafica *boxplot* los resultados de correlación entre las variables de igual signo. Las acciones que realiza la rutina *Correlacion_x_e.m* para esta sección se describen en el Anexo 0.

5.4. CALIBRACIÓN DE MODELOS

Con los datos procesados, se obtienen series temporales por cada estación meteorológica y mes en estudio. Una de las series corresponde a la precipitación observada (P_{obs}), las demás series corresponden a datos de las variables atmosféricas (VAs) que fueron recogidos desde el punto de grilla de mejor correlación con la P_{obs} (ver sección 5.3.1). Los datos se agrupan de forma mensual de acuerdo con lo expuesto en la sección 5.3.2. Luego, se procede a someter las series temporales a una regresión lineal multivariada por cada estación y mes en el periodo *reforecast* (1995-2010), utilizando la P_{obs} como predictando y las VAs como predictores. Obteniendo así, mediante una combinación lineal, modelos que pronostican la precipitación diaria, para cada mes y estación.

REGRESIÓN LINEAL MULTIVARIADA

Dado que se cuenta con más de una variable predictora, nos encontramos en el caso de una regresión lineal multivariada. Las variables independientes seleccionadas son PRATE, Q700, Z850, PRMSL, Z500, T850 y TMP5FC (éstas pueden variar según los análisis de correlaciones). Por otro lado, la precipitación observada en cada una de las estaciones corresponde a la variable dependiente, es decir, es la variable por predecir (predictando).

De esta manera, el modelo arroja una precipitación pronosticada (P_{pron}) y se puede expresar por la siguiente ecuación:

$$P_{pron} = \beta_0 + \beta_1 PRATE + \beta_2 X_2 + \dots + \beta_n X_n$$

Donde las constantes $\beta_0, \dots, \beta_n$ corresponden a las constantes a calibrar y $X_2, \dots, X_n$ corresponden a las variables atmosféricas utilizadas. La variable PRATE es la única variable que se encuentra presente en todos los modelos. Si se suma la variable aleatoria de error se puede obtendría la precipitación observada.

$$P_{obs} = P_{pron} + e$$

Para obtener la calibración de cada de modelo por mes y estación se ingresan los datos a R-Studio, software que proporciona una interfaz gráfica para ejecutar R (RStudio_Team, 2020). Todas las variables deben superar el criterio de tolerancia (sección 3.3.2) para que puedan ser introducidas en la ecuación, independientemente del método de entrada especificado. Tampoco se introduce una variable si esto provoca que la tolerancia de otra ya presente en el modelo se sitúe por debajo del criterio de tolerancia. Para este análisis se utilizó el código *Modelos_regresion_mensual.R* de elaboración propia (Anexo 0.).

5.5. VERIFICACIÓN DE SUPUESTOS

Para verificar la validez de las regresiones lineales multivariadas obtenidas en la sección anterior, se deben verificar que éstas cumplan con los supuestos de linealidad indicados en la sección 3.3.3. A continuación, se muestran los valores de los índices y técnicas utilizados para verificar la evaluación de las condiciones de linealidad, basada en la información de Draper & Smith (1998). Para este análisis se utilizó el código *Test_supuestos.R* de elaboración propia (Anexo 0.).

- **Linealidad:** Se generan diagramas de regresión parcial entre la precipitación observada diaria y cada VA, para cada estación y mes de estudio. Buscando de manera visual una relación con tendencia lineal de los datos para verificar el cumplimiento del supuesto.
- **Independencia:** Se aplica el estadístico Durbin-Watson DW. El supuesto de independencia se cumplirá cuando el estadístico se encuentre dentro del siguiente rango: $1,5 < DW < 2,5$.
- **Homocedasticidad:** Se generan diagramas de dispersión de pronósticos y residuos. Se verifica que la varianza de los residuos se mantenga constante para que se cumpla el supuesto. Además, si complementa con el cálculo del estadístico Breusch-Pagan, para cumplir el supuesto deben tener una significancia menor a 0,05.
- **Normalidad:** Para el cumplimiento de este supuesto se trabaja con tres requisitos:
 - Gráficos P-P de los residuos, se busca que los residuos cumplan con un comportamiento de $N(0, \sigma^2)$ (requisito de una distribución normal).
 - Gráficos Q-Q de las variables que componen la regresión lineal. En este punto, se busca que las variables cumplan con un comportamiento de $N(0, \sigma^2)$ (requisito de una distribución normal).
 - Test de Kolmogórov-Smirnov o Shapiro Wilk, para cumplir el supuesto, deben tener una significancia menor a 0,05.
- **No-colinealidad:** Se aplica el factor de incremento de varianza (FIV). Se cumple el supuesto de no-colinealidad cuando el valor es menor a 10 ($FIV < 10$).

5.6. VALIDACIÓN DE MODELOS

El análisis de los modelos de regresión lineal multivariada consiste en analizar la exactitud de los modelos. Para ello se analizan el coeficiente de correlación de Pearson (r) y el error absoluto medio (MAE) obtenidos en cada uno de los 174 modelos, uno por cada estación y mes de análisis, para el periodo 2011-2017.

Para determinar el r y el MAE en el periodo de validación (operational) se calcula primero la precipitación pronosticada (P_{pron}) en el periodo de validación, lo cual consiste en evaluar los modelos de regresión obtenidos en el periodo de calibración, pero ahora evaluados con los datos del periodo de validación, esto para los 174 modelos.

El análisis se realiza por medio de resultados numéricos, comparando rangos y valores promedio, como también un análisis de los *boxplot* de los r y los MAEs de ambos periodos. También, se comparan los resultados con un estudio similar realizado en la misma zona de estudio por Vega (2018). La evaluación de los modelos y la construcción de las gráficas fue realizada con el *software* Matlab por medio del código *Validacion_correlacion.m* de elaboración propia (Anexo 0.).

6. RESULTADOS

Los resultados de este estudio se dividen en tres fases de acuerdo con la metodología presentada en la sección 5 (ver Figura 11). La primera fase, consiste en un análisis cluster de la precipitación observada (P_{obs}) en las estaciones meteorológicas, formando grupos (clusters) de estaciones entre las 29 estaciones en estudio (ver Tabla 8). La segunda fase, corresponde a la calibración de los modelos de regresión lineal multivariada, para ello se realizan tres procesos: primero el análisis de correlaciones, lo segundo es calibrar los modelos para el periodo *reforecast*, y por último se verifican los supuestos de los modelos. La tercera y última fase, es la validación de los modelos, que consiste en un análisis de la dependencia lineal entre la precipitación pronosticada (P_{pron}) y la precipitación observada (P_{obs}) en las estaciones meteorológicas, como también determinar la precisión de los modelos mediante el coeficiente de correlación de Pearson (r) y el error absoluto medio (MAE).

6.1. CLUSTERS DE ESTACIONES METEOROLÓGICAS

Se realiza un análisis cluster con los datos de precipitación observada (P_{obs}) en las 29 estaciones meteorológicas en estudio (ver Tabla 8). Se consideran solo los meses de análisis, abril a septiembre, del periodo completo (1995-2017). Con esto se podrá agrupar las estaciones en clusters, determinando elementos atípicos y representativos en cada uno de estos grupos.

6.1.1. MÉTODO *SINGLE*

Primero se utiliza el método *Single*, el cual permite identificar elementos (estaciones meteorológicas) atípicos con respecto a los demás elementos. Las estaciones meteorológicas atípicas, debido a esta condición, no podrían ser utilizadas como un elemento representativo para el cluster al que pertenecen. En la Figura 12, se observa el dendograma del método *Single*, donde se visualiza que las últimas estaciones en unirse a un cluster son: 28, 1, 14, 23, 11 y 29. En particular, la estación 29 es la última estación en incorporarse a un cluster, puesto que su unión genera un aumento significativo en la distancia euclidiana de similitud (distancia entre los clusters), siendo así una estación fuertemente atípica.

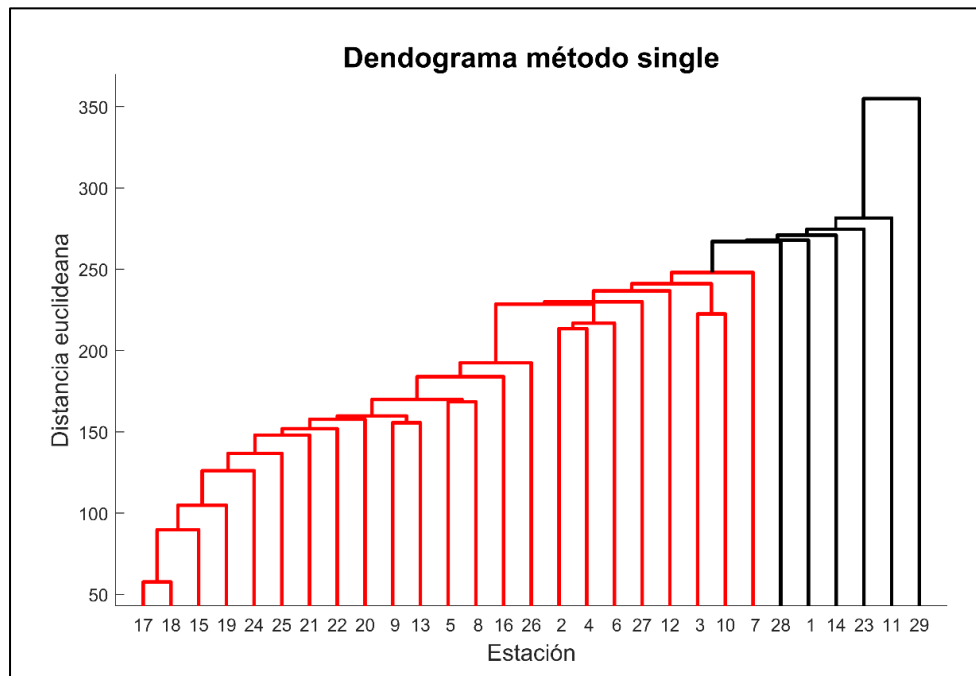


Figura 12: Dendrograma método Single. Fuente: Elaboración propia

6.1.2. MÉTODO WARD

Por medio del método Ward, se busca un número de clusters apropiados para la situación. En la Figura 13, se encuentra el dendrograma del método con todos los posibles grupos que se podrían formar dada la distancia euclidiana, de acuerdo al número de grupos formados (eje Y). Para efecto de claridad, se han marcado con distintos colores (azul, rojo y verde) la formación de tres clusters de forma arbitraria. La utilización de estos grupos nos lleva a un análisis espacial con delimitaciones claras de ubicación de los clusters (ver Figura 14). Seleccionar un número menor de clusters, dos en este caso, llevaría a una formación de dos grupos más heterogéneos asociado a un aumento significativo de la distancia euclidiana.

Estos tres clusters de estaciones, se observan en la Figura 14 de forma espacial, y se le ha dado un nombre característico a cada uno de acuerdo con la ubicación de la mayoría de sus estaciones. De esta forma, el cluster rojo es el cluster costa, el cluster azul es el cluster valle (valle central) y el cluster verde es el cluster cordillera.

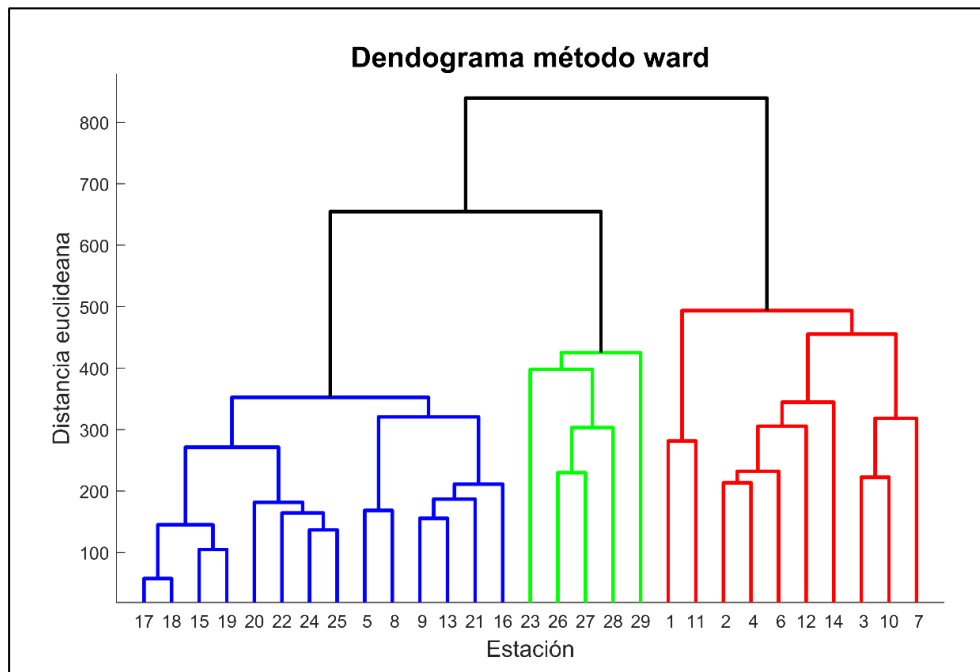


Figura 13: Dendrograma método Ward
Fuente: Elaboración propia

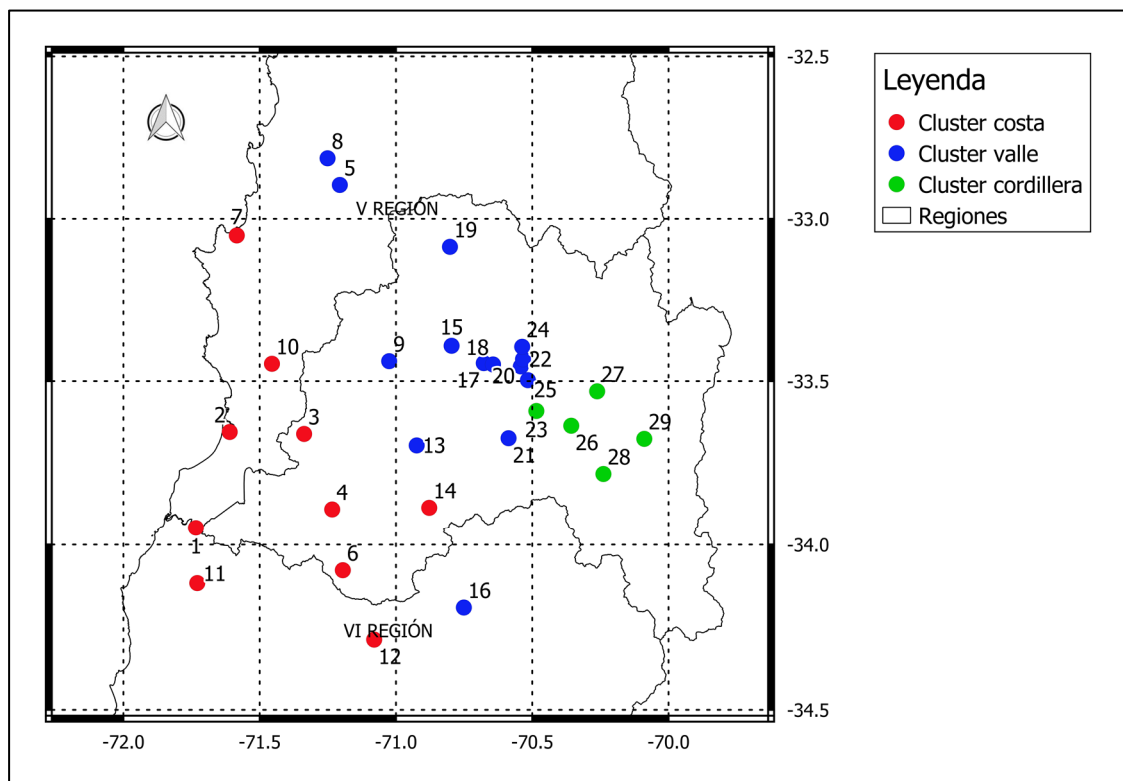


Figura 14: Representación espacial de las estaciones meteorológicas agrupadas en 3 cluster. Fuente: Elaboración propia

6.2. CORRELACIONES

Las correlaciones son determinadas mediante el coeficiente de correlación de Pearson (r) entre series temporales de datos de tres formas distintas: espacial, mensual y entre variables. Se han utilizado las series completas de datos (1995-2017), considerando los días pertenecientes a los meses de análisis (abril a septiembre) y con precipitación observada diaria con un valor mayor a cero ($P_{obs} > 0$).

6.2.1. CORRELACIÓN ESPACIAL

Los datos de las variables atmosféricas (VA) recogidos desde NCEP se extraen en una grilla con datos recopilados a nivel mundial (1° de latitud por 1° de longitud), la cual ha sido recortada en el cuadrante de interés, donde cada punto de esta grilla posee una serie temporal de la variable medida en ese punto.

En el estudio de Vega (2018) se utiliza el método *el vecino más cercano* (Sibson, 1981), para seleccionar el punto de grilla de la variable atmosférica más cercana a la estación meteorológica. También, existen otros procedimientos similares que utilizan puntos cercanos a la estación (IDW y bilineal). Particularmente, en este estudio se utilizará un método diferente, se observará el comportamiento de la correlación entre una variable atmosférica (VA) y la precipitación observada (P_{obs}) en cada punto de la grilla para un cuadrante amplio con 42 o 63 puntos, dependiendo de la variable. De tal forma de identificar el punto de mejor correlación utilizando el periodo completo de datos (1995-2017), lo que no corresponde necesariamente al punto más cercano de la estación (ver la Figura 15 y Figura 16).

Posteriormente, se ubican las correlaciones en los puntos de grilla de la VA. Este análisis se realizó para las 29 estaciones en estudio, pero solo se mostrarán los resultados acotados a tres estaciones representativas de los tres grupos de estaciones (estaciones 4, 17 y 26) de acuerdo, a lo determinado en el análisis cluster (sección 6.1).

En los análisis posteriores, se utilizará la serie temporal de la variable perteneciente al punto de mayor correlación, dependiendo únicamente de la estación en análisis. De este modo, existe una serie temporal de cada variable de acuerdo con cada estación. Además, este punto de grilla es el mismo incluso en la división de datos en los dos periodos: *Reforecast* (calibración) y *Operational* (validación), como a su vez, es el mismo para cada mes y año.

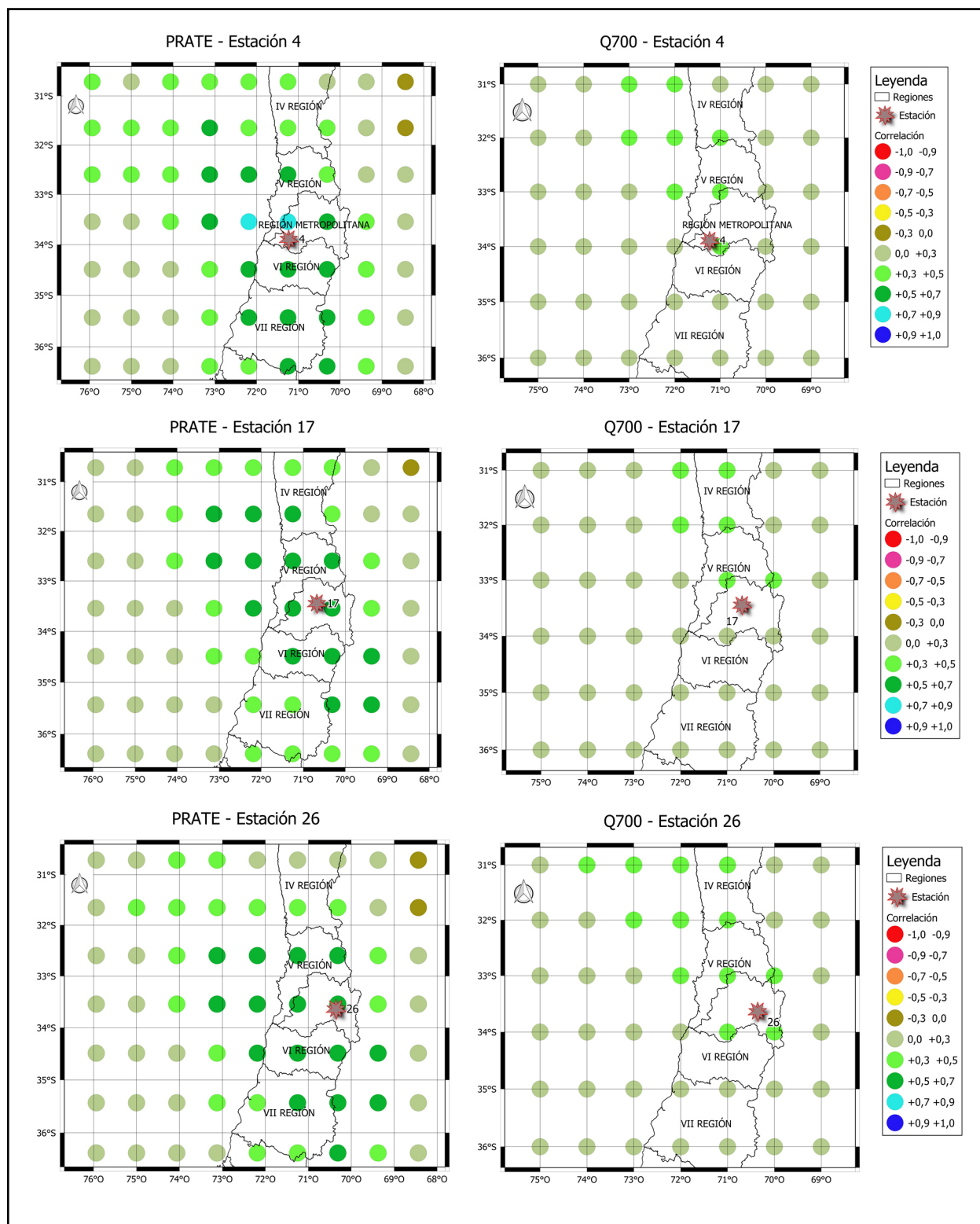


Figura 15: Correlación espacial entre la precipitación observada en las estaciones 4, 17 y 26; y las variables PRATE y Q700. Fuente: Elaboración propia.

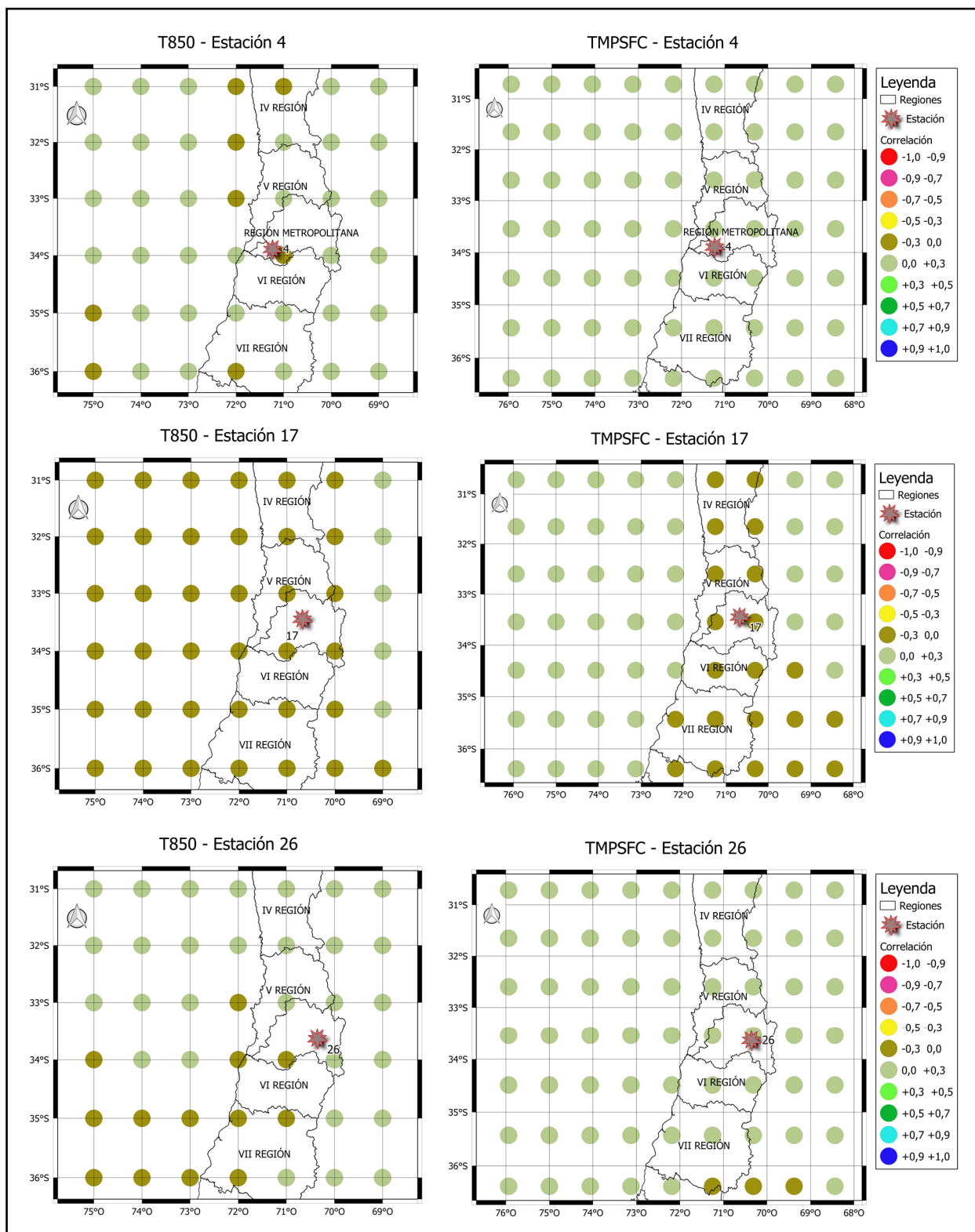


Figura 16: Correlación espacial entre precipitación observada en las estaciones 4, 17 y 26; y las variables de temperatura T850 y TMPSFC. Fuente: Elaboración propia.

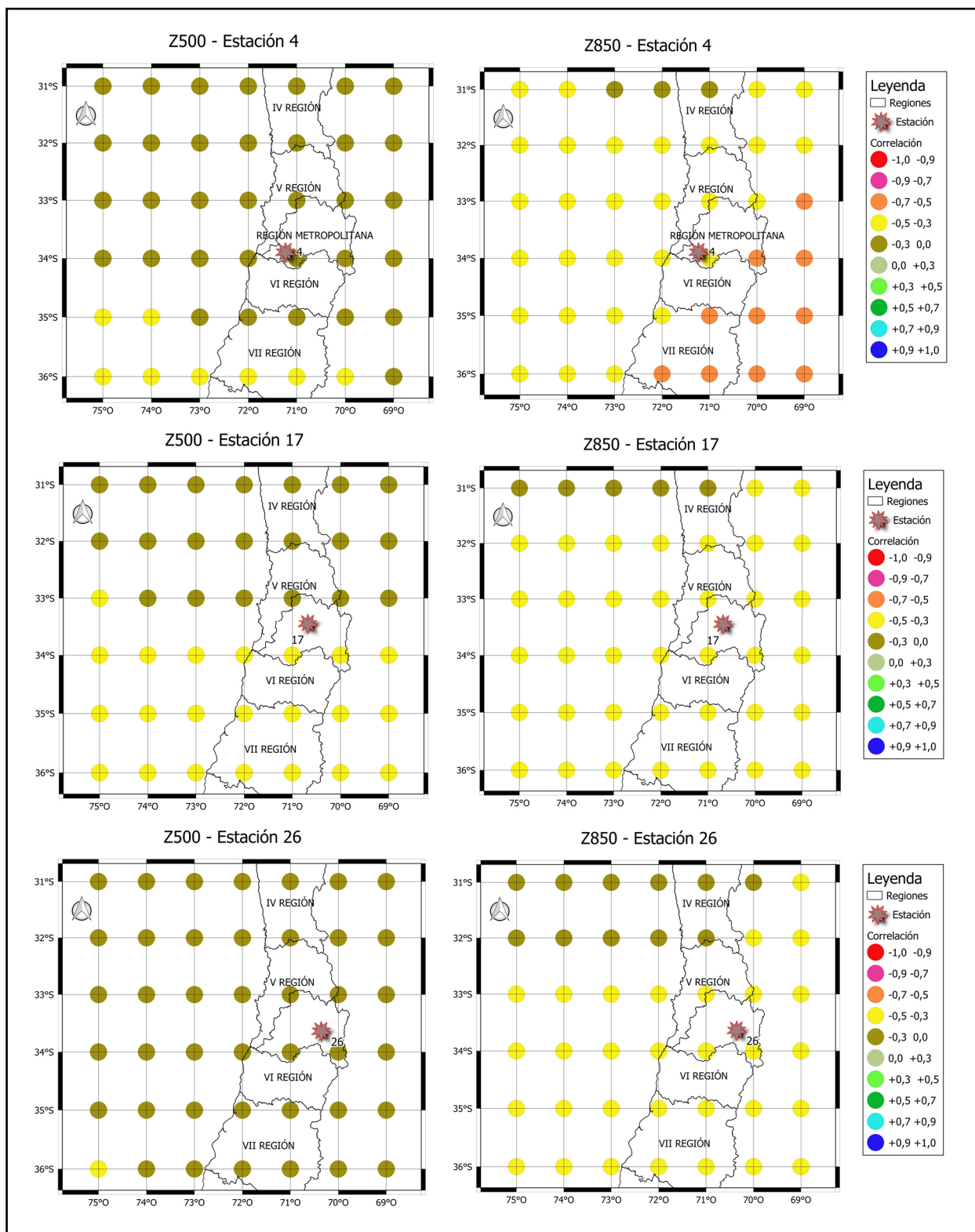


Figura 17: Correlación espacial entre precipitación observada en las estaciones 4, 17 y 26; y las variables de altura geopotencial Z500 y Z850. Fuente: Elaboración propia

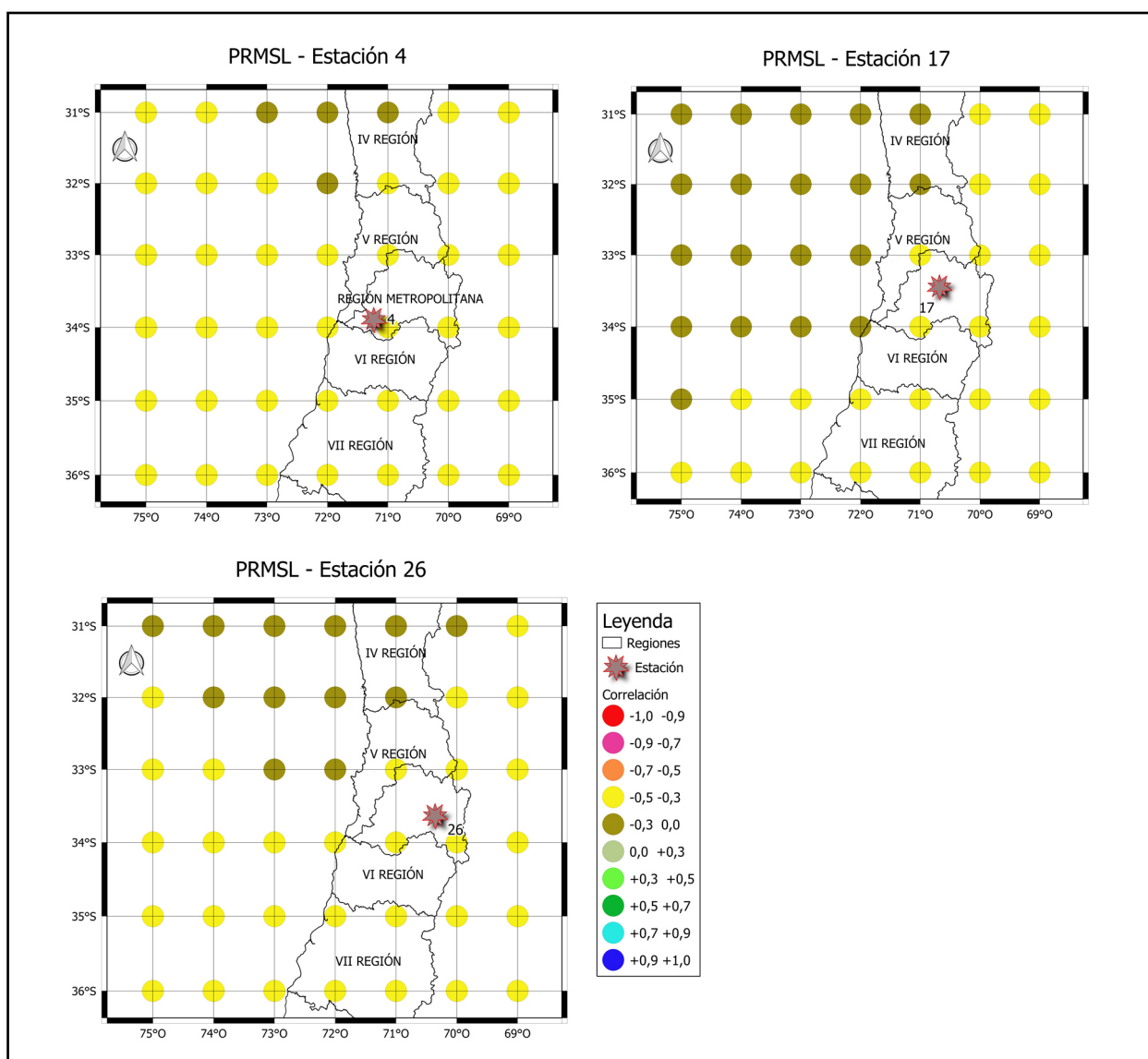


Figura 18: Correlación espacial entre precipitación observada en las estaciones 4, 17 y 26; y las variables de presión PRMSL. Fuente: Elaboración propia.

6.2.2. CORRELACIÓN MENSUAL

Al realizar el análisis de correlación espacial (6.2.1), se selecciona para cada variable atmosférica (VA), una serie temporal desde el punto de mayor correlación, con una correlación positiva o negativa según la variable en análisis. En esta sección, se determinará la correlación a nivel mensual para los 6 meses en estudio, de abril a septiembre, entre la precipitación observada en cada una de las 29 estaciones (Tabla 8) y cada una de las 7 variables atmosféricas en estudio. Sin embargo, no se acotará a las 3 estaciones representativas del análisis cluster, puesto que es posible mostrar de forma gráfica los resultados de las 29 estaciones ayudando, además, al análisis general. A

continuación, se muestran las gráficas de correlación entre cada variable atmosférica y la precipitación observada para cada estación y mes (Figura 19, Figura 21 a Figura 26). Para profundizar en unos análisis, se presenta en la Figura 20 dos gráficas entre la precipitación observada diaria y la variable PRATE, particularmente en la estación 13 para los meses de abril y agosto.

➤ VARIABLE DE PRECIPITACIÓN

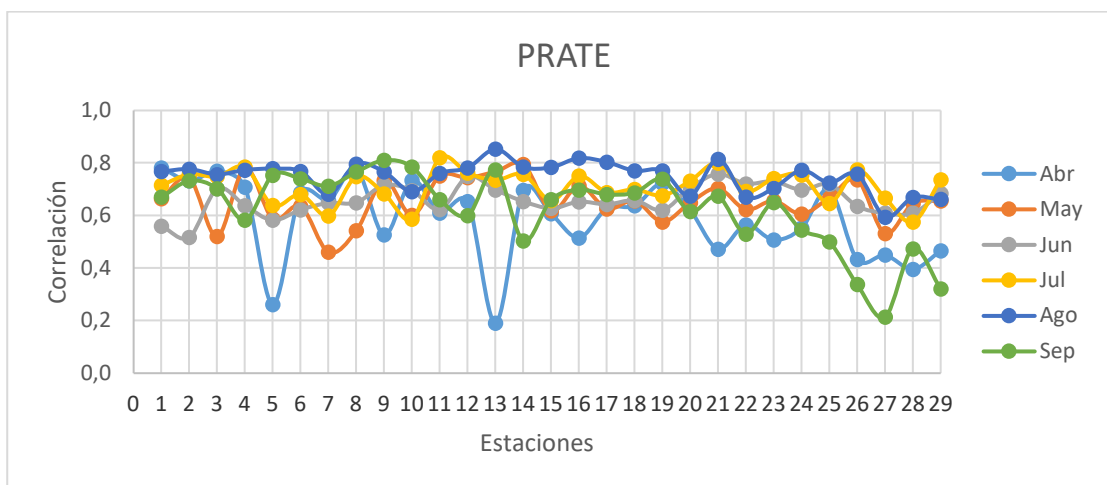


Figura 19: Correlación mensual entre la precipitación observada y la variable PRATE para las 29 estaciones. Fuente: Elaboración propia.

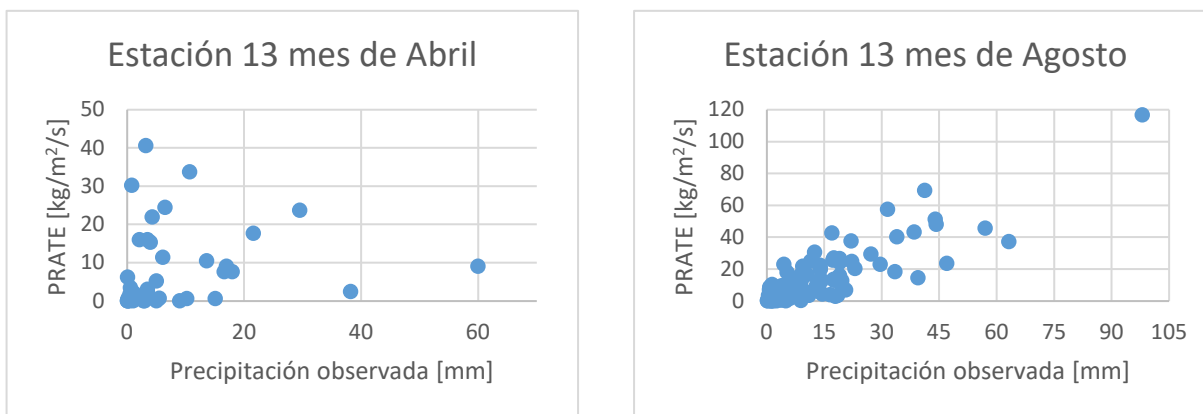


Figura 20: Gráfica de los valores de PRATE y Pobs para los meses de abril y agosto en la estación 13. Fuente: Elaboración propia

➤ VARIABLE DE HUMEDAD

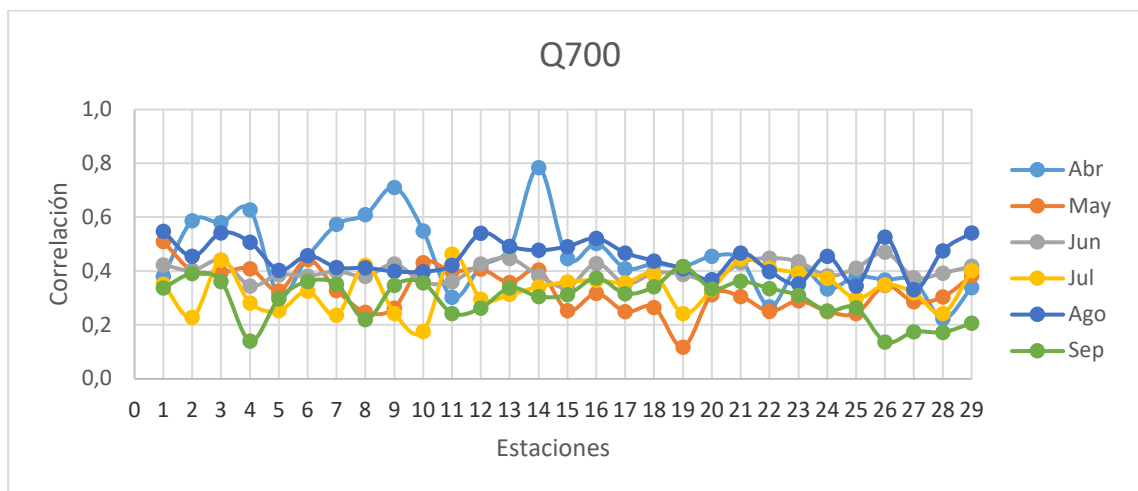


Figura 21: Correlación mensual entre la precipitación observada y la variable Q700 para las 29 estaciones. Fuente: Elaboración propia.

➤ VARIABLES DE PRESIÓN:

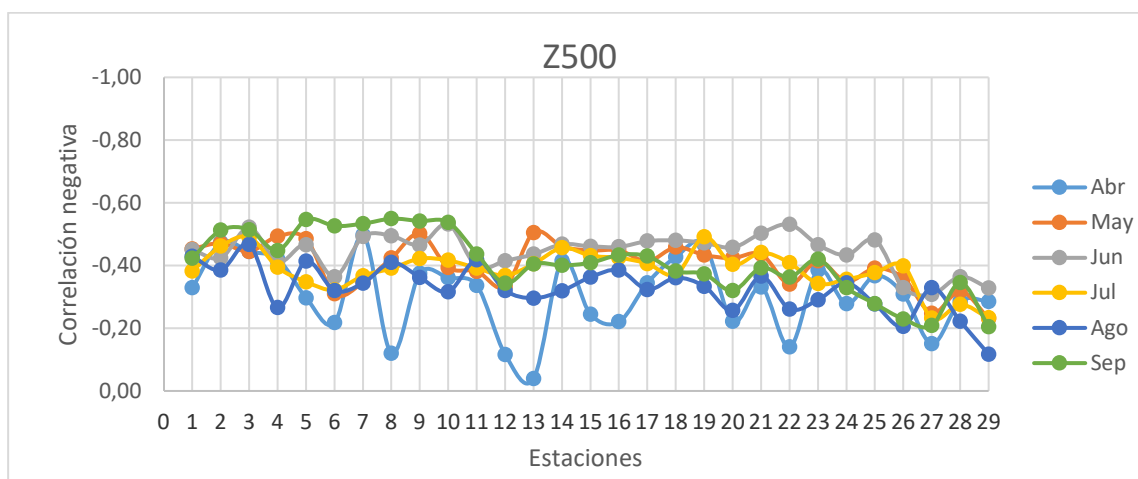


Figura 22: Correlación mensual entre la precipitación observada y la variable Z500 para las 29 estaciones. Fuente: Elaboración propia.

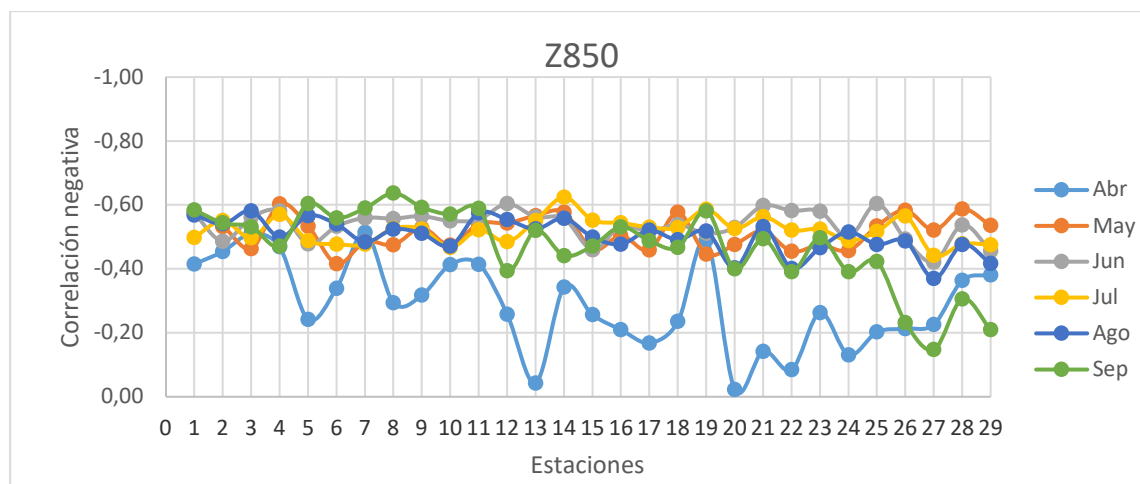


Figura 23: Correlación mensual entre la precipitación observada y la variable Z850 para las 29 estaciones. Fuente: Elaboración propia.

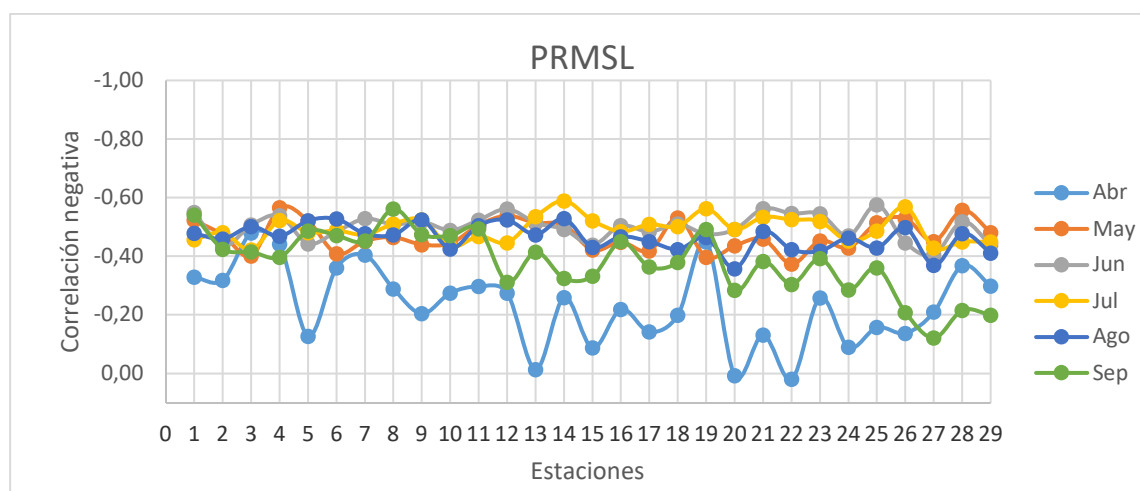


Figura 24: Correlación mensual entre la precipitación observada y la variable PRMSL para las 29 estaciones. Fuente: Elaboración propia.

➤ VARIABLES DE TEMPERATURA

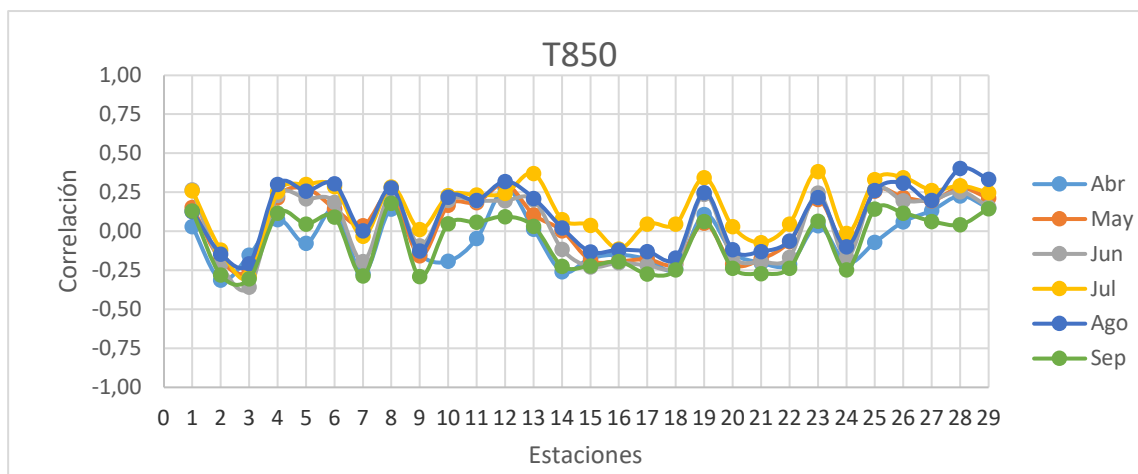


Figura 25: Correlación mensual entre la precipitación observada y la variable T850 para las 29 estaciones. Fuente: Elaboración propia.

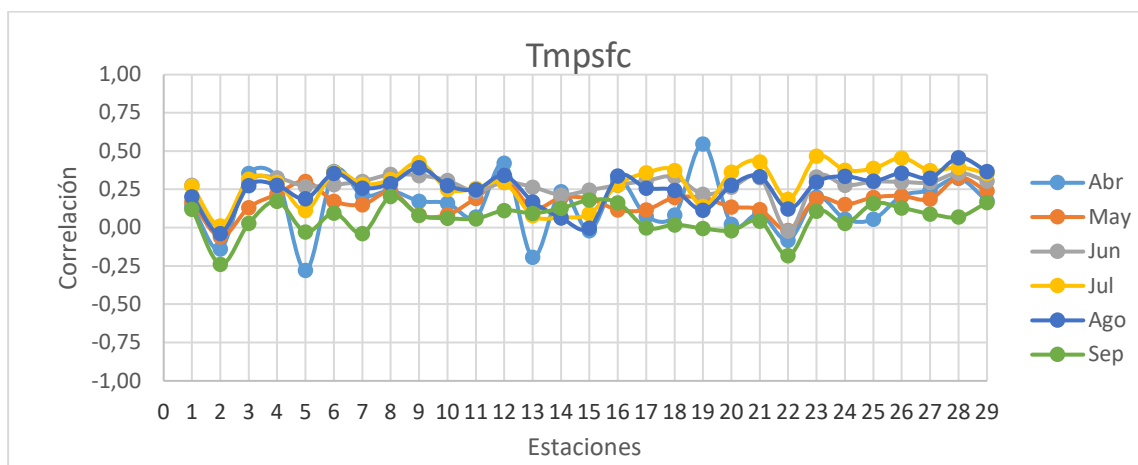


Figura 26: Correlación mensual entre la precipitación observada y la variable TMPsfc para las 29 estaciones. Fuente: Elaboración propia.

6.2.3. CORRELACIÓN ENTRE VARIABLES

En esta sección se ha determinado la correlación entre variables, para verificar que cada una entregue información independiente como predictor de la precipitación observada. De entregar la misma información, se descarta una de las variables asociadas al modelo. En la sección 6.2.2, se ha observado que las variables atmosféricas de precipitación y humedad (PRATE y Q700 respectivamente) presentan una correlación positiva con la precipitación observada (P_{obs}), por otro lado, que las variables de presión Z500, Z850 y PRMSL presentan correlaciones negativas con P_{obs} .

En la Figura 27, se presenta un *boxplot* con la correlación entre variables, analizando entre éstas las del mismo signo de correlación. Se utilizan las series de datos del periodo completo (1995-2017) para las 29 estaciones en estudio, donde la serie temporal corresponde al punto de grilla de mejor correlación. Se cuenta entonces con 29 datos de correlación en el cruce entre cada par de variables.

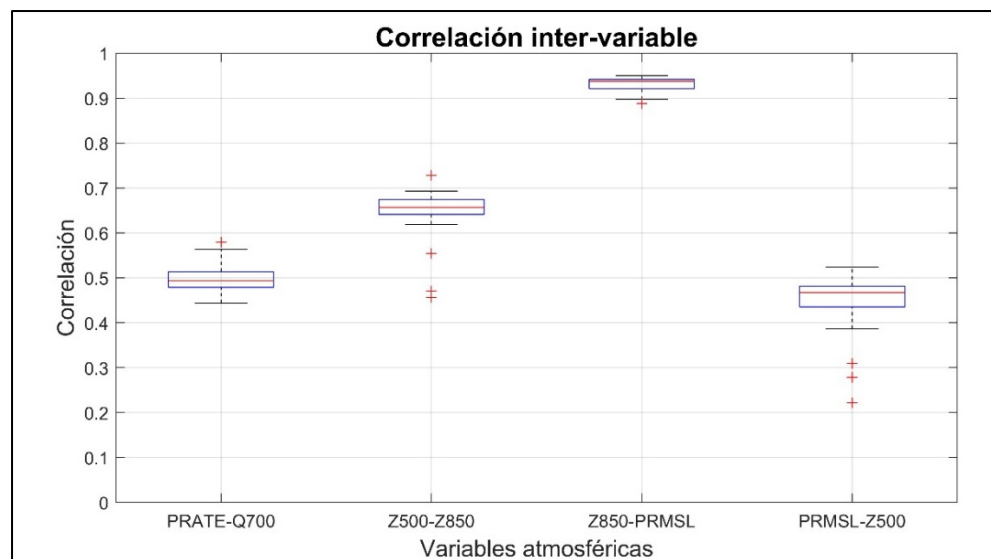


Figura 27: Boxplot correlaciones entre variables. Fuente: Elaboración propia.

Dados los resultados, las variables Z850 y PRMSL aportan información similar (linealmente dependiente). Por ende, se debe descartar una variable, por lo que no se continuará con la variable PRMSL como variable regresora de los modelos.

Para complementar, también se observa de manera grafica el cruce de las series temporales entre las variables atmosféricas en la Figura 41, Figura 42 y Figura 43 (Anexo 1.E), notando que entre las variables PRMSL y Z850 se forma una relación linealmente dependiente entre ellas.

6.3. CALIBRACIÓN DE LOS MODELOS

Luego de la selección de las variables atmosféricas (VAs) a utilizar, se procede a generar modelos de regresión multivariada para la proyección de la precipitación diaria observada. Para este proceso, se utiliza el *software* R y se establece que utilice siempre la variable PRATE como punto de partida en la generación del modelo de regresión. Luego, por medio del método hacia delante (*forward*), se busca la adición de nuevas variables al modelo. En las Tabla 11, Tabla 12 y Tabla 13 se observan los modelos de regresión obtenidos para cada estación, en los distintos meses de estudio (abril a septiembre), con sus respectivos valores de r y error absoluto medio (MAE). En las casillas donde no hay valor, se debe a que el *software* no agregó la variable atmosférica puesto que no contribuía a mejorar el ajuste. Se generó un modelo para cada una de las 29 estaciones y para cada uno de los 6 meses de estudio, 174 en total. En promedio estos modelos poseen un coeficiente de correlación de Pearson de 0,68, correspondiente una correlación media (Ver Tabla 5).

En la Tabla 14, se muestra de forma resumida las variables que son incluidas en los modelos de regresión, notar que hay modelos donde solo se incorporó una variable, mientras que en otros modelos se incorporaron dos variables además de PRATE.

Tabla 11 Constantes de cada modelo de regresión obtenido para la serie temporal fase Reforecast (Calibración), para los meses de abril y mayo, separados por estación meteorológica. Fuente: Elaboración propia.

Estación meteorológica	Abril							Mayo						
	β_0	P _{RATE}	Q700	Z850	Z500	r	MAE	β_0	P _{RATE}	Q700	Z850	Z500	r	MAE
1	167	1,42	-	-	-0,029	0,81	4,0	180	0,826	2436	-	-0,032	0,67	7,1
2	190	1,12	-	-	-0,033	0,77	3,6	212	1,486	-	-	-0,037	0,70	9,1
3	303	0,40	-1474	-	-0,053	0,74	3,7	477	0,103	-	-0,14	-0,046	0,49	9,0
4	213	0,86	-	-	-0,037	0,74	3,8	360	1,047	-	-	-0,062	0,66	10,3
5	169	1,57	-	-0,11	-	0,70	3,4	507	0,623	-	-	-0,089	0,58	10,0
6	5,4	0,75	-	-	-	0,71	5,1	6,4	0,946	-	-	-	0,64	12,1
7	177	0,66	-	-	-0,030	0,60	4,5	405	0,544	-	-0,26	-	0,45	16,8
8	223	1,16	-	-0,15	-	0,67	4,3	643	0,822	-	-	-0,113	0,53	13,7
9	174	-0,05	1756	-	-0,030	0,59	3,0	193	0,725	-2440	-	-0,034	0,76	6,9
10	272	1,03	-1799	-0,18	-	0,73	3,7	1,30	0,945	4791	-	-	0,53	11,5
11	429	0,87	-2630	-	-0,074	0,73	6,8	3,13	1,416	-	-	-	0,66	13,1
12	7,1	0,38	-	-	-	0,60	6,9	4,52	0,874	-	-	-	0,64	13,8
13	0,35	-0,02	3051	-	-	0,38	4,7	211	0,651	-	-	-0,037	0,73	7,9
14	0,24	0,15	4318	-	-	0,55	5,4	282	0,767	-	-	-0,049	0,71	10,0
15	204	0,42	-	-	-0,036	0,55	3,8	178	0,436	-1459	-	-0,031	0,67	5,4
16	0,30	0,47	2364	-	-	0,40	6,2	154	0,639	-	-	-0,027	0,74	6,3
17	231	0,73	-	-	-0,040	0,62	4,7	199	0,587	-1771	-	-0,034	0,67	6,7
18	255	0,77	-	-	-0,044	0,64	4,6	192	0,55	-1429	-0,12	-	0,72	6,2
19	2,9	0,70	-	-	-	0,71	3,1	315	0,353	-	-	-0,056	0,58	7,6
20	160	0,69	-	0,08	-0,049	0,53	5,0	200	0,669	-	-	-0,035	0,66	7,3
21	300	0,07	-	-	-0,052	0,40	5,7	189	0,739	-1735	-	-0,032	0,74	7,1
22	3,2	1,21	-	-	-	0,64	4,5	6,08	0,64	-	-	-	0,62	8,6
23	460	0,35	-	-	-0,080	0,54	6,7	4,48	1,106	-	-	-	0,65	12,6
24	160	0,83	-	-	-0,028	0,57	4,6	4,81	0,621	-	-	-	0,64	7,7
25	173	1,03	-	-	-0,030	0,74	4,1	162	0,502	-	-0,10	-	0,67	7,8
26	400	0,32	-	-	-0,069	0,55	6,0	210	0,929	-	-0,14	-	0,70	8,9
27	100	0,16	-	-0,06	-	0,32	5,8	246	0,462	-	-0,16	-	0,58	8,7
28	225	0,28	-	-0,14	-	0,50	7,2	261	0,692	-	-0,17	-	0,70	8,6
29	186	0,27	-	-0,12	-	0,49	5,9	-94	0,819	-	-0,14	0,053	0,62	9,9

Tabla 12 Constantes de cada modelo de regresión obtenido para la serie temporal reforecast (Calibración), para los meses de junio y Julio, separados por estación meteorológica. Fuente: Elaboración propia

Estación meteorológica	Junio							Julio						
	β_0	P _{RATE}	Q700	Z850	Z500	r	MAE	β_0	P _{RATE}	Q700	Z850	Z500	r	MAE
1	278	0,445	-	-0,18	-	0,66	8,9	129,5	1,141	-	-0,08	-	0,77	7,2
2	350,3	0,367	3938	-	-0,062	0,54	11,8	300,5	1,142	-2255	-	-0,052	0,78	7,3
3	122	0,36	-	-0,08	-	0,68	6,9	175,7	0,453	-	-	-0,031	0,74	4,7
4	283,6	0,655	-	-0,18	-	0,66	13,8	220,8	1,164	-	-	-0,038	0,83	8,1
5	416,5	0,461	2734	-	-0,073	0,61	11,1	271,6	0,951	-	-	-0,047	0,67	8,3
6	170,9	0,494	-	-0,11	-	0,68	11,1	6,385	0,637	-	-	-	0,73	7,6
7	455,3	0,731	-	-	-0,079	0,61	16,8	233,9	0,721	-	-0,15	-	0,58	13,4
8	576,6	1,00	-	-	-0,101	0,67	14,6	328,5	1,459	-	-	-0,058	0,77	9,4
9	292	0,521	-	-	-0,051	0,75	9,4	168,7	0,37	-	-	-0,029	0,68	6,0
10	307,8	0,852	-	-	-0,054	0,70	10,2	394,5	0,881	-	-	-0,070	0,62	9,1
11	301,7	0,612	-	-0,20	-	0,67	13,7	0,252	1,472	2792	-	-	0,83	10,1
12	228,8	0,603	-	-0,15	-	0,75	13,3	3,503	0,829	-	-	-	0,75	9,7
13	233	0,524	2626	-	-0,041	0,71	11,5	177,9	0,671	-	-	-0,031	0,75	7,6
14	586,3	0,678	-	-	-0,103	0,69	15,5	264,7	0,803	-	-0,17	-	0,78	10,6
15	208,8	0,391	-	-	-0,036	0,67	7,5	162,4	0,368	-	-0,11	-	0,70	5,4
16	167,2	0,354	1675	-	-0,029	0,68	7,4	126,2	0,502	-	-	-0,022	0,77	5,5
17	234	0,468	-	-	-0,041	0,68	8,4	121,9	0,466	-	-0,08	-	0,69	5,8
18	220,7	0,513	-	-	-0,039	0,70	8,1	131,8	0,501	-	-0,09	-	0,72	5,6
19	300	0,449	-	-	-0,053	0,65	9,7	183,7	0,419	-2743	-0,12	-	0,77	5,0
20	186,4	0,728	-	-	-0,033	0,73	8,8	117,6	0,644	-	-0,08	-	0,74	5,8
21	231,3	0,534	-	-	-0,040	0,77	9,3	4,687	0,645	-	-	-	0,79	6,3
22	133,1	0,49	-	-0,08	-	0,72	8,5	167,8	0,594	-	-	-0,029	0,71	6,8
23	294,2	0,751	3089	-	-0,052	0,76	13,9	6,988	0,955	-	-	-	0,74	10,0
24	5,448	0,705	-	-	-	0,72	8,9	76,91	0,679	-	-0,05	-	0,76	6,2
25	158,3	0,511	2308	-0,10	-	0,77	9,2	117,2	0,435	-	-0,07	-	0,65	6,4
26	4,0	0,698	4529	-	-	0,69	12,0	11,72	0,983	-2783	-0,15	0,040	0,77	7,3
27	8,852	0,753	-	-	-	0,64	11,5	-136	0,782	-3967	-0,08	0,046	0,67	7,7
28	528,6	0,471	5121	-	-0,093	0,64	14,3	10,7	0,797	-	-	-	0,59	11,3
29	4,686	1,319	-	-	-	0,70	16,0	3,045	1,29	-	-	-	0,78	9,6

Tabla 13 Constantes de cada modelo de regresión obtenido para la serie temporal del periodo Reforecast (Calibración), para los meses de agosto y septiembre, separados por estación meteorológica. Fuente: Elaboración propia

Estación meteorológica	Agosto							Septiembre						
	β_0	P _{RATE}	Q700	Z850	Z500	r	MAE	β_0	P _{RATE}	Q700	Z850	Z500	r	MAE
1	197,1	1,051	2927	-	-0,035	0,80	7,7	157,6	0,848	-	-0,10	-	0,68	7,5
2	4,335	1,163	-	-	-	0,79	7,8	2,54	1,10	-	-	-	0,73	5,9
3	146	0,598	-	-	-0,026	0,74	5,0	3,083	0,46	-	-	-	0,73	6,5
4	9,643	1,234	-	-	-	0,80	9,5	8,613	0,791	-	-	-	0,59	8,8
5	279,6	0,799	-	-	-0,049	0,77	8,2	149,6	0,65	-	-	-0,026	0,80	4,2
6	7,693	0,812	-	-	-	0,77	9,2	184,3	0,556	-	-0,12	-	0,77	6,7
7	252,6	1,172	-6282	-0,16	-	0,70	11,6	7,168	1,077	-	-	-	0,68	9,4
8	351,2	0,919	-	-	-0,062	0,79	8,5	137,9	0,81	-	-0,09	-	0,81	5,3
9	5,599	0,667	-2247	-	-	0,76	6,4	1,595	0,572	-	-	-	0,82	4,5
10	184	1,069	-	-	-0,032	0,79	7,9	205,3	1,125	-	-	-0,036	0,80	6,3
11	370,6	0,855	-	-	-0,065	0,77	10,1	336,6	0,80	-3993	-0,22	-	0,70	9,8
12	4,653	0,925	-	-	-	0,79	11,3	8,258	0,414	-	-	-	0,56	8,3
13	3,048	0,748	-	-	-	0,84	6,2	1,426	0,613	-	-	-	0,78	4,9
14	5,394	0,951	-	-	-	0,80	8,8	8,309	0,382	-	-	-	0,43	7,4
15	2,656	0,506	-	-	-	0,76	5,0	2,697	0,338	-	-	-	0,66	3,7
16	-29,2	0,616	2727	0,10	-0,020	0,84	5,3	108,5	0,425	-	-0,07	-	0,70	5,5
17	2,834	0,684	-	-	-	0,78	5,9	79,96	0,336	-	-0,05	-	0,69	4,6
18	3,612	0,687	-	-	-	0,75	6,3	54,09	0,349	1562	-0,04	-	0,72	4,4
19	3,707	0,5	-	-	-	0,73	5,9	0,446	0,271	2717	-	-	0,77	3,7
20	3,949	0,946	-	-	-	0,68	8,1	3,389	0,645	-	-	-	0,60	5,9
21	4,496	0,749	-	-	-	0,79	7,2	2,643	0,436	1953	-	-	0,68	5,8
22	7,507	0,617	-	-	-	0,66	8,1	6,645	0,411	-	-	-	0,50	6,8
23	6,713	1,013	-	-	-	0,68	11,7	130,8	0,547	-	-0,08	-	0,66	8,0
24	4,974	0,822	-	-	-	0,75	7,0	88,48	0,341	-	-0,06	-	0,55	6,4
25	-68,8	0,761	-2558	-0,14	0,050	0,73	8,1	110,5	0,264	-	-0,07	-	0,51	7,4
26	-116	1,058	-	-0,12	0,055	0,78	8,0	8,387	0,37	-	-	-	0,30	7,5
27	269,3	0,641	1568	-	-0,047	0,72	6,2	185,2	0,051	3767	0,07	-0,052	0,33	6,8
28	7,449	1,013	-	-	-	0,66	11,8	251,9	0,468	-	-	-0,043	0,49	9,3
29	-0,61	1,086	5183	-	-	0,67	13,9	9,548	0,486	-	-	-	0,30	10,9

Tabla 14 Agrupación de la cantidad de modelos de regresión de precipitación en función de las variables seleccionadas por el método hacia adelante.

	PRATE	Q700 PRATE	Z850 PRATE	Z500 PRATE	Z850 Q700 PRATE	Z500 Q700 PRATE	Z500 Z850 PRATE	Z500 Z850 Q700 PRATE
Abril	4	3	5	12	1	3	1	0
Mayo	6	1	5	9	1	5	2	0
Junio	3	1	7	11	1	6	0	0
Julio	6	1	9	9	1	1	0	2
Agosto	16	2	0	5	1	2	1	2
Septiembre	13	2	8	3	2	0	0	1
Total	48	10	34	49	7	17	4	5
%	28%	6%	20%	28%	4%	10%	2%	3%

6.4. VERIFICACIÓN DE SUPUESTOS

Para la verificación de los 174 modelos de regresión lineal multivariada generados, en el periodo de calibración (1995-2010), se deben cumplir los supuestos señalados en la sección 3.3.3. La Tabla 15, indica las condiciones para que cada supuesto se cumpla. Este ítem se realizó utilizando los resultados del *software* R.

Tabla 15 Herramientas y criterios de verificación del cumplimiento de los supuestos de linealidad.

Supuesto	Herramientas Graficas	Herramientas numéricas	
		Test	Cumplimiento
Linealidad	Diagramas de regresión parcial.	-	-
Independencia	-	Durbin-Watson	$1,5 < DW < 2,5$
Homocedasticidad	Diagrama de dispersión de pronósticos y residuos tipificados.	Breush-Pagan	p-valor < 0.05
Normalidad	Graficas P-P y Q-Q.	Shapiro-Wilk o Kolmogorov-Smirnov	p-valor < 0.05
No-colinealidad	-	FIV	$FIV < 10$

Considerando que existe mucha información a mostrar, se presentan los resultados solo para tres estaciones de forma representativa. Para ello, se utilizan las estaciones representativas de los tres clusters (sección 6.1): estación 4 Los Guindos para el cluster costa, estación 17 Quinta Normal Santiago para el cluster valle y estación 26 San José de Maipo Retén para el cluster cordillera.

6.4.1. LINEALIDAD

Al observar la primera columna de la Figura 41, Figura 42 y Figura 43 presente en el Anexo E, se muestra la relación entre la variable a pronosticar (P_{obs}) y cada una de las cinco variables regresoras. Como resultado, se demuestra de forma gráfica que las variables de presión (PRMSL, Z500 y Z850) poseen una tendencia lineal negativa con respecto a la P_{obs} , mientras que para PRATE y Q700 se observan resultados con una tendencia lineal positiva, acorde con las correlaciones obtenidas.

6.4.2. INDEPENDENCIA

Para verificar la independencia de los residuos del modelo de regresión lineal multivariada, se debe comprobar que el valor del *test* Durbin-Watson (DW) debe oscilar entre 1,5 y 2,5. En el Anexo D se encuentran los valores del estadístico DW para cada uno de los modelos generados. En la Tabla 16 (valores numéricos en Anexo D) se muestra si los modelos cumplen con la condición de independencia, donde “Sí” indica que el modelo cumple con el supuesto de independencia de residuos y “No” que no cumple con este supuesto.

Tabla 16: Cumplimiento del test de independencia Durbin-Watson

Estación pluviométrica		Abr	May	Jun	Jul	Ago	Sep
1	RAPEL	Sí	Sí	Sí	Sí	Sí	Sí
2	SANTO DOMINGO AD.	Sí	No	Sí	Sí	Sí	Sí
3	ESTERO PUANGUE RUTA 78	Sí	Sí	Sí	Sí	Sí	Sí
4	LOS GUINDOS	Sí	Sí	Sí	Sí	Sí	Sí
5	QUILLOTA	Sí	Sí	No	Sí	Sí	Sí
6	BARRERA LONCHA	Sí	Sí	Sí	Sí	Sí	Sí
7	RODELILLO	Sí	Sí	Sí	Sí	Sí	Sí
8	LO ROJAS	Sí	Sí	Sí	Sí	Sí	No
9	LOS PANGUILES	Sí	Sí	Sí	Sí	Sí	Sí
10	LAGUNILLAS	No	Sí	Sí	Sí	Sí	Sí
11	LITUECHE	Sí	Sí	Sí	Sí	Sí	Sí
12	COLTAUCO	Sí	Sí	Sí	Sí	Sí	Sí
13	EL VERGEL	Sí	Sí	Sí	Sí	Sí	Sí
14	LAGUNA ACULEO	Sí	Sí	Sí	Sí	Sí	Sí
15	PUDAHUEL SANTIAGO	Sí	Sí	Sí	Sí	Sí	Sí
16	RANCAGUA (CACHAPOAL-DCP)	Sí	Sí	Sí	Sí	Sí	Sí
17	QUINTA NORMAL SANTIAGO	Sí	Sí	Sí	Sí	Sí	Sí
18	TERRAZAS OFICINA CENTRAL DGA	Sí	Sí	Sí	Sí	Sí	Sí
19	HUECHUN EMBALSE	Sí	Sí	Sí	Sí	Sí	Sí
20	EULOGIO SANCHEZ AD	Sí	Sí	Sí	Sí	Sí	Sí
21	PIRQUE	Sí	Sí	Sí	Sí	Sí	Sí
22	QUEBRADA RAMON	Sí	No	Sí	Sí	Sí	Sí
23	LA OBRA RECINTO REMOS	No	Sí	Sí	Sí	Sí	Sí
24	CERRO CALAN	Sí	Sí	Sí	Sí	Sí	Sí
25	ANTUPIREN	Sí	Sí	Sí	Sí	Sí	Sí
26	SAN JOSE DE MAIPO RETEN	Sí	Sí	Sí	Sí	Sí	Sí
27	MAITENES BOCATOMA	Sí	Sí	Sí	Sí	Sí	Sí
28	SAN GABRIEL	Sí	Sí	Sí	Sí	Sí	Sí
29	EL YESO EMBALSE	Sí	Sí	Sí	Sí	Sí	Sí

6.4.3. HOMOCEDASTICIDAD

Para verificar el supuesto de homocedasticidad, se deben analizar los residuos de los modelos de regresión lineal multivariada. El fin de este supuesto, es corroborar que los residuos posean una varianza constante. En la Figura 28 y Figura 29 se muestran dos casos distintos de resultados gráficos. La Figura 28 corresponde al mes de junio en la estación 17 (Quinta Normal), se observa una tendencia lineal de los residuos (línea roja), por lo que no cumple con el supuesto de homocedasticidad. Mientras que en la Figura 29 se presentan los resultados para la estación 7 (Rodelillo), donde la varianza de los residuos si es constante (línea roja), por lo que si se cumple el supuesto de homocedasticidad.

La

Tabla 17, muestra los modelos de regresión lineal que cumplen con el criterio de homocedasticidad por medio del del test de Breush-Pagan (*software R*) para obtener un análisis cuantitativo del cumplimiento de este supuesto. Donde “Sí” indica que el modelo cumple con el supuesto de homocedasticidad y “No” que no cumple con este supuesto. Los valores de los p-valores se puede encontrar en el Anexo D.

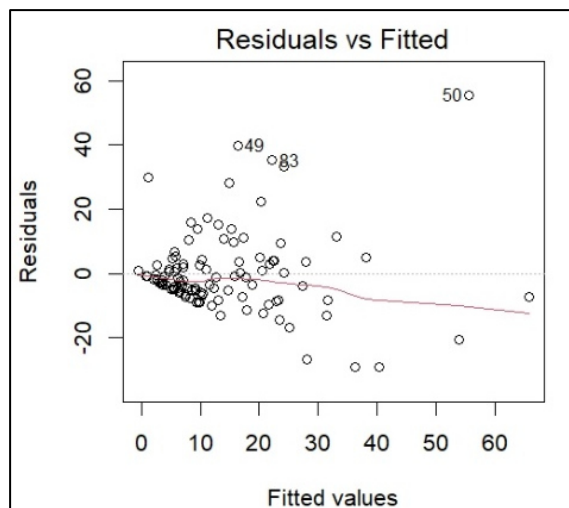


Figura 28: Residuos del modelo de regresión en la estación 17 (Quinta Normal), mes de junio. Fuente: Elaboración propia.

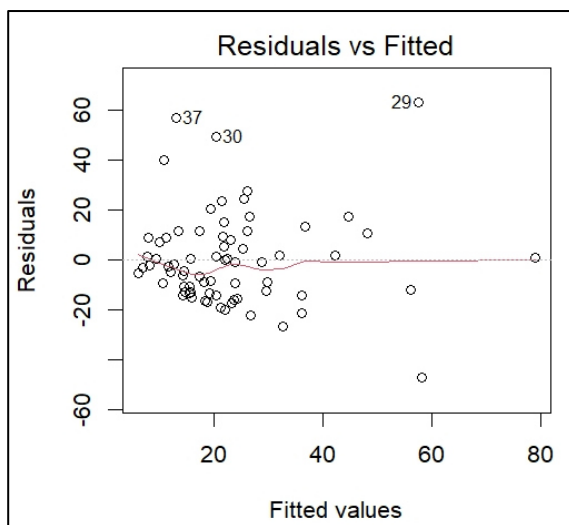


Figura 29: Residuos del modelo de regresión en la estación 7 (Rodelillo), mes de julio. Fuente: Elaboración propia.

Tabla 17: cumplimiento de la homocedasticidad de los residuos del modelo lineal para cada estación y mes. Fuente: elaboración propia.

Estación pluviométrica		Abr	May	Jun	Jul	Ago	Sep
1	RAPEL	No	No	No	No	No	No
2	SANTO DOMINGO AD.	No	No	Sí	No	No	No
3	ESTERO PUANGUE RUTA 78	Sí	Sí	No	No	Sí	No
4	LOS GUINDOS	Sí	Sí	No	Sí	Sí	Sí
5	QUILLOTA	Sí	Sí	No	Sí	No	Sí
6	BARRERA LONCHA	Sí	No	No	No	No	No
7	RODELILLO	Sí	Sí	No	Sí	Sí	No
8	LO ROJAS	Sí	Sí	No	No	No	No
9	LOS PANGUILES	No	No	No	Sí	Sí	No
10	LAGUNILLAS	No	No	No	No	Sí	No
11	LITUECHE	No	No	No	No	No	No
12	COLTAUCO	Sí	No	No	No	No	Sí
13	EL VERGEL	No	No	No	No	No	No
14	LAGUNA ACULEO	Sí	No	No	No	Sí	Sí
15	PUDAHUEL SANTIAGO	No	No	No	No	No	No
16	RANCAGUA (CACHAPOAL-DCP)	Sí	No	No	No	No	No
17	QUINTA NORMAL SANTIAGO	No	No	No	No	No	No
18	TERRAZAS OFICINA CENTRAL DGA	No	No	No	Sí	No	No
19	HUECHUN EMBALSE	Sí	No	No	Sí	No	No
20	EULOGIO SANCHEZ AD	No	No	No	Sí	Sí	Sí
21	PIRQUE	No	No	No	Sí	No	Sí
22	QUEBRADA RAMON	Sí	Sí	No	Sí	Sí	Sí
23	LA OBRA RECINTO REMOS	No	No	No	Sí	Sí	Sí
24	CERRO CALAN	Sí	No	No	Sí	Sí	Sí
25	ANTUPIREN	Sí	No	No	Sí	No	Sí
26	SAN JOSE DE MAIPO RETEN	Sí	No	No	Sí	Sí	Sí
27	MAITENES BOCATOMA	Sí	No	No	No	No	Sí
28	SAN GABRIEL	Sí	No	No	Sí	Sí	Sí
29	EL YESO EMBALSE	No	No	No	No	No	Sí

6.4.4. NORMALIDAD

La normalidad consiste en que todas las muestras deben estar distribuidas en una forma normal, tanto las variables de precipitación observada diaria (P_{obs}) como las variables atmosféricas (VA). Para verificar este supuesto, se pueden utilizar las gráficas Q-Q o el estadístico de Shapiro-Wilk. Pero antes de aplicar estas herramientas se presentan para su respectivo análisis, las series temporales de las variables en la Figura 30, en específico de la estación 17 (Quinta Normal) para el mes de junio.

Se presentan dos casos de graficas Q-Q, la Figura 31 correspondiente a la estación 17 mes de junio, y la Figura 32 de la estación 7 (Rodelillo) en el mes de julio; para ambos casos no se observa una tendencia lineal, es decir, no cumple con la normalidad.

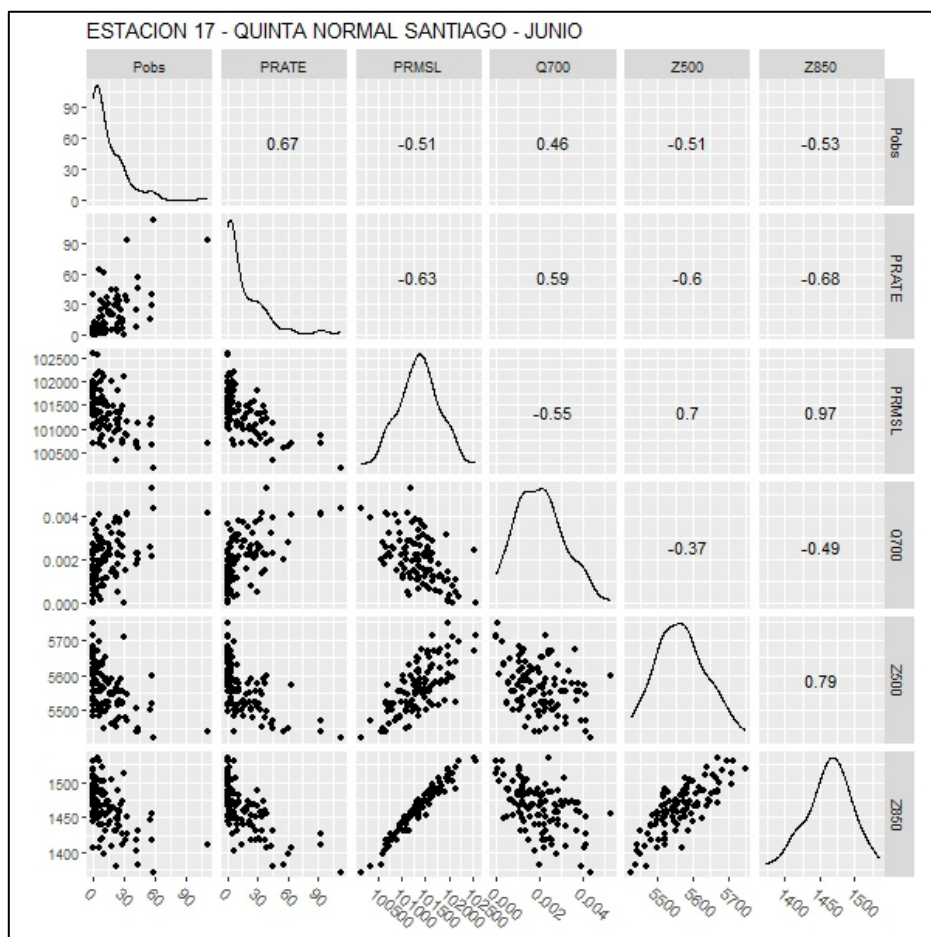


Figura 30 : Gráfica y correlación inter variable en estación 17.
Fuente: Elaboración propia

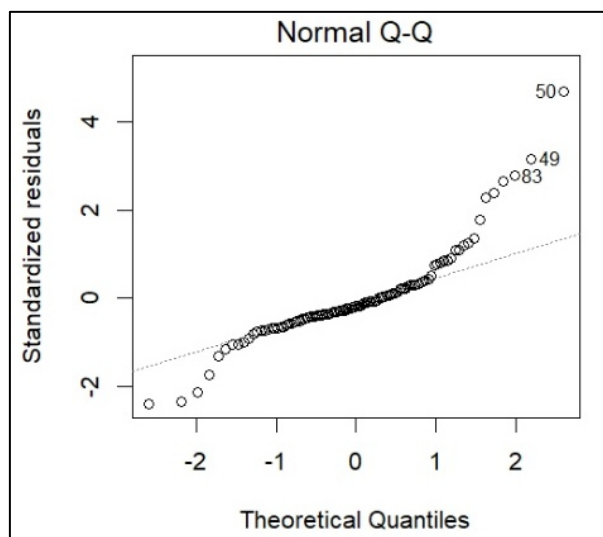


Figura 31: Gráfica Q-Q de la estación 17 (Quinta Normal) en el mes de junio. Fuente: Elaboración propia.

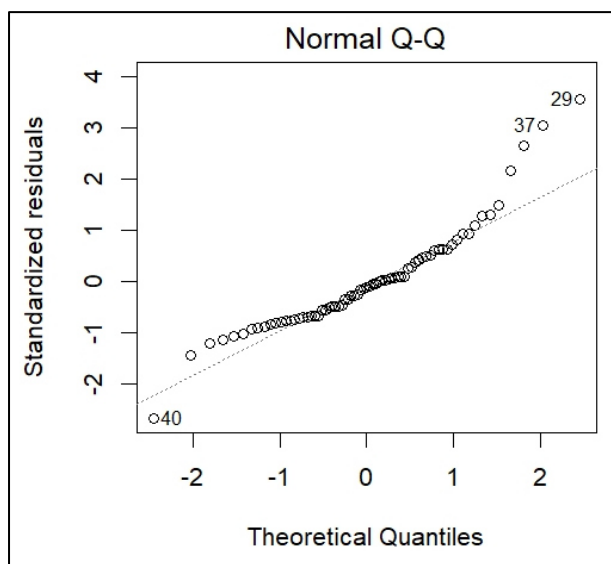


Figura 32: Gráfica Q-Q de la estación 7 (Rodelillo) en el mes de julio. Fuente: Elaboración propia.

En la Tabla 18 se muestra si los modelos de regresión cumplen con la condición de normalidad mediante el test de Shapiro-Wilk (valores numéricos en Anexo D), donde “Sí” indica que el modelo cumple con el supuesto de normalidad y “No” que no cumple con este supuesto.

Tabla 18: Test de normalidad. Fuente: elaboración propia

Estación pluviométrica		Abr	May	Jun	Jul	Ago	Sep
1	RAPEL	No	No	No	No	No	No
2	SANTO DOMINGO AD.	No	No	No	No	No	No
3	ESTERO PUANGUE RUTA 78	No	No	No	No	No	No

4	LOS GUINDOS	No	No	No	No	No	No
5	QUILLOTA	Sí	No	No	No	No	No
6	BARRERA LONCHA	No	No	No	No	No	No
7	RODELILLO	No	No	No	No	No	No
8	LO ROJAS	Sí	No	No	No	Sí	No
9	LOS PANGUILES	Sí	No	No	No	No	No
10	LAGUNILLAS	Sí	No	No	No	No	No
11	LITUECHE	Sí	No	No	No	No	No
12	COLTAUCO	No	No	No	No	No	No
13	EL VERGEL	No	No	No	No	No	No
14	LAGUNA ACULEO	No	No	No	No	No	No
15	PUDAHUEL SANTIAGO	No	No	No	No	No	No
16	RANCAGUA (CACHAPOAL-DCP)	No	No	No	No	No	No
17	QUINTA NORMAL SANTIAGO	No	No	No	No	No	No
18	TERRAZAS OFICINA CENTRAL DGA	No	No	No	No	No	No
19	HUECHUN EMBALSE	No	No	No	No	No	No
20	EULOGIO SANCHEZ AD	No	No	No	No	No	No
21	PIRQUE	No	No	No	No	No	No
22	QUEBRADA RAMON	No	No	No	No	No	No
23	LA OBRA RECINTO REMOS	No	No	No	No	No	No
24	CERRO CALAN	No	No	No	No	No	No
25	ANTUPIREN	Sí	No	No	No	No	No
26	SAN JOSE DE MAIPO RETEN	Sí	No	No	No	No	No
27	MAITENES BOCATOMA	No	No	No	Sí	No	No
28	SAN GABRIEL	No	No	No	No	No	No
29	EL YESO EMBALSE	No	No	No	No	No	No

6.4.5. NO COLINEALIDAD

Antes de generar los modelos de regresión lineal multivariada, se observa en el análisis de correlación (ver sección 6.2.3) entre las variables Z850 y PRMSL, que éstas poseen una alta correlación entre ellas. Por esta razón, se descartó la utilización de ambas en el modelo de regresión, utilizando solo Z850 y no PRMSL. De este modo, los modelos contemplan las variables PRATE, Q700, Z850 y Z500 (ver sección 6.3).

Para la verificación de este supuesto se utiliza el test de factor de inflación de la varianza (FIV). Los resultados de la verificación de no-colinealidad mediante el test de FIV se encuentran en el Anexo D. Se obtiene que todos los modelos generados no poseen colinealidad entre las variables atmosféricas utilizadas para la predicción de los valores de precipitación observada.

6.5. VALIDACIÓN DE LOS MODELOS

La validación de los modelos de regresión lineal multivariada tiene como objetivo verificar una buena performance del modelo, prediciendo valores fuera del periodo de ajuste. Esto se verifica mediante el coeficiente de correlación de Pearson (r , correlación lineal temporal) y el error absoluto medio (MAE, precisión)

En esta sección, se procede a determinar las correlaciones r y los MAEs para la fase validación (2011-2017), la que consiste en utilizar los modelos anteriormente obtenidos en la fase de calibración, pero ahora evaluados con los datos de las series temporales del periodo de validación. Los resultados son mostrados por medio de *boxplot* para su análisis.

6.5.1. CORRELACIÓN DE LOS MODELOS

En la Tabla 19 se presentan los valores del coeficiente de correlación de Pearson (r) para ambos periodos (*reforecast*/calibración y *operational*/validación), entre la precipitación observada diaria (P_{obs}) y la precipitación pronosticada (P_{pron}) por los modelos.

En la Tabla 19, los valores se muestran con distintos colores dependiendo de su nivel de correlación de acuerdo a la Tabla 5: muy baja con color rojo (0,0 – 0,3), baja color café (0,3 – 0,5), media color verde (0,5 – 0,7), alta color azul (0,7 – 0,9) y muy alta color morado (0,9 – 1,0). Para facilitar la comparación se presentan los valores de correlación en dos gráficos *boxplot*, un *boxplot* para el periodo de calibración (Figura 33) y uno para el periodo de validación (Figura 34).

Tabla 19: Correlación entre P_{pron} y P_{obs} para el periodo Operational. Los colores corresponden a los niveles de correlación de la Tabla 5: muy baja (rojo), baja (café), media (verde), alta (azul) y muy alta (morado). Fuente: Elaboración propia.

Estaciones meteorológicas		Correlación fase Operational					
		Abr	May	Jun	Jul	Ago	Sep
1	RAPEL	0,82	0,75	0,39	0,33	0,58	0,94
2	SANTO DOMINGO AD.	0,81	0,88	0,75	0,76	0,74	0,73
3	ESTERO PUANGUE RUTA 78	0,65	0,54	0,82	0,78	0,76	0,85
4	LOS GUINDOS	0,74	0,97	0,62	0,55	0,69	0,17
5	QUILLOTA	0,00	0,82	0,69	0,54	0,93	0,17
6	BARRERA LONCHA	0,48	0,76	0,25	0,36	0,74	0,26
7	RODELILLO	0,88	0,67	0,81	0,60	0,61	0,78
8	LO ROJAS	0,68	0,81	0,65	0,71	0,88	-0,03
9	LOS PANGUILES	0,73	0,87	0,56	0,71	0,82	0,70
10	LAGUNILLAS	0,48	0,88	0,80	0,68	0,29	0,67
11	LITUECHE	0,48	0,91	0,51	0,73	0,82	0,66
12	COLTAUCO	0,86	0,93	0,75	0,82	0,68	0,80
13	EL VERGEL	0,46	0,94	0,75	0,73	0,92	0,80
14	LAGUNA ACULEO	0,91	0,94	0,52	0,80	0,71	0,84
15	PUDAHUEL SANTIAGO	0,52	0,78	0,52	0,72	0,88	0,62

16	RANCAGUA (CACHAPOAL-DCP)	0,84	0,78	0,57	0,68	0,81	0,52
17	QUINTA NORMAL SANTIAGO	0,61	0,76	0,54	0,83	0,90	0,32
18	TERRAZAS OFICINA CENTRAL DGA	0,68	0,81	0,52	0,74	0,86	0,07
19	HUECHUN EMBALSE	0,36	0,81	0,58	0,57	0,89	-0,10
20	EULOGIO SANCHEZ AD	0,66	0,77	0,59	0,78	0,73	0,39
21	PIRQUE	0,42	0,69	0,69	0,85	0,88	0,28
22	QUEBRADA RAMON	0,49	0,71	0,54	0,67	0,76	0,60
23	LA OBRA RECINTO REMOS	0,52	0,90	0,54	0,82	0,83	0,32
24	CERRO CALAN	0,53	0,61	0,61	0,76	0,88	0,27
25	ANTUPIREN	0,67	0,81	0,61	0,79	0,81	0,31
26	SAN JOSE DE MAIPO RETEN	0,25	0,90	0,45	0,93	0,77	0,43
27	MAITENES BOCATOMA	0,47	0,69	0,39	0,60	0,18	0,24
28	SAN GABRIEL	0,34	0,69	0,61	0,51	0,69	0,29
29	EL YESO EMBALSE	0,50	0,88	0,60	0,50	0,74	0,39

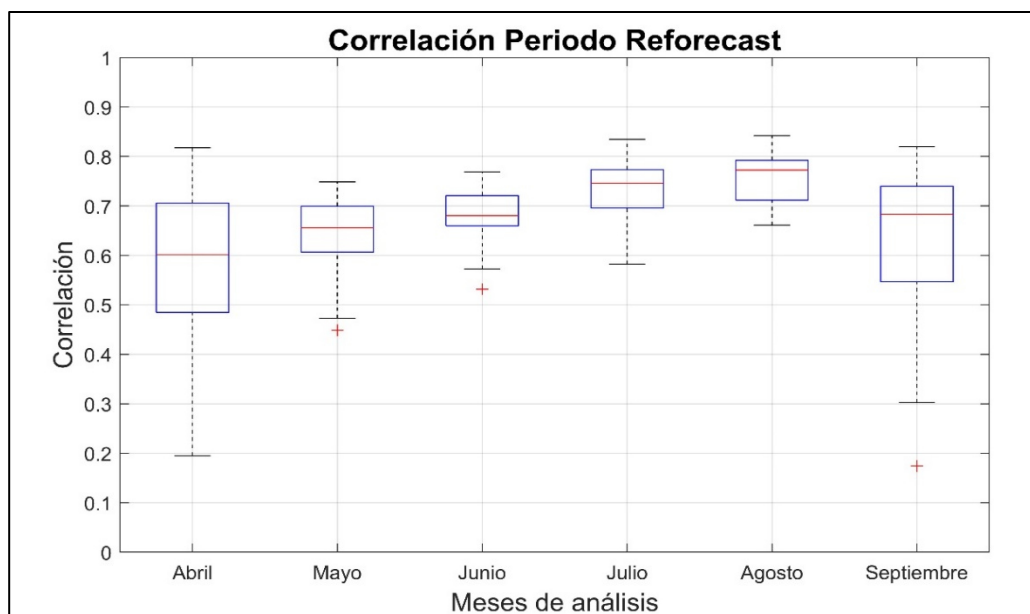


Figura 33: Correlación de los modelos periodo reforecast (calibración).
Fuente: Elaboración propia

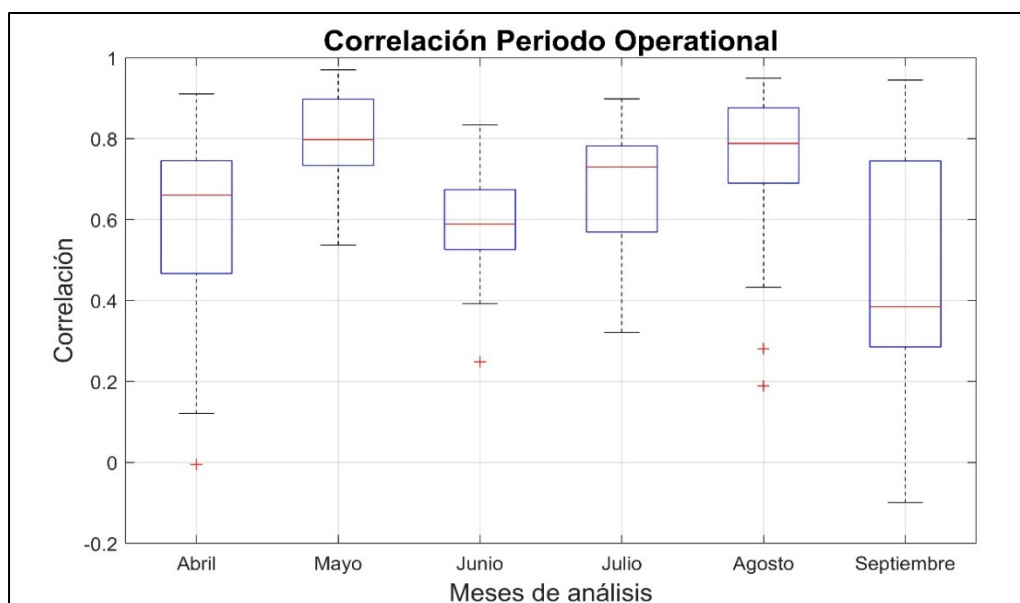


Figura 34: Correlación de los modelos periodo operacional (validación).
Fuente: Elaboración propia.

6.5.2. ERROR ABSOLUTO MEDIO DE LOS MODELOS

Se utiliza el error absoluto medio (MAE), para cuantificar la precisión de los modelos de regresión lineal multivariada, el valor del MAE indica la magnitud del error (residuo) entre el pronóstico del modelo y la variable pronosticada. Es decir, un error con unidades de milímetros (mm) entre la precipitación observada y la pronosticada. El análisis se realiza para ambos periodos: *reforecast* y *operational*.

En la Tabla 20, se presenta el MAE para cada uno de los 174 modelos correspondiente a cada estación, a cada mes y a cada periodo. Para facilitar la comparación se presentan los valores del MAE en dos gráficos *boxplot*, un *boxplot* para el periodo de calibración (Figura 35) y uno para el periodo de validación (Figura 36).

Tabla 20: MAE entre los modelos y Pobs para el periodo Operational (validación). Colores de acuerdo con el resultado: rojo (15-20 mm), verde (10-15 mm), azul (5-10 mm) y morado (0-5 mm). Fuente: Elaboración propia.

Estaciones meteorológicas		MAE fase Operational					
		Abr	May	Jun	Jul	Ago	Sep
1	RAPEL	4,2	10,6	14,3	9,9	8,9	4,1
2	SANTO DOMINGO AD.	4,1	7,3	7,2	5,5	5,1	3,1
3	ESTERO PUANGUE RUTA 78	7,8	8,6	6,5	3,6	4,6	4,5
4	LOS GUINDOS	9,1	10,0	10,2	8,9	9,4	5,4
5	QUILLOTA	13,6	7,5	9,6	7,9	5,6	5,5
6	BARRERA LONCHA	5,0	14,4	11,3	9,1	8,3	7,1

7	RODELILLO	3,6	12,8	8,4	12,6	9,8	7,0
8	LO ROJAS	11,8	11,4	11,5	7,4	8,0	6,9
9	LOS PANGUILES	6,3	5,4	9,0	4,9	4,8	2,7
10	LAGUNILLAS	11,1	9,6	7,1	5,8	9,0	5,1
11	LITUECHE	10,2	10,2	12,9	8,1	7,1	5,2
12	COLTAUCO	12,0	10,6	9,6	8,3	9,4	6,4
13	EL VERGEL	10,6	5,6	8,1	5,4	4,0	3,2
14	LAGUNA ACULEO	11,1	7,1	13,8	8,2	7,3	8,3
15	PUDAHUEL SANTIAGO	6,3	4,7	7,3	3,8	3,3	1,8
16	RANCAGUA (CACHAPOAL-DCP)	3,9	7,1	7,6	4,3	5,2	3,5
17	QUINTA NORMAL SANTIAGO	5,8	5,2	8,2	3,7	3,5	3,4
18	TERRAZAS OFICINA CENTRAL DGA	2,0	4,1	7,3	4,8	5,2	3,4
19	HUECHUN EMBALSE	5,2	5,3	8,0	4,2	4,6	4,6
20	EULOGIO SANCHEZ AD	6,2	5,5	8,4	4,2	6,0	3,8
21	PIRQUE	8,8	7,0	7,8	5,4	5,1	5,3
22	QUEBRADA RAMON	9,4	8,1	7,6	6,7	6,4	6,0
23	LA OBRA RECINTO REMOS	12,5	8,0	13,1	8,4	8,0	7,0
24	CERRO CALAN	6,8	7,1	8,6	4,5	5,5	4,9
25	ANTUPIREN	6,5	5,4	8,7	5,7	6,1	5,4
26	SAN JOSE DE MAIPO RETEN	10,3	6,2	12,1	4,8	7,6	7,0
27	MAITENES BOCATOMA	8,0	6,2	11,1	7,4	8,0	6,4
28	SAN GABRIEL	9,1	9,6	12,8	12,3	7,0	6,0
29	EL YESO EMBALSE	7,0	6,7	12,0	9,0	8,7	7,9

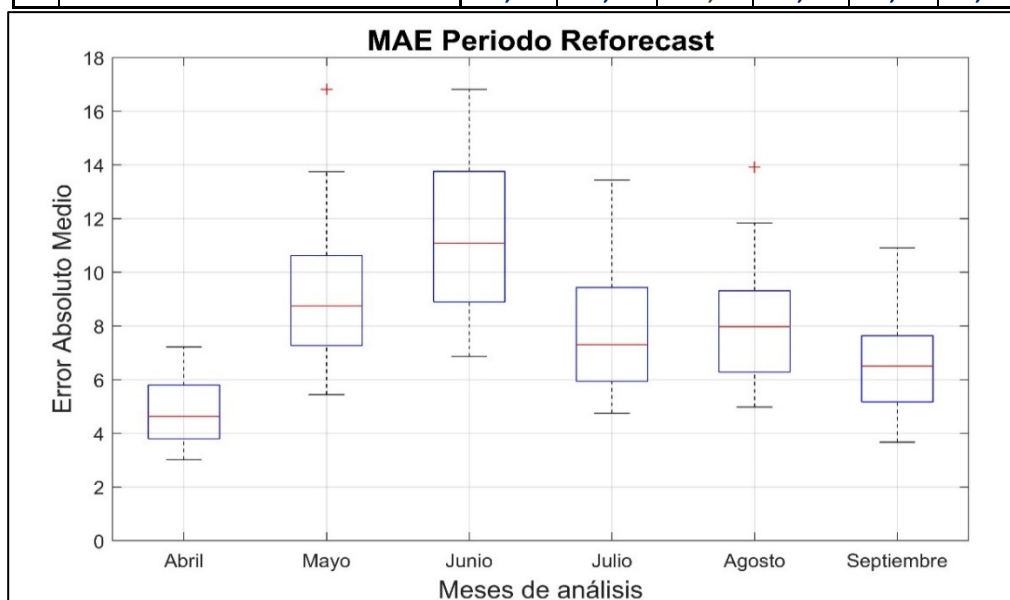


Figura 35: MAE de los modelos periodo reforecast (calibración).

Fuente: Elaboración propia.

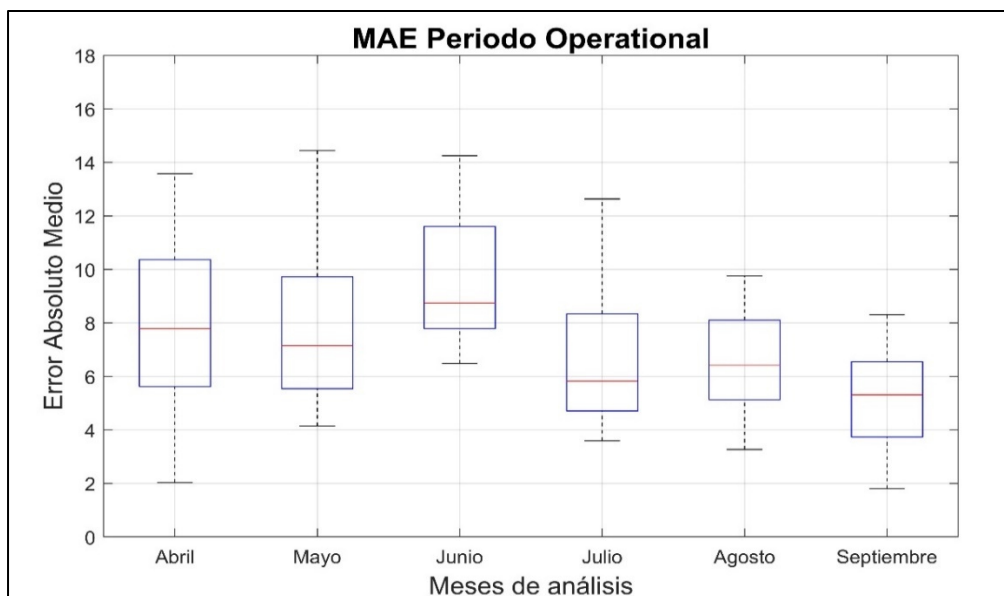


Figura 36: MAE de los modelos periodo operacional (validación).
Fuente: Elaboración propia.

7. DISCUSIÓN DE RESULTADOS

En esta sección se procede a discutir los resultados presentados en la sección anterior, manteniendo el orden de los puntos tratados.

7.1. ANÁLISIS CLUSTER

El método *Single* (Figura 12) permitió anticipar el comportamiento de estaciones meteorológicas atípicas, por ejemplo, el comportamiento de la estación 14 (ver Figura 14), ya que perteneciendo al cluster costa se ubica dentro del cluster valle sin una explicación espacial al respecto.

Por otra parte, el método *Single* también advirtió sobre las estaciones 1 y 11, éstas en el método Ward se unen en el último paso al cluster costa (Figura 13), lo que coincide con su ubicación espacial cercana entre ellas y apartada del resto (ver Figura 14). Las otras tres estaciones consideradas atípicas (23, 28 y 29) pertenecen al cluster cordillera, de esta forma el 60% de las estaciones del cluster cordillera son atípicas, lo cual puede deberse a su gran variabilidad de altura dentro del pequeño grupo de cinco estaciones. Referente a la altitud, el cluster cordillera presenta una desviación estándar de 595 m, superior a la desviación en el cluster costa de 98 m, o el cluster valle de 230 m. Cabe notar, que la estación 29 es fuertemente atípica de acuerdo con el método *Single*, y que también es la estación con mayor altitud (2483 msnm), bastante superior al resto, puesto que, la segunda estación ubicada a mayor altitud es la estación 28 (1256 msnm).

Debido a la gran cantidad de gráficos generados, en los análisis posteriores, se ha seleccionado una estación representativa de cada cluster, con el fin de entregar información detallada y acotada a estas tres estaciones, sin perjuicio de entregar los resultados generales para cada estación. Se escogen las primeras estaciones asociadas a cada cluster de acuerdo con la Figura 13, descartando las estaciones atípicas entregadas por el método *Single* (Figura 12). De este modo, se escoge como estación representativa del cluster costa la estación 4 (Los Guindos), para el cluster valle la estación 17 (Quinta Normal Santiago) y para el cluster cordillera la estación 26 (San José de Maipo Retén). Además, estas estaciones seleccionadas se ubican en un lugar central al cluster que pertenecen.

7.2. CORRELACIONES

7.2.1. CORRELACIÓN ESPACIAL

Respecto a las correlaciones espaciales (Figuras 15 a la 18) entre la precipitación observada diaria y cada una de las variable atmosféricas, se observan cuatro situaciones con respecto a la ubicación del punto de mejor correlación.

Primero, la gráfica de correlación espacial entre la P_{obs} y la variable atmosférica de precipitación PRATE (Figura 15), evidencia que el punto de mejor correlación se encuentra cercano a la estación (con valores entre 0,59 a 0,76). En este caso, los métodos que utilizan los puntos más cercanos son válidos.

Segundo, la variable de humedad Q700 (Figura 15) muestra que el punto de mejor correlación se encuentra al norte de la estación aproximadamente a 1,5 o 2 grados de latitud, es decir, 1 o 2 puntos de la grilla hacia el norte desde el punto más cercano por el oeste de la estación, con valores en los puntos de mayor correlación entre 0,33 a 0,44.

Tercero, las variables atmosféricas de presión, Z500, Z850 y PRMSL (Figura 17 y Figura 18), presentan una correlación negativa coherente con lo discutido en la sección 3.4.5, con valores entre -0.26 a -0.44 para Z500, -0.41 a -0.56 para Z850 y de -0.36 a -0.49 para PRMSL. A diferencia de las variables PRATE y Q700, en las gráficas espaciales para las VAs de presión no se observa ningún caso con un punto de mejor correlación encerrado por una isolínea, pero si se observa una tendencia creciente de la correlación, mostrada por las isolíneas hacia el sur de la estación en el caso de Z500 y al sureste en el caso de Z850 y PRMSL.

Como cuarta y última situación, para las variables de temperatura, T850 y TMP5FC (Figura 16), no se observan puntos claros de mayor correlación ni tendencia espacial, con valores entre -0.23 a 0.19 para T850 y -0.09 a 0.19 para TMP5FC.

7.2.2. CORRELACIÓN MENSUAL

A continuación, se analizan los resultados de correlación mensual entre la precipitación observada (P_{obs}) diaria y cada una de las variables atmosféricas trabajadas, de acuerdo al tipo de variable: precipitación, humedad, presión y temperatura.

La correlación entre P_{obs} y la variable de precipitación PRATE (Figura 19) presenta resultados medios o altos (ver Tabla 5) en los meses de mayo a agosto, mientras que para abril se obtienen

valores muy bajos en la estación 5 y 13, como también se observa una muy baja correlación para el mes de septiembre en la estación 27. Estos bajos valores de correlación se deben a una dispersión de los datos como se observa en la Figura 20, para el caso de la estación 13 en el mes de abril. Por otro lado, en la misma Figura 20 se observa el caso para el mes de agosto, con una alta correlación de los datos. La baja correlación en abril y septiembre puede deberse a una menor cantidad de días de lluvia; 2,2 días en promedio en el mes de abril y 3,9 días en promedio en el mes de septiembre, mientras los demás meses precipita en promedio 5,5 días al mes. Los meses con menor cantidad de datos (abril y septiembre) tienen menor peso para escoger el punto de mayor correlación, pudiendo ser el motivo de los resultados con mayor dispersión.

Con respecto a la variable atmosférica de humedad Q700 (Figura 21), la mayoría de los valores se encuentran en un rango bajo de correlación (ver Tabla 5). Esta variable no presenta inconvenientes en el mes de abril, por el contrario, se observan valores medios e incluso altos para este mes, por lo que esta variable podría ayudar a corregir los pronósticos para el mes de abril.

Para el caso de las variables atmosféricas de presión (Z500, Z850 y PRMSL) los resultados de correlación son negativos (Figura 22, Figura 23 y Figura 24). Las correlaciones en el mes de abril son irregulares para las variables de presión, al igual que en el caso de la variable PRATE. Por otra parte, las gráficas de Z850 y PRMSL son muy similares, lo que podría advertir de una posible colinealidad, es decir, que ambas variables tengan un comportamiento parecido, aportando con información similar al modelo de regresión lineal multivariada.

Para las variables de temperatura (T850 y TMP5FC) el resultado de correlación mensual no es muy favorable (Figura 25 y Figura 26), ya que los resultados obtenidos son bajos (Tabla 5) y varían entre valores positivos y negativos según estación meteorológica. Este comportamiento es más notorio para la variable T850, por lo que la variable TMP5FC podría ser una variable menos deficiente como predictor que la variable T850 para esta zona de estudio. Este comportamiento, puede deberse a sus resultados mostrados en la Figura 16, puesto que no se logró identificar con claridad el punto de mayor correlación. Puesto que ambas variables presentan correlaciones bajas, se descartarán estas dos variables de temperatura como potenciales predictores de la precipitación observada para la zona de estudio.

7.2.3. CORRELACIÓN ENTRE VARIABLES

En la Figura 27, se observa una muy alta correlación (Tabla 5) entre las variables Z850 y PRMSL, mientras que todos los demás resultados expresan correlaciones medias o bajas. Estos resultados, indican que las variables Z850 y PRMSL estarían entregando al modelo una información muy parecida, con correlaciones entre las series temporales de las variables Z850 y PRMSL mayores a 0,9. En la construcción de los modelos una de las dos variables siempre se eliminaría o no sería agregada para evitar la colinealidad. Es decir, tenemos dos variables que aportan de igual manera a los modelos que se buscan desarrollar, de hecho, serían series temporales muy semejantes si se normalizaran las series de ambas variables.

Para evaluar la calidad de ambas variables como potenciales predictores de la precipitación se calcula el promedio de las correlaciones entre cada una de estas dos VAs de presión (Z850 y PRMSL) y la P_{obs} , para las 29 estaciones entre los meses abril a septiembre. Obteniendo que la correlación

promedio de la variable Z850 es de -0,51 y para la variable PRMSL es de -0,44. Por lo anterior, a pesar de que la variable PRMSL es una buena alternativa como potencial predictor de la precipitación, se descarta, ya que entrega información similar a la variable Z850, pero ésta última proporciona mejores correlaciones.

7.3. CALIBRACIÓN

Los modelos fueron contruidos incluyendo obligatoriamente a la variable PRATE, con respecto a la utilización de las otras variables se resume en la Tabla 14. Se indica que el 28% de los modelos no incorpora más variables, predominantemente en los meses de agosto y septiembre. Mientras la variable Q700 se hace presente en el 22% de los modelos, la variable Z850 en el 29% de los modelos y la variable Z500 está en el 43%. Si contáramos solo con las variables PRATE y Z850 se tendría el 56% de los casos con los modelos de mejor correlación. Por otro lado, si se contara con PRATE, Z850 y Z500 se tendrían el 78% de los casos.

En el estudio de Vega (2018) se utilizaron las variables PRATE, Q700 y Z850. Sin embargo, en el presente estudio se muestra que la construcción de modelos con las variables anteriormente señaladas son la mejor opción en el 57% de los casos, pero se ha incorporado la variable Z500, la cual mejora la corrección de los modelos en el 43% de los casos restantes.

7.4. VERIFICACIÓN DE SUPUESTOS

7.4.1. LINEALIDAD

Las variables regresoras (PRATE, Q700, PRMSL, Z500 y Z850) presentan una relación lineal con la precipitación observada diaria, lo cual es favorable para la generación de los modelos. La variable de precipitación PRATE y la variable de humedad Q700 presentan una relación positiva con la P_{obs} , acorde con el fenómeno físico. Mientras que las variables de presión (PRATE, Z500 y Z850) presentan una relación lineal negativa, lo que implica que la presencia de una baja de presión aumenta la probabilidad de precipitación.

7.4.2. INDEPENDENCIA

De un total de 174 modelos generados, solamente seis de ellos no cumplen con el supuesto de independencia, que requiere un modelo de regresión lineal multivariado. De estos resultados, indican tres situaciones: (1) para 23 de las estaciones, el supuesto se cumple para todos los meses de estudio, (2) para 6 estaciones el supuesto no se cumple solo en un mes y (3) el supuesto se cumple para todas las estaciones en los meses de julio y agosto (Tabla 16).

7.4.3. HOMOCEDASTICIDAD

Los resultados del test de Breush-Pagan (

Tabla 17), indican que para este supuesto solo 63 de los 174 modelos obtenidos cumplen con que los residuos posean una varianza constante. Esto quiere decir, que existen valores del modelo generado con una gran diferencia con el valor de la precipitación observada. La precedencia de este tipo de diferencia genera que la varianza del modelo no sea constante.

Los modelos obtenidos son deterministas y no cumplen en su mayoría con la homocedasticidad, en estos casos en Wilby (2002) se señala que se pueden reproducir estocásticamente la distribución de las varianzas, inflando su valor para obtener resultados de varianza constante, cumpliendo así, con la homocedasticidad. Sin embargo, la aplicación de esta metodología escapa del alcance de este estudio.

7.4.4. NORMALIDAD

La Figura 30 de las series temporales de las variables atmosféricas Q700, Z500 y Z850 muestra claramente una distribución normal de las series. A diferencia de esto, para la variable precipitación observada (P_{obs}) y la variable atmosférica PRATE el comportamiento no es normal, pero sí similar entre ellas, debido a que ambas son variables de precipitación.

Al observar las gráficas Q-Q de la Figura 31 y Figura 32, se deduce que los modelos propuestos para las estaciones en los meses indicados en las figuras no cumplen con el supuesto de normalidad, esto se demuestra ya que, las gráficas Q-Q no presentan una línea recta y poseen curvas en los extremos, esto podría ser mejorado con el ajuste de funciones por tramos en estudios posteriores. Esto reafirma el comportamiento no normal tanto de la variable a predecir (P_{obs}) como para la variable de predicción (PRATE).

Al observar la Tabla 18, solo 8 de los 174 modelos cumplen con el supuesto de normalidad, es decir, que las variables utilizadas para la generación del modelo no siguen una distribución normal. Esto es claro, al observar las distribuciones de PRATE, variable atmosférica obligatoria, para todos los modelos que tienen una clara distribución no normal. Los modelos de regresión lineal multivariada, se basan en que sus variables presenten una distribución normal, y al no cumplir este supuesto, permite que otros supuestos tampoco se cumplan.

7.4.5. NO COLINEALIDAD

Realizar un análisis entre las variables permitió descartar a priori la variable PRMSL y continuar con otra variable que aportó información similar, la variable Z850. La incorporación de este paso, permite que los modelos generados se construyan con variables que aporten distinta información. De no haber realizado este paso, probablemente se habrían obtenido algunos modelos con PRMSL y otros con Z850, generando una mayor variación dentro de los 174 modelos resultantes, sin una mejora significativa en la correlación.

7.5. VALIDACIÓN

7.5.1. CORRELACIÓN DE LOS MODELOS

En el periodo de calibración, se obtuvieron resultados del r con un valor promedio de 0,68. Mientras, que en el periodo de validación el promedio de los valores del r disminuyó a 0,64, de todas maneras, se mantuvo en el nivel de correlación en un nivel medio (r medio 0,5 – 0,7).

Los resultados de la Tabla 19 que se muestran gráficamente en la Figura 33, advierten dos situaciones con respecto a los resultados a nivel mensual, por un lado, se tiene un grupo de meses (mayo a agosto) con valores de correlación de al menos 0,45 y un promedio de 0,70. Por otra parte, los meses de abril y septiembre presentan correlaciones inferiores, en abril desde 0,32 y un promedio de 0,60, y para septiembre con valores de correlación desde 0,30 y un promedio de 0,63.

La correlación más baja en los meses de abril y septiembre en comparación a los otros meses tiene relación con la menor influencia en la selección del punto de grilla a escoger, puesto que estos dos meses tienen la menor cantidad de días con registros de precipitación observada. Por otro lado, en la Figura 33 se observa que para los meses abril y septiembre es más grande el tamaño de la caja, indicando mayor dispersión en los valores del r . Las complicaciones de estos meses fueron visibilizadas en la sección 6.2.2, obteniéndose modelos menos precisos para estos meses (abril y septiembre).

Los resultados de correlación en el periodo calibración, pueden ser comparados con el estudio realizado por Vega en 2018. En aquel estudio, se utilizaron las mismas 29 estaciones meteorológicas, en 5 meses de análisis (de mayo a septiembre), por medio de dos modelos distintos: uno lineal y otro no lineal. Solo se comparan los modelos lineales, puesto que los modelos no lineales escapan al interés de este estudio. Los resultados lineales obtenidos en Vega (2018) a través del *software* SPSS para los meses del análisis, tienen los siguientes valores promedio de correlación (r): mes de mayo 0,46, junio 0,53, julio 0,46, agosto 0,43 y septiembre 0,41. En este estudio los valores para dichos meses en el mismo orden son: 0,64, 0,69, 0,73, 0,76 y 0,63. Al comparar ambos resultados, hay un aumento de la correlación de los modelos desde 0,46 a 0,69 en promedio. Esto se debe a tres diferencias entre ambos estudios, primero, Vega (2018) utilizó el punto de grilla más cercano a la estación, mientras que aquí se utiliza el punto de mejor correlación. Segundo, Vega (2018) aplica una transformación logarítmica a las variables de precipitación. Tercero, este estudio incorpora la variable Z500 en la construcción de los modelos.

En el periodo de validación se mantiene la dispersión de las correlaciones para los meses de abril y septiembre, reafirmando que en estos meses la correlación toma valores más dispersos. El promedio de la correlación para el mes de abril es 0,58, para septiembre 0,46, mientras que para los meses de mayo a agosto es 0,71 (Figura 34).

El caso más desfavorable entre los meses desde mayo a agosto, se presenta en la estación 27 en el mes de agosto. Donde el coeficiente de correlación en el periodo de validación fue de 0,18 (ver Tabla 19). Para el mes de abril el caso más desfavorable se presenta en la estación 5 (Quillota) con correlación 0,0; lo particular de este modelo, es que presenta una constante beta de PRATE con valor 1,57, siendo la constante que acompaña a PRATE de mayor valor entre los 174 modelos. Esto indica que la variable PRATE pesa más en el resultado de la precipitación pronosticada, pudiendo

ser este el motivo del mal ajuste. Por otro lado, en el mes de septiembre hay un resultado negativo en la estación 19, lo que indica que el modelo es inversamente proporcional, por lo que no es aplicable el modelo en este caso.

7.5.2. ERROR ABSOLUTO MEDIO DE LOS MODELOS

En el periodo de calibración resulta en promedio, un MAE de 8,0 mm. Esto indica, que en general los valores pronosticados varían en 8,0 mm con respecto al valor observado de precipitación. En el periodo de validación, este valor promedio del MAE disminuyó a 7,3 mm, lo que indica el mismo orden de magnitud para ambos periodos.

A nivel de estaciones, se obtienen valores menores a 10 mm en todos los meses en ambos periodos en doce estaciones (4, 9, 15, 16, 17, 18, 19, 20, 21, 22, 24 y 25). Los resultados más desfavorables se presentan en la estación 7 (Rodelillo), para los meses de mayo y junio, en el periodo de calibración, con un valor del MAE de 16,8 mm (Tabla 20).

En la Figura 35, se presenta los valores del MAE para el periodo de calibración. En el cual, se observa que el mes con menor MAE corresponde al mes de abril, con un promedio de 4,9 mm, y el siguiente mes con menor MAE es el mes de septiembre con un valor promedio de 6,6 mm. Mientras que el mes de junio, es el mes con menor precisión en la fase de calibración, con un promedio del MAE de 11,1 mm.

De forma similar, los valores del MAE obtenidos para el periodo de validación se encuentran en la Figura 36. Lo primero que llama la atención, es el aumento del MAE para el mes de abril con un valor promedio de 7,9 mm, sumado a un aumento en la dispersión. Mientras, que en septiembre el valor promedio del MAE disminuye a 5,2 mm. El MAE más alto persiste para el mes de junio en el periodo de validación con un valor de 9,7 mm, lo cual podría deberse, a que es el mes que presenta valores más altos de precipitación diaria en comparación a los otros meses. Es decir, mayor cantidad de días con precipitación que podrían considerarse extremas y que escapan de un comportamiento normal, generando un comportamiento irregular de los residuos (ver sección 6.4.3). En general, los modelos de regresión para meses con mayor precipitación (junio, julio y agosto) poseen mejores correlaciones, pero los valores del MAE son más altos.

Los resultados del MAE en el periodo de validación, pueden ser comparados con los resultados del estudio de Vega (2018). En este análisis, el estudio de Vega utiliza los meses de mayo a septiembre y contempló una separación de los datos, definiendo dos grupos: uno correspondiente a precipitaciones moderadas y otro con precipitaciones extremas. El valor de precipitación que separa estos dos grupos varía por cada estación y mes de análisis. Los resultados del MAE para los modelos en precipitaciones moderadas se encuentran en un rango de 5 a 13 mm, mientras que en precipitaciones extremas los resultados fueron entre 15 a 25 mm. En este estudio los datos no fueron separados en grupos de precipitación moderada y extrema, por lo que, los resultados del MAE para el conjunto completo de datos se encuentran en un rango de 1,8 a 14,4 mm. Los resultados del MAE, en este estudio, durante la fase de validación se asemejan a los valores del MAE en precipitaciones moderadas de Vega(2018). Aunque este estudio contempla también los datos extremos.

En la Figura 37, se muestran los valores promedios del MAE para el periodo de validación en cada una de las 29 estaciones meteorológicas, agrupando los valores en colores en tres rangos. Esta gráfica espacial también muestra, las coordenadas de los puntos de ambas grillas de las variables de CFSv2 (grilla NCEP y grilla T126). Esta representación espacial se asemeja al resultado obtenido en el análisis cluster (ver Figura 14), donde al cruzar ambos gráficos se puede decir, que en general el cluster valle presenta los mejores resultados del MAE para este periodo de validación, el siguiente en obtener mejores resultados es el cluster cordillera y por último el cluster costa.

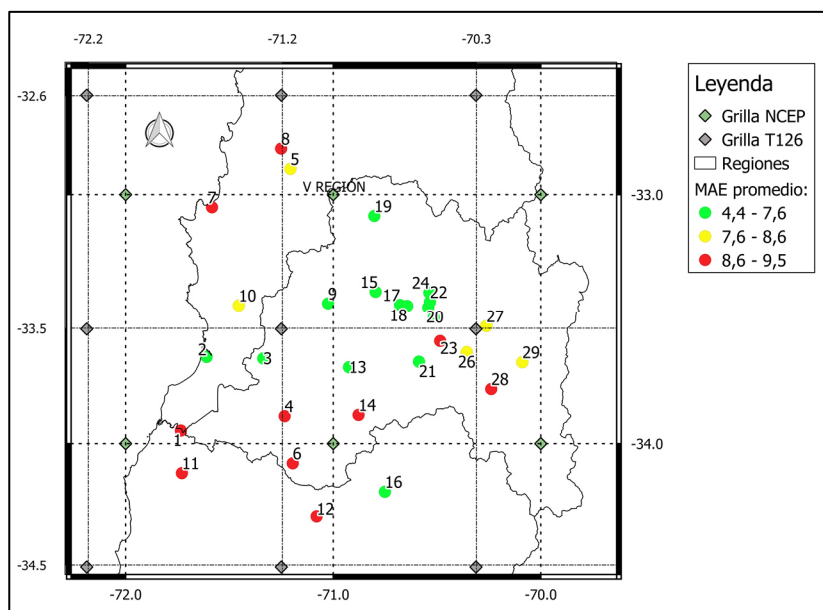


Figura 37: MAE promedio para cada estación, periodo de validación.
Fuente: Elaboración propia.

8. CONCLUSIONES

Para una mayor claridad, las conclusiones mantienen el orden en el que se desarrolló este estudio: análisis cluster entre estaciones meteorológicas, análisis de correlaciones, calibración de modelos de regresión lineal multivariada, verificación de supuestos, análisis de validación de los modelos y conclusiones generales. Señalar que lo concluido hace referencia únicamente a la zona de estudio.

La agrupación de las estaciones meteorológicas condujo a tres grupos (clusters) de estaciones: costa, valle y cordillera. Las 29 estaciones en la zona de estudio, en general, se agrupan claramente por su ubicación geográfica, por lo que es probable que entre estaciones cercanas tengan un comportamiento similar con respecto a la precipitación observada. El límite entre los clusters valle y cordillera fue delimitado por la latitud 70,5° oeste, mientras que la limitación espacial entre el cluster costa y el cluster valle es una línea diagonal, desde el noroeste al sureste.

El cluster cordillera ubicado entre la latitud 70,0° y 70,5° oeste, presentó un 60% de estaciones atípicas, lo atípico en sus series temporales de precipitación podría guardar relación con la mayor diferencia de altitudes entre las estaciones en este cluster. Siendo la estación más atípica, la estación de mayor altitud *El Yeso Embalse* (2483 msnm). Otra observación, fue lo atípico de la estación *Laguna Acúleo*, agrupándose con las estaciones del cluster valle, pero ubicada entre las estaciones del cluster costa, este comportamiento atípico puede deberse a la posibilidad de la generación de un microclima debido a la presencia de la laguna.

En el análisis de correlación espacial entre las variables atmosféricas (VAs) y la precipitación observada diaria (P_{obs}), se determinó que utilizar los puntos más cercanos de la grilla a la estación meteorológica solo aplica para la variable de precipitación PRATE. En el caso de la variable de humedad Q700, se presentan las mejores correlaciones en puntos de la grilla ubicados entre 1° a 2° de latitud hacia el norte, con una leve inclinación hacia el oeste. En el caso de las variables de presión (PRMSL, Z850 y Z500) solo se puede mencionar que los puntos de grilla más idóneos se encontraran hacia el sur de la estación. Las variables de temperatura en la zona de estudio (TMPSFC y T850), no entregan una tendencia con respecto al punto de mayor correlación. Con respecto a los signos de la correlación: la correlación es positiva (directamente proporcional) para las variables PRATE y Q700, es negativa (inversamente proporcional) para las variables de presión (PRMSL, Z850 y Z500), mientras que para las variables de temperatura no se identifica con claridad un signo predominante en la correlación.

El análisis de correlación mensual mostró una irregularidad en las correlaciones para los meses de abril y septiembre, pudiendo deberse a que estos dos meses presentan una menor cantidad de días con registro de P_{obs} mayor a cero. Por otra parte, en este análisis se reafirmó que las variables de temperatura (TMPSFC y T850) no logran tener un comportamiento claro del signo de su correlación, por lo que no son predictoras confiables y fueron descartadas para la construcción de los modelos. Este comportamiento irregular se hace más notorio en la variable T850, si se quisiera utilizar una variable de temperatura, TMPSFC sería una mejor opción que T850.

La variable que entrego mayor información para corregir el modelo de desagregación estadística, desde la variable de precipitación PRATE a la precipitación observada, es la variable Z500, puesto que permitió corregir la mayor cantidad de modelos, aportando en la construcción del 43% de ellos. Mientras que la variable Z850 en el 29% y la variable Q700 en el 22% de los casos. Si contáramos solo con las variables PRATE y Z500 se tendría el 56% de los casos con modelos de mejor correlación, este porcentaje mejora a un 78% si se incorpora la variable Z850, la tercera variable a considerar es la variable Q700.

Mediante el análisis de correlación entre variable, se concluye que las variables de presión Z850 y PRMSL nos entregan una información meteorológica similar. Las correlaciones entre sus series temporales en el rango muy alto (0,9 – 1,0), pero Z850 es más recomendable que la variable PRMSL en la zona de estudio, ya que presentó un mejor comportamiento, con una correlación promedio con la P_{obs} de -0,51, mientras que para la variable PRMSL fue de -0,44. Por esta razón, solo se consideró Z850 en la construcción de los modelos, subsanando así una posible colinealidad.

Por medio del análisis de linealidad de los modelos se reafirmaron los signos de las correlaciones entre la P_{obs} y las VAs. El supuesto de no colinealidad es válido para todos los modelos, y el supuesto de independencia de los residuos es válido para la mayoría de los modelos (168 de los 174 modelos).

El test de normalidad se ha cumplido solo en 8 de los 174 modelos, puesto que las variables precipitación observada y PRATE no tienen un comportamiento normal, la transformación de los datos de precipitación por medio de una función logarítmica no fue incluido en este trabajo, sin embargo, con ello se estaría más cerca de cumplir con este supuesto, en desmedro de valores más altos de correlación. Por lo que, en este estudio, se priorizó la correlación entre las variables y en los modelos.

Que los modelos no sean normales también afectó a los residuos, en el test de homocedasticidad 63 de los 174 modelos cumplen con el supuesto, esto implica que la mayoría de los modelos poseen una varianza de los residuos que no es constante, es decir, los valores de los residuos difieren para distintos valores de las variables independientes.

Aunque los modelos no cumplieron con el test de normalidad, los resultados de correlación fueron más altos en comparación a un estudio similar realizado por Vega (2018), en el que se aplicó la transformación logarítmica a los datos de precipitación. Otras diferencias con aquel estudio fueron (1) la selección del punto de grilla y (2) la incorporación de la variable Z500. Estas diferencias generaron un aumento de la correlación de los modelos de 0,46 en el caso de Vega (2018) a un 0,69 en este trabajo, según resultados correspondientes a los mismos meses de análisis.

Al comparar los resultados del MAE en el periodo de validación, los resultados en Vega (2018) están en un rango de 5 a 13 mm para las precipitaciones moderadas y 15 a 25 mm para precipitaciones extremas. En este estudio no se realizó esta separación y los resultados del MAE se obtuvieron en un rango de 1,8 a 14,4 mm (utilizando conjunto de datos completo que considera las precipitaciones moderadas y extremas). Se concluye que los rangos del MAE en el periodo de validación están asociados con el orden de asociación de los clusters; el cluster valle fue el primero en asociarse como cluster y es el que presenta menores valores del MAE, continuando, el segundo cluster en asociarse fue el cluster cordillera y es el siguiente con valores más altos en el MAE, y por último se asoció el cluster costa, obteniendo los valores más altos.

Finalmente, se ha logrado estudiar el pronóstico de precipitación en la zona centro de Chile, mediante la generación de 174 modelos de desagregación, utilizando variables atmosféricas, correlacionadas para su calibración, con un promedio de 0,68. Además, la desagregación desde la variable de precipitación PRATE de CFSv2 a una estación meteorológica que mide la precipitación observada, puede ser corregida mediante la incorporación de tres potenciales variables atmosféricas de acuerdo al siguiente orden: Z500, Z850 y Q700.

9. REFERENCIAS

- Anandhi, A. (2011). *Uncertainties in downscaled relative humidity for a semi-arid region in India*.
- Anuchaivong, P. (2017). Statistical downscaling for rainfall forecasts using modified constructed analog method in Thailand. *The Scientific World Journal*.
- Araya, D. (2020). *Climate change projections of temperature and precipitation in Chile*.

- Bardossy, A. (1992). WSpace-time model for daily rainfall using atmospheric circulation patterns. *Water resources research*, 1247-1259.
- BCN. (2020). *Biblioteca del Congreso Nacional de Chile*. Obtenido de <https://www.bcn.cl/siit/nuestropais/>
- Borén, R. (1995). *Método analógico de simulación de escenarios climáticos a escala comarcal*. Madrid.
- Castro, L. (2012). *Análisis Clúster como técnica de análisis exploratorio de registros múltiples en datos meteorológicos*. Cali, Colombia: Universidad del Valle.
- De la Fuente, S. (2011). Analisis de conglomerados. Universidad Autónoma de Madrid. Obtenido de <http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/SEGMENTACION/CONGLOMERADOS/conglomerados.pdf>
- Draper, N., & Smith, H. (1998). *Applied Regression Analysis*.
- Figueroa, R. (2008). *Efectos del cambio climático en la disponibilidad de recursos hídricos a nivel de cuenca - Implementación de un modelo integrado a nivel superficial y subterráneo*.
- Garreaud, R. (2011). *Cambio climático: bases físicas e impacto en Chile*.
- Hair, J. (1999). *Análisis Multivariante*.
- Hertig, E. (2019). *The impact of soil moisture on precipitation downscaling in the Euro Mediterranean area*.
- Houze, R. (1993). *Cloud Dynamics*.
- Huth, R. (1999). *Statistical downscaling in central Europe: evaluation of methods and potential predictors*.
- INE. (2017). *Resultados Censo 2017*. Obtenido de <http://resultados.censo2017.cl>
- IPPC. (1990). *Climate Change*.
- Kilsby, C. (1998). Predicting rainfall statistics in england and wales using atmospheric circulation variables. *International Journal of Climatology*.
- Kim, S. (2015). *ppcor: An R Package for a Fast Calculation to Semi-Partial Correlation Coefficients*.
- Liou, K. N. (2002). *An Introduction to Atmospheric Radiation*.
- MathWorks. (2019). Obtenido de <https://es.mathworks.com/help/matlab/ref/corrcoef.html>
- Menéndez, C. (2010). *CLARIS Project: towards climate downscaling in South*.
- National Centers for Environmental Information. (s.f.). *Climate Forecast System*. Obtenido de <https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/climate-forecast-system-version2-cfsv2>

- NCEP, N. C. (2011). *The NCEP climate forecast system versión 2*. Obtenido de <https://cfs.ncep.noaa.gov/cfsv2/docs.html>
- NOAA. (s.f.). *NOAA Celebrates 200 Years*. Obtenido de https://celebrating200years.noaa.gov/breakthroughs/climate_model/modeling_schematic.html
- NOAA, & NCEP. (2010). *CFS version 2 Information*. Obtenido de <https://cfs.ncep.noaa.gov/cfsv2.info/>
- Oliva, C. (2015). *Métodos para la segmentación de datos longitudinales. Aplicación a datos de rendimientos de cultivos en Argentina*. Buenos Aires: Tesis de Licenciatura en Ciencias Matemáticas.
- Pardo, A., & Ruiz, M. (2005). *Analisis de datos con SPSS 13 Base*. Madrid: Mc Graw Hill. Obtenido de <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/GuiaSPSS/18reglin.pdf>
- Pizarro, R. (2008). *Análisis del comportamiento y agresividad de las precipitaciones en la zona central de Chile*.
- RStudio_Team. (2020). *RStudio: Integrated Development for R*. Obtenido de RStudio, PBC, Boston: <http://www.rstudio.com/>
- Saha, S. (2014). *The NCEP Climate Forecast System Versión 2*.
- Sibson, R. (1981). A brief description of natural neighbor interpolation. En J. Wiley, & Sons, *Interpolating Multivariate Data* (págs. 21-36). New York.
- Souvignet, M. (2010). *Statistical downscaling of precipitation and temperature in north-central Chile: an assessment of possible climate change impacts in an arid Andean watershed*.
- Spak, S. (2007). A comparison of statistical and dynamical downscaling.
- Storch., V. (1993). Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime. *Journal of Climate*, 1161-1171.
- Trenberth. (2003). *The changing character of precipitation*.
- Trenberth. (2005). *Trends and variability in column-integrated atmospheric water vapor*.
- Trzaska, S., & Schnarr, E. (2014). *A review of downscaling methods for climate change projection*.
- Vega, H. (2018). *Pronostico de precipitaciones extremas usando un análisis multivariante*. Valparaíso.
- Wallace, J. M., & Hobbs, P. V. (2006). *Atmospheric Science*.
- Wilby, R. (2002). SDSM — a decision support tool for the assessment of regional. *Environmental Modelling & Software*.
- Wilby, R. (2004). *Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods*.

- Wilby, R. (2009). *A review of climate risk information for adaptation and development planning*.
- Wilby, R., & Wigley, T. (1997). Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography*.
- Wilks, & Wilby. (1999). *The weather generation game: a review of stochastic weather models*.
- Wu, W. (2012). Statistical downscaling of climate forecast system seasonal predictions for. *Atmospheric Research*.
- Zorita, E., & von Storch, H. (1999). *The Analog Method as a Simple Statistical Downscaling Technique*.

ÍNDICE DE FIGURAS

Figura 1: Modelo Numérico de Predicción, Fuente: (NOAA, NOAA Celebrates 200 Years, s.f.)	10
Figura 2: Tipos de pronósticos CFSv2 Reforecast. Fuente: (NOAA & NCEP, CFS version 2 Information, 2010)	11
Figura 3: Tipos de pronósticos CFSv2 Operational. Fuente: (NOAA & NCEP, CFS version 2 Information, 2010)	12
Figura 4: Grilla de un Modelo Numérico de Predicción típico. Fuente: (IPPC, 1990)	12
Figura 5: Conceptualización de agregación y desagregación entre modelos hidrológicos y atmosféricos. Fuente: (Trzaska & Schnarr, 2014)	13
Figura 6 : Ejemplo de un campo de predicción y áreas de estudio. El dominio del modelo regional se encuentra en gris. Fuente: (Spak, 2007)	14
Figura 7: Configuración simple de una nube que se mueve sobre una superficie. Fuente: (Liou, 2002)	26
Figura 8: Campo vectorial de fuerza de gradiente de presión de flotabilidad para una parcela flotante en el plano x-z. Fuente: (Houze, 1993).....	26
Figura 9: Estaciones pluviométricas. Fuente: Elaboración propia	35
Figura 10: Grillas del modelo CFSv2, NCEP y Gaussiana T126. Fuente: Elaboración propia.....	37
Figura 11: Esquema de la metodología de trabajo. Fuente: elaboración propia.....	38
Figura 12: Dendograma método Single. Fuente: Elaboración propia	45
Figura 13: Dendograma método Ward Fuente: Elaboración propia.....	46
Figura 14: Representación espacial de las estaciones meteorológicas agrupadas en 3 cluster. Fuente: Elaboración propia	46
Figura 15: Correlación espacial entre la precipitación observada en las estaciones 4, 17 y 26; y las variables PRATE y Q700. Fuente: Elaboración propia	48
Figura 16: Correlación espacial entre precipitación observada en las estaciones 4, 17 y 26; y las variables de temperatura T850 y TMP5FC. Fuente: Elaboración propia.	49
Figura 17: Correlación espacial entre precipitación observada en las estaciones 4, 17 y 26; y las variables de altura geopotencial Z500 y Z850. Fuente: Elaboración propia.....	50

Figura 18: Correlación espacial entre precipitación observada en las estaciones 4, 17 y 26; y las variables de presión PRMSL. Fuente: Elaboración propia.	51
Figura 19: Correlación mensual entre la precipitación observada y la variable PRATE para las 29 estaciones. Fuente: Elaboración propia.	52
Figura 20: Gráfica de los valores de PRATE y Pobs para los meses de abril y agosto en la estación 13. Fuente: Elaboración propia.	52
Figura 21: Correlación mensual entre la precipitación observada y la variable Q700 para las 29 estaciones. Fuente: Elaboración propia.	53
Figura 22: Correlación mensual entre la precipitación observada y la variable Z500 para las 29 estaciones. Fuente: Elaboración propia.	53
Figura 23: Correlación mensual entre la precipitación observada y la variable Z850 para las 29 estaciones. Fuente: Elaboración propia.	54
Figura 24: Correlación mensual entre la precipitación observada y la variable PRMSL para las 29 estaciones. Fuente: Elaboración propia.	54
Figura 25: Correlación mensual entre la precipitación observada y la variable T850 para las 29 estaciones. Fuente: Elaboración propia.	55
Figura 26: Correlación mensual entre la precipitación observada y la variable TMP5FC para las 29 estaciones. Fuente: Elaboración propia.	55
Figura 27: Boxplot correlaciones entre variables. Fuente: Elaboración propia.	56
Figura 28: Residuos del modelo de regresión en la estación 17 (Quinta Normal), mes de junio. Fuente: Elaboración propia.	62
Figura 29: Residuos del modelo de regresión en la estación 7 (Rodelillo), mes de julio. Fuente: Elaboración propia.	62
Figura 30 : Gráfica y correlación inter variable en estación 17. Fuente: Elaboración propia	64
Figura 31: Gráfica Q-Q de la estación 17 (Quinta Normal) en el mes de junio. Fuente: Elaboración propia.	64
Figura 32: Gráfica Q-Q de la estación 7 (Rodelillo) en el mes de julio. Fuente: Elaboración propia.	65
Figura 33: Correlación de los modelos periodo reforecast (calibración). Fuente: Elaboración propia	68
Figura 34: Correlación de los modelos periodo operational (validación). Fuente: Elaboración propia.	68
Figura 35: MAE de los modelos periodo reforecast (calibración). Fuente: Elaboración propia.	70
Figura 36: MAE de los modelos periodo operational (validación). Fuente: Elaboración propia.	70
Figura 37: MAE promedio para cada estación, periodo de validación. Fuente: Elaboración propia.	77
Figura 38 Grafica del comportamiento del modelo de regresión para la estación 4 – Julio. Fuente: Elaboración propia	92
Figura 39 Grafica del comportamiento del modelo de regresión para la estación 17 – Julio. Fuente: Elaboración propia	93
Figura 40 Grafica del comportamiento del modelo de regresión para la estación 26 – Julio. Fuente: Elaboración propia	94
Figura 41: Gráfica y correlación entre variables en estación 4. Fuente: Elaboración propia	100
Figura 42: Gráfica y correlación entre variables en estación 19. Fuente: Elaboración propia	101
Figura 43: Gráfica y correlación entre variables en estación 26. Fuente: Elaboración propia	102

ÍNDICE DE TABLAS

Tabla 1: Desagregación espacial estadística: categoría, método, variables predictoras y de predicción, ventajas y desventajas. Fuente: (Trzaska & Schnarr, 2014)	16
Tabla 2: Fortalezas y debilidades de desagregación estadística y dinámica. Fuente: (Wilby R. , 2002) & (Trzaska & Schnarr, 2014).....	17
Tabla 3: Variables y correlación parcial de una desagregación estadística en Canadá. (Wilby R. , 2002).	27
Tabla 4: Correlación parcial entre variables predictoras y la precipitación en una desagregación estadística en Tailandia (Anuchaivong, 2017).....	28
Tabla 5: Niveles de correlación. Fuente: (Anuchaivong, 2017).....	28
Tabla 6: Variables y correlación parcial en una desagregación estadística en la VI Región, Chile (Figueroa, 2008)	29
Tabla 7: Variables y correlación parcial en una desagregación estadística en la IV Región, Chile (Souvignet, 2010).	30
Tabla 8: Estaciones pluviométricas en la zona de estudio. Fuente: Elaboración propia	34
Tabla 9: Variables modelo CFSv2. Fuente: (NCEP, 2011)	36
Tabla 10: Variables atmosféricas de CFSv2 a utilizar. Fuente: (NCEP, 2011)	39
Tabla 12 Constantes de cada modelo de regresión obtenido para la serie temporal fase Reforecast (Calibración), para los meses de abril y mayo, separados por estación meteorológica. Fuente: Elaboración propia.	57
Tabla 13 Constantes de cada modelo de regresión obtenido para la serie temporal reforecast (Calibración), para los meses de junio y Julio, separados por estación meteorológica. Fuente: Elaboración propia	58
Tabla 14 Constantes de cada modelo de regresión obtenido para la serie temporal del periodo Reforecast (Calibración), para los meses de agosto y septiembre, separados por estación meteorológica. Fuente: Elaboración propia.....	59
Tabla 15 Agrupación de la cantidad de modelos de regresión de precipitación en función de las variables seleccionadas por el método hacia adelante.	59
Tabla 16 Herramientas y criterios de verificación del cumplimiento de los supuestos de linealidad.	60
Tabla 17: Cumplimiento del test de independencia Durbin-Watson.....	61
Tabla 18: cumplimiento de la homocedasticidad de los residuos del modelo lineal para cada estación y mes. Fuente: elaboración propia.....	63
Tabla 19: Test de normalidad. Fuente: elaboración propia	65
Tabla 20: Correlación entre <i>Ppron</i> y Pobs para el periodo Operational. Los colores corresponden a los niveles de correlación de la Tabla 5: muy baja (rojo), baja (café), media (verde), alta (azul) y muy alta (morado). Fuente: Elaboración propia.....	67
Tabla 21: MAE entre los modelos y Pobs para el periodo Operational (validación). Colores de acuerdo con el resultado: rojo (15-20 mm), verde (10-15 mm), azul (5-10 mm) y morado (0-5 mm). Fuente: Elaboración propia.	69
Tabla 22: Puntos de grilla de máxima correlación para cada estación por cada variable	90
Tabla 23: Test Durbin-Watson	91
Tabla 24: p-valores para el test de significancia Breush-Pagan de la homocedasticidad de los residuos del modelo lineal para cada estación y mes. Fuente: elaboración propia.....	95

Tabla 25: Cumplimiento de supuesto de normalidad de las variables para cada estación y mes. Fuente: elaboración propia.....	96
Tabla 26: p-valores para el test de significancia Shapiro-Wilk de la normalidad de las variables para cada estación y mes. Fuente: elaboración propia.....	97
Tabla 27 P-valores para el test de Shapiro Wilk para verificar la normalidad de los modelos lineales. Fuente: elaboración propia.....	98
Tabla 28: Test de colinealidad valores de FIV. Fuente: elaboración propia.....	99

ANEXOS

A. REGRESIÓN LINEAL

1. REGRESIÓN LINEAL SIMPLE

Una regresión lineal simple para un caso determinista es descrita por $y = \beta_0 + \beta_1 x$, donde el valor actual valor observado y es una función de x con los parámetros β_0 y β_1 . Esto se puede generalizar como un modelo probabilístico para la variable aleatoria Y de la siguiente forma:

$$Y = \beta_0 + \beta_1 X + \epsilon \quad (13)$$

La salida del modelo incluye un error aleatorio (ϵ) normalmente distribuido ($E(\epsilon) = 0$; $Var(\epsilon) = \sigma_\epsilon^2$). X se llama predictor e Y es el predictando. Para una muestra de tamaño n con datos observados de la forma (x_i, y_i) para un $i = 1, \dots, n$ la suma de los errores cuadráticos (SSE) se define como:

$$SSE = \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x)]^2 = \sum_{i=1}^n \epsilon_i^2 \quad (14)$$

Los parámetros β_0, β_1 se estiman buscando el menor error cuadrático. Los estimadores se denotan por $\hat{\beta}_0$ y $\hat{\beta}_1$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (15)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (16)$$

La varianza del estimado de $\hat{\beta}_1$ es

$$s_{\hat{\beta}_1}^2 = \frac{s_\epsilon^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (17)$$

$$s_\epsilon^2 = \frac{\sum_{i=1}^n \epsilon_i^2}{n - 2} \quad (18)$$

Notar que el divisor $n - 2$ es número de grados de libertad asociados a la estimación y dos parámetros.

2. REGRESIÓN LINEAL MULTIVARIADA

La regresión lineal multivariada para múltiples variables de $x_i^1, x_i^2, \dots, x_i^S$ es descrita por:

$$Y_i = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \dots + \beta_S x_i^S + \epsilon_i \quad (19)$$

donde S es el número de predictores de interés. De forma matricial, se expresa de la siguiente manera:

$$Y_i = \vec{x}_i^T \boldsymbol{\beta} + \epsilon_i \quad (20)$$

donde $\vec{x}_i = [1, x_i^1, x_i^2, \dots, x_i^S]^T$ y $\boldsymbol{\beta} = [\beta_0, \beta_1, \dots, \beta_S]^T$.

La suma de los cuadrados de los residuos y su estimación del mínimo cuadrado de $\boldsymbol{\beta}$ vienen dadas por:

$$SSE = \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \vec{\mathbf{x}} \boldsymbol{\beta} - \boldsymbol{\beta}^T \vec{\mathbf{x}}^T \mathbf{y} + \boldsymbol{\beta}^T \vec{\mathbf{x}}^T \vec{\mathbf{x}} \boldsymbol{\beta} \quad (21)$$

donde $\vec{\mathbf{x}} = [1, \dots, 1; x_1^1, \dots, x_n^1; x_1^2, \dots, x_n^2; \dots; x_1^S, \dots, x_n^S]$ (matriz de dimensiones $n \times S + 1$), e $\mathbf{y} = [y_1, \dots, y_n]^T$ con n numero de datos.

El estimador de mínimos cuadrados se obtiene como:

$$\hat{\boldsymbol{\beta}} = (\vec{\mathbf{x}}^T \vec{\mathbf{x}})^{-1} \vec{\mathbf{x}}^T \mathbf{y} \quad (22)$$

Además, la variación del error $Var(\epsilon) = \sigma_\epsilon^2$ se estima con

$$\sigma_\epsilon^2 = \frac{SSE}{n - (k + 1)} \quad (23)$$

B. RUTINAS EN SOFTWARE

1. *CLUSTER.M*

- Lee las series de precipitación observada de las 29 estaciones en el periodo 1995-2017.
- Filtra los datos descartando los días en los que no existe precipitación para ninguna de las estaciones.

- Se aplica el análisis cluster con el método *Single* para obtener estaciones atípicas, obteniendo un dendograma.
- Se aplica el análisis cluster con el método Ward para generar los grupos (clusters), obteniendo un dendograma.

2. PROCESO DE DATOS DE LAS VARIABLES ATMOSFÉRICAS

Procesamiento de datos CFSv2. Fuente: Departamento de ingeniería hidráulica y ambiental de la Pontificia Universidad Católica de Chile.

FASE REFORECAST 1995-2010	FASE OPERATIONAL 2011-2017
<i>URL_lista_retro.m</i>	<i>URL_lista_Operational.m</i>
• Genera la lista con las URL de descarga para cada una de las variables de CFSv2. • La lista se almacena en un archivo de texto (*.txt) con todas las URL necesarias para ser utilizada por el siguiente código.	
<i>downgrd2_transnc_cfsv2reforecast.m</i>	<i>downgrd2_transnc_cfsv2operational.m</i>
• Utiliza la lista de las URL para realizar la descarga de archivos por medio de la aplicación WGET. • Transforma los archivos del formato GRIB 2 (*.grb2) al formato NETCDF (*.nc). • Extrae los datos de la zona de estudio desde la grilla mundial. • Los archivos de salida en formato “.nc” son guardados en carpetas por mes y año.	
<i>CFSv2Reforecast_variable_diaria.m</i>	<i>CFSv2Operational_variable_diaria.m</i>
• Lee los archivos NETCDF archivados por mes y año. • Procesa los pronósticos de las variables y los agrega a nivel diario; el valor acumulado diario en el caso PRATE, el valor mínimo diario para Q700 y el valor promedio diario para las otras 5 variables (Tabla 10). • Los pronósticos son guardados en archivos *.mat; según su respectivo mes y año.	
<i>CFS_joinmat_diaria.m</i>	<i>CFSv2Operational.m</i>
• Lee los archivos organizados por mes y año. • Acopla los pronósticos de forma mensual entre los años del periodo 1995-2010. • Almacena los datos en una carpeta por cada variable en un archivo *.mat	• Lee los archivos organizados por mes y año. • Procesa los pronósticos de las variables llevándolos a un nivel diario entre los años del periodo 2011-2017 (<i>read_cfsv2_2011_2017.m</i>). • Almacena los datos en una carpeta por cada variable en un archivo *.mat

3. *CORRELACION_X_E.M*

3.1. CORRELACIÓN ESPACIAL

- Lee los pronósticos de cada variable desde los archivos *.mat para ambos periodos (reforecast y operational)
- Lee los datos de precipitación observada.
- Filtra los datos para los días con precipitación, se descartan los datos en los días que la precipitación es cero.
- Grafica las correlaciones espaciales de la serie completa (datos de abril a septiembre).
- Determina los puntos de grilla de mejor correlación y los guarda en un archivo *.mat.
- Guarda las series de estos puntos para cada estación y cada mes en cada periodo (reforecast y operational) en archivos *.mat.
- Grafica los resultados de forma espacial entre P_{obs} y cada VA, para cada estación meteorológica.

3.2. CORRELACIÓN MENSUAL

- Escoge las series temporal para cada variable de acuerdo al punto de grilla de mejor correlación.
- Separa cada serie temporal para cada mes de análisis (abril a septiembre).
- Calcula el r entre cada una de las VA y la P_{obs} en cada estación para cada mes, para datos con P_{obs} mayor a cero dentro del periodo completo (1995 – 2017).
- Exporta los resultados de correlación a un archivo Excel.

3.3. CORRELACIÓN ENTRE VARIABLES

- Calcula la correlación entre las VAs por cada estación y mes de análisis. Utiliza los datos del periodo completo (1995-2017) en los días que $P_{obs} > 0$.
- Grafica las correlaciones entre variables en un boxplot. Donde cada cruce de variables considera las 29 estaciones.

4. *MODELOS_REGRESION_MENSUAL.R*

- Genera los modelos de regresión (aplicando el método *stepwise*) utilizando los puntos de mejor correlación de cada variable atmosférica y graficas de apoyo para la verificación de los supuestos del modelo. Considerando las series temporales por cada mes del periodo *Reforecast* (1995 – 2010).
- Calcula el r y el MAE entre P_{pron} y P_{obs} , para cada uno de los modelos generados por estación y por mes.

5. *TEST_SUPUESTOS.R*

- Calcula para cada uno de las variables y modelos de regresión obtenidos los test: DW, Shaprio Wilk, Breusch-Pagan y FIV.

- Genera gráficos por estación de cada uno de los modelos lineales, para cada mes. Estos gráficos de residuos, desviación estándar de residuos, Q-Q. Complementado el análisis de cumplimiento de supuestos.

6. *VALIDACION_CORRELACION.M*

- Lee las constantes de los modelos de regresión obtenidos para cada mes y estación periodo *Reforecast* (1995 – 2010)
- Se evalúan los modelos utilizando las series temporales de periodo *Operational* (2011-2017).
- Calcula el coeficiente de Pearson de cada modelo respecto a P_{obs} y los MAE para cada uno de los modelos por estación y por mes.
- Gráfica en boxplot con los resultados de r y MAE, por periodo y mes de análisis.

C. PUNTOS DE GRILLA DE MÁXIMA CORRELACIÓN

Tabla 21: Puntos de grilla de máxima correlación para cada estación por cada variable

Estación Pluviométrica	P _{RATE}		Q700		Z500		Z850	
	Lat	Lon	Lat	Lon	Lat	Lon	Lat	Lon
1	-33,5	-72,2	-32	-72	-36	-75	-36	-71
2	-33,5	-72,2	-33	-71	-36	-75	-35	-70
3	-33,5	-71,2	-32	-71	-36	-75	-35	-70
4	-33,5	-72,2	-32	-72	-36	-75	-36	-70
5	-32,6	-72,2	-32	-72	-36	-75	-35	-70
6	-33,5	-71,2	-33	-71	-36	-75	-35	-70
7	-33,5	-72,2	-32	-72	-36	-75	-36	-71
8	-32,6	-72,2	-32	-72	-36	-75	-35	-70
9	-33,5	-71,2	-31	-72	-36	-75	-35	-70

10	-32,6	-72,2	-31	-72	-36	-75	-35	-70
11	-34,5	-72,2	-33	-72	-36	-75	-36	-71
12	-34,5	-71,2	-32	-72	-36	-75	-36	-70
13	-33,5	-71,2	-31	-72	-36	-75	-35	-70
14	-33,5	-71,2	-32	-72	-36	-75	-36	-70
15	-32,6	-71,2	-32	-71	-36	-75	-34	-70
16	-33,5	-71,2	-32	-71	-36	-75	-35	-70
17	-32,6	-71,2	-32	-71	-36	-75	-33	-69
18	-32,6	-71,2	-31	-72	-36	-75	-34	-70
19	-32,6	-71,2	-32	-71	-36	-75	-33	-69
20	-33,5	-70,3	-31	-71	-36	-75	-33	-69
21	-33,5	-71,2	-32	-71	-36	-75	-34	-69
22	-32,6	-71,2	-31	-71	-36	-75	-34	-70
23	-33,5	-71,2	-32	-71	-36	-75	-34	-69
24	-32,6	-71,2	-31	-71	-36	-75	-33	-69
25	-32,6	-71,2	-31	-72	-36	-75	-33	-69
26	-33,5	-70,3	-32	-72	-36	-75	-35	-70
27	-33,5	-70,3	-32	-72	-36	-70	-36	-70
28	-34,5	-70,3	-32	-72	-36	-70	-35	-69
29	-34,5	-70,3	-32	-72	-36	-70	-36	-70

D. VERIFICACIÓN DE SUPUESTOS DE MODELOS LINEALES

1. INDEPENDENCIA

Tabla 22: Test Durbin-Watson

Estación pluviométrica		Abril	Mayo	Junio	Julio	Agosto	Sept.
1	RAPEL	1,70	1,67	2,09	1,85	2,14	2,03
2	SANTO DOMINGO AD.	2,02	2,56	2,22	1,98	1,74	2,13
3	ESTERO PUANGUE RUTA 78	1,56	1,56	2,20	2,02	1,92	2,04
4	LOS GUINDOS	1,75	2,30	2,00	1,88	1,84	2,28
5	QUILLOTA	2,00	1,69	1,44	2,16	2,08	1,86
6	BARRERA LONCHA	2,31	1,98	2,15	2,32	1,91	2,05
7	RODELILLO	1,91	2,29	2,17	1,90	1,83	2,33
8	LO ROJAS	1,68	1,83	1,62	1,92	2,35	1,49
9	LOS PANGUILES	1,86	2,24	1,89	2,15	1,73	1,98
10	LAGUNILLAS	1,36	2,27	1,88	2,24	1,62	2,13
11	LITUECHE	1,68	1,76	1,95	1,82	1,83	2,40
12	COLTAUCO	2,32	2,40	1,78	2,46	1,63	2,27

13	EL VERGEL	2,31	2,26	1,94	2,07	1,93	1,97
14	LAGUNA ACULEO	2,03	2,03	1,75	2,21	2,18	1,91
15	PUDAHUEL SANTIAGO	2,36	2,12	1,81	2,25	1,94	2,07
16	RANCAGUA (CACHAPOAL-DCP)	1,85	2,36	2,17	2,09	1,74	1,85
17	QUINTA NORMAL SANTIAGO	2,36	1,78	1,79	2,17	1,91	2,22
18	TERRAZAS OFICINA CENTRAL DGA	1,95	2,06	1,62	1,94	2,07	1,53
19	HUECHUN EMBALSE	1,86	1,72	1,89	2,01	1,62	2,18
20	EULOGIO SANCHEZ AD	2,32	2,02	1,95	2,08	1,56	2,31
21	PIRQUE	1,92	2,06	2,07	2,01	1,78	1,85
22	QUEBRADA RAMON	1,92	1,31	2,12	1,68	1,61	1,96
23	LA OBRA RECINTO REMOS	1,47	1,86	1,75	2,23	1,61	1,72
24	CERRO CALAN	1,74	1,96	1,86	2,06	1,60	2,03
25	ANTUPIREN	2,30	1,68	1,85	2,14	1,66	1,70
26	SAN JOSE DE MAIPO RETEN	1,93	2,05	1,84	1,98	1,72	1,56
27	MAITENES BOCATOMA	1,62	1,86	2,04	1,77	1,65	1,94
28	SAN GABRIEL	1,94	1,99	2,04	1,95	1,72	1,92
29	EL YESO EMBALSE	1,76	1,86	1,51	1,84	1,95	1,82

2. HOMOCEDASTICIDAD

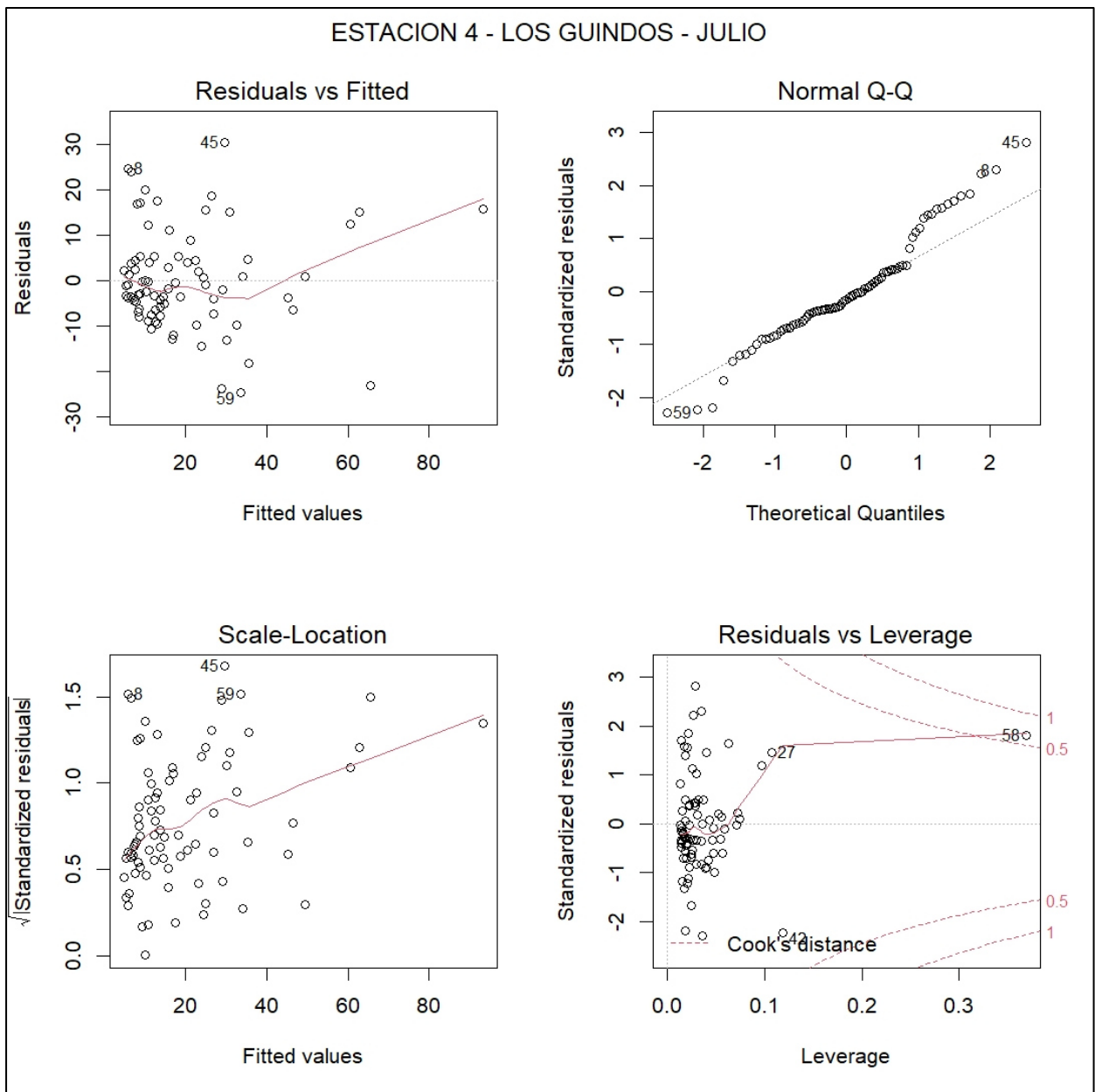


Figura 38 Grafica del comportamiento del modelo de regresión para la estación 4 – Julio.
Fuente: Elaboración propia

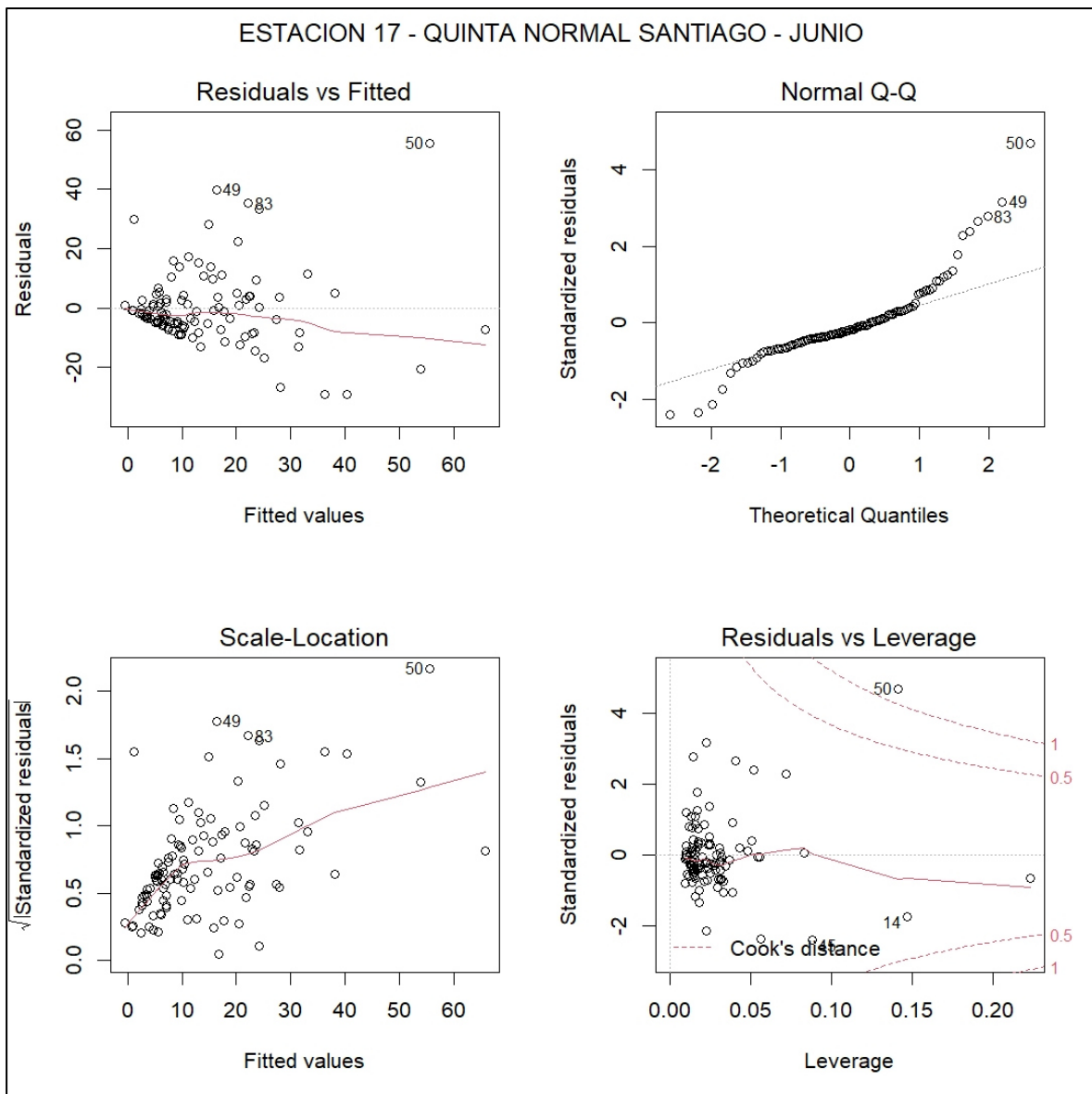


Figura 39 Grafica del comportamiento del modelo de regresión para la estación 17 – Julio.
Fuente: Elaboración propia

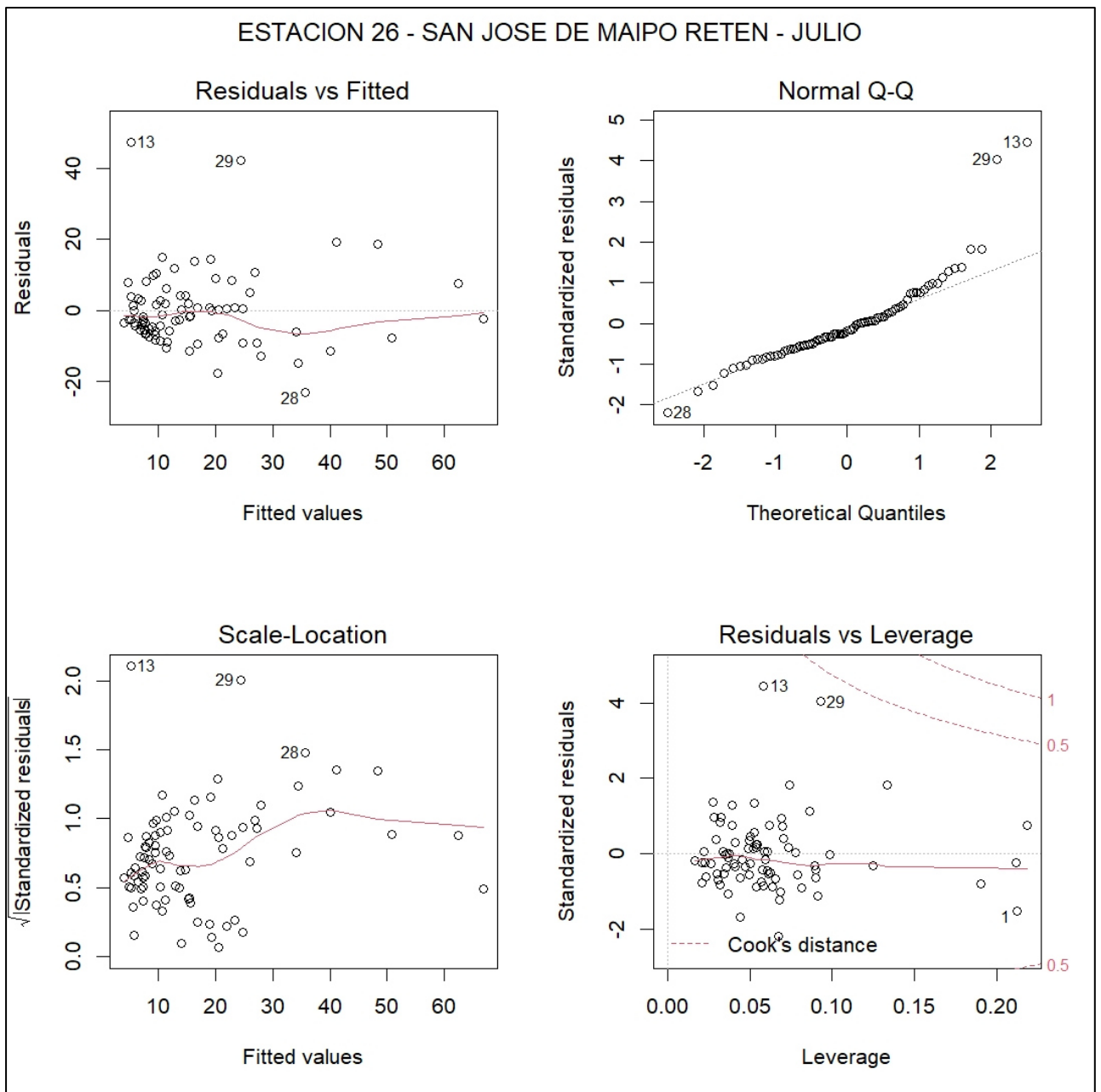


Figura 40 Grafica del comportamiento del modelo de regresión para la estación 26 – Julio.

Fuente: Elaboración propia

Tabla 23: p-valores para el test de significancia Breush-Pagan de la homocedasticidad de los residuos del modelo lineal para cada estación y mes. Fuente: elaboración propia.

Estación pluviométrica		Abril	Mayo	Junio	Julio	Agosto	Sept.
1	RAPEL	0,034	0,045	0,000	0,000	0,000	0,000
2	SANTO DOMINGO AD.	0,044	0,000	0,092	0,000	0,000	0,000
3	ESTERO PUANGUE RUTA 78	0,609	0,072	0,000	0,001	0,064	0,000
4	LOS GUINDOS	0,139	0,217	0,000	0,058	0,095	0,171
5	QUILLOTA	0,060	0,253	0,009	0,065	0,041	0,092
6	BARRERA LONCHA	0,716	0,000	0,000	0,000	0,001	0,000
7	RODELILLO	0,476	0,186	0,000	0,085	0,421	0,040
8	LO ROJAS	0,066	0,328	0,000	0,004	0,043	0,000
9	LOS PANGUILES	0,000	0,000	0,000	0,070	0,742	0,000
10	LAGUNILLAS	0,013	0,028	0,000	0,003	0,351	0,000
11	LITUECHE	0,001	0,012	0,000	0,000	0,000	0,000
12	COLTAUCO	0,935	0,000	0,000	0,000	0,000	0,051
13	EL VERGEL	0,025	0,001	0,003	0,000	0,003	0,000
14	LAGUNA ACULEO	0,267	0,023	0,001	0,000	0,200	0,894
15	PUDAHUEL SANTIAGO	0,014	0,000	0,000	0,012	0,000	0,022
16	RANCAGUA (CACHAPOAL-DCP)	0,141	0,000	0,000	0,000	0,001	0,000
17	QUINTA NORMAL SANTIAGO	0,014	0,000	0,000	0,020	0,001	0,001
18	TERRAZAS OFICINA CENTRAL DGA	0,404	0,005	0,000	0,267	0,000	0,000
19	HUECHUN EMBALSE	0,012	0,000	0,000	0,212	0,026	0,018
20	EULOGIO SANCHEZ AD	0,001	0,000	0,000	0,386	0,480	0,225
21	PIRQUE	0,005	0,000	0,004	0,135	0,024	0,358
22	QUEBRADA RAMON	0,155	0,051	0,011	0,874	0,747	0,846
23	LA OBRA RECINTO REMOS	0,006	0,000	0,001	0,605	0,281	0,183
24	CERRO CALAN	0,400	0,000	0,000	0,180	0,448	0,310
25	ANTUPIREN	0,108	0,000	0,000	0,458	0,035	0,344
26	SAN JOSE DE MAIPO RETEN	0,734	0,007	0,032	0,521	0,323	0,573
27	MAITENES BOCATOMA	0,469	0,002	0,000	0,000	0,036	0,177
28	SAN GABRIEL	0,067	0,015	0,000	0,323	0,082	0,133
29	EL YESO EMBALSE	0,004	0,000	0,000	0,000	0,000	0,410

3. NORMALIDAD

NORMALIDAD POR VARIABLE

Tabla 24: Cumplimiento de supuesto de normalidad de las variables para cada estación y mes. Fuente: elaboración propia.

		Abril					Mayo					Junio					Julio					Agosto					Septiembre				
Estación pluviométrica		Pobs	PRATE	Q700	Z850	Z500	Pobs	PRATE	Q700	Z850	Z500	Pobs	PRATE	Q700	Z850	Z500	Pobs	PRATE	Q700	Z850	Z500	Pobs	PRATE	Q700	Z850	Z500	Pobs	PRATE	Q700	Z850	Z500
1	RAPEL	No	No	No	Si	Si	No	No	No	No	Si	No	No	No	No	Si	No	No	No	Si	No	No	No	No	Si	No	No	No	No	No	Si
2	SANTO DOMINGO AD.	No	No	Si	Si	Si	No	No	No	Si	Si	No	No	No	No	Si	No	No	Si	Si	Si	No	No	No	Si	No	No	No	Si	No	Si
3	ESTERO PUANGUE RUTA 78	No	No	Si	Si	Si	No	No	No	Si	Si	No	No	No	Si	No	No	No	Si	No	No	No	No	No	Si	No	No	No	Si	Si	Si
4	LOS GUINDOS	No	No	Si	Si	Si	No	No	No	Si	Si	No	No	No	No	Si	No	No	No	Si	Si	No	No	No	Si	Si	No	No	Si	Si	Si
5	QUILLOTA	No	No	Si	Si	Si	No	No	No	Si	Si	No	No	No	Si	Si	No	No	No	Si	Si	No	No	No	Si	No	No	No	Si	No	Si
6	BARRERA LONCHA	No	No	Si	Si	Si	No	No	Si	Si	Si	No	No	Si	No	Si	No	No	Si	Si	No	No	No	Si	Si	No	No	Si	No	Si	Si
7	RODEILLO	No	No	No	Si	Si	No	No	Si	Si	Si	No	No	No	Si	Si	No	No	No	Si	No	No	No	Si	Si	No	No	No	Si	No	Si
8	LO ROJAS	No	No	Si	Si	Si	No	No	Si	Si	Si	No	No	No	Si	Si	No	No	No	Si	No	No	No	Si	Si	No	No	Si	No	Si	Si
9	LOS PANGUILES	No	No	No	Si	Si	No	No	No	Si	Si	No	No	No	Si	No	No	No	No	Si	No	No	No	Si	Si	No	No	No	No	No	Si
10	LAGUNILLAS	No	No	Si	Si	Si	No	No	No	Si	Si	No	No	No	No	Si	No	No	No	Si	Si	No	No	No	Si	Si	No	No	No	No	Si
11	LITUECHE	No	No	No	Si	Si	No	No	No	Si	Si	No	No	No	No	Si	No	No	No	Si	Si	No	No	No	Si	Si	No	No	No	Si	Si
12	COLTAUCO	No	No	Si	Si	Si	No	No	No	Si	Si	No	No	No	No	Si	No	No	No	Si	No	No	No	Si	Si	No	No	Si	Si	Si	Si
13	EL VERGEL	No	No	No	Si	Si	No	No	Si	Si	Si	No	No	No	Si	Si	No	No	No	Si	No	No	No	Si	Si	No	No	No	No	No	Si
14	LAGUNA ACULEO	No	No	Si	No	Si	No	No	No	Si	Si	No	No	No	No	Si	No	No	No	Si	No	No	No	Si	Si	No	No	Si	No	Si	Si
15	PUDAHUEL SANTIAGO	No	No	Si	Si	Si	No	No	Si	Si	Si	No	No	Si	Si	Si	No	No	No	Si	Si	No	No	No	Si	Si	No	No	Si	No	Si
16	RANCAGUA (CACHAPOAL-DCP)	No	No	Si	Si	Si	No	No	No	Si	Si	No	No	No	No	Si	No	No	No	Si	No	No	No	Si	No	No	No	No	No	No	Si
17	QUINTA NORMAL SANTIAGO	No	No	Si	Si	Si	No	No	Si	Si	Si	No	No	Si	Si	Si	No	No	No	Si	Si	No	No	No	Si	Si	No	No	Si	Si	Si
18	TERRAZAS OFICINA CENTRAL DGA	No	No	Si	Si	Si	No	No	Si	Si	Si	No	No	No	Si	Si	No	No	No	No	Si	No	No	No	Si	Si	No	No	Si	No	Si
19	HUECHUN EMBALSE	No	No	Si	Si	Si	No	No	Si	Si	Si	No	No	No	Si	Si	No	No	No	Si	No	No	No	Si	Si	No	No	Si	Si	Si	Si
20	EULOGIO SANCHEZ AD	No	No	Si	Si	No	No	No	Si	Si	Si	No	No	Si	Si	Si	No	No	No	Si	Si	No	No	No	Si	Si	No	No	No	Si	Si
21	PIRQUE	No	No	Si	Si	Si	No	No	Si	Si	Si	No	No	Si	Si	Si	No	No	No	Si	No	No	No	Si	Si	No	No	Si	Si	Si	Si
22	QUEBRADA RAMON	No	No	Si	Si	Si	No	No	No	Si	Si	No	No	Si	Si	Si	No	No	No	Si	No	No	No	Si	Si	No	No	No	No	No	Si
23	LA OBRA RECINTO REMOS	No	No	Si	Si	Si	No	No	Si	Si	Si	No	No	Si	Si	Si	No	No	No	Si	No	No	No	Si	Si	No	No	No	No	Si	Si
24	CERRO CALAN	No	No	No	Si	Si	No	No	No	Si	Si	No	No	Si	Si	Si	No	No	No	Si	No	No	No	Si	Si	No	No	No	No	No	Si
25	ANTUPIREN	No	No	No	Si	Si	No	No	No	Si	Si	No	No	No	Si	Si	No	No	No	Si	No	No	No	Si	Si	No	No	No	No	Si	Si
26	SAN JOSE DE MAIPO RETEN	No	No	No	Si	Si	No	No	No	Si	Si	No	No	No	No	Si	No	No	Si	Si	Si	No	No	No	Si	Si	No	No	No	No	Si
27	MAITENES BOCATOMA	No	No	No	No	Si	No	No	No	Si	Si	No	No	No	No	Si	No	No	No	Si	Si	No	No	No	Si	Si	No	No	No	No	Si
28	SAN GABRIEL	No	No	No	Si	Si	No	No	No	Si	Si	No	No	No	No	Si	No	No	No	Si	Si	No	No	No	Si	Si	No	No	No	No	Si
29	EL YESO EMBALSE	No	No	No	Si	Si	No	No	No	Si	Si	No	No	No	No	Si	No	No	No	Si	Si	No	No	No	No	Si	No	No	Si	No	Si

Tabla 25: p-valores para el test de significancia Shapiro-Wilk de la normalidad de las variables para cada estación y mes. Fuente: elaboración propia.

Estación pluviométrica		Abril					Mayo					Junio					Julio					Agosto					Septiembre				
		Pobs	PRATE	Q700	Z850	Z500	Pobs	PRATE	Q700	Z850	Z500	Pobs	PRATE	Q700	Z850	Z500	Pobs	PRATE	Q700	Z850	Z500	Pobs	PRATE	Q700	Z850	Z500	Pobs	PRATE	Q700	Z850	Z500
1	RAPEL	0,000	0,000	0,007	0,174	0,925	0,000	0,000	0,000	0,020	0,717	0,000	0,000	0,000	0,001	0,082	0,000	0,000	0,000	0,101	0,017	0,000	0,000	0,000	0,188	0,022	0,000	0,000	0,004	0,014	0,949
2	SANTO DOMINGO AD.	0,000	0,000	0,092	0,572	0,579	0,000	0,000	0,005	0,537	0,066	0,000	0,000	0,013	0,003	0,453	0,000	0,000	0,057	0,156	0,053	0,000	0,000	0,000	0,115	0,000	0,000	0,000	0,145	0,001	0,294
3	ESTERO PUANGUE RUTA 78	0,002	0,007	0,651	0,080	0,320	0,000	0,000	0,006	0,259	0,915	0,000	0,000	0,049	0,344	0,022	0,000	0,000	0,000	0,952	0,032	0,000	0,000	0,012	0,851	0,040	0,000	0,000	0,099	0,222	0,814
4	LOS GUINDOS	0,016	0,000	0,237	0,111	0,278	0,000	0,000	0,022	0,475	0,450	0,000	0,000	0,000	0,010	0,555	0,000	0,000	0,005	0,129	0,057	0,000	0,000	0,000	0,925	0,436	0,000	0,000	0,099	0,205	0,400
5	QUILLOTA	0,000	0,000	0,084	0,220	0,405	0,000	0,000	0,038	0,713	0,386	0,000	0,000	0,002	0,080	0,519	0,000	0,000	0,013	0,074	0,095	0,000	0,000	0,001	0,847	0,023	0,000	0,000	0,064	0,018	0,827
6	BARRERA LONCHA	0,029	0,001	0,664	0,143	0,775	0,000	0,000	0,176	0,677	0,343	0,000	0,000	0,138	0,005	0,491	0,000	0,000	0,050	0,210	0,049	0,000	0,000	0,011	0,508	0,421	0,000	0,000	0,615	0,022	0,510
7	RODELILLO	0,000	0,000	0,049	0,051	0,173	0,000	0,000	0,288	0,312	0,624	0,000	0,000	0,001	0,081	0,318	0,000	0,000	0,006	0,225	0,030	0,000	0,000	0,000	0,858	0,336	0,000	0,000	0,339	0,011	0,983
8	LO ROJAS	0,003	0,000	0,088	0,342	0,420	0,000	0,000	0,069	0,725	0,097	0,000	0,000	0,002	0,138	0,784	0,000	0,000	0,003	0,521	0,010	0,000	0,000	0,001	0,561	0,260	0,000	0,000	0,286	0,026	0,111
9	LOS PANGUILES	0,000	0,000	0,003	0,179	0,851	0,000	0,000	0,022	0,086	0,211	0,000	0,000	0,000	0,264	0,037	0,000	0,000	0,000	0,383	0,032	0,000	0,000	0,000	0,261	0,351	0,000	0,000	0,001	0,001	0,540
10	LAGUNILLAS	0,000	0,000	0,073	0,078	0,092	0,000	0,000	0,043	0,473	0,799	0,000	0,000	0,000	0,029	0,201	0,000	0,000	0,000	0,249	0,088	0,000	0,000	0,000	0,636	0,179	0,000	0,000	0,016	0,039	0,992
11	LITUECHE	0,000	0,000	0,002	0,345	0,553	0,000	0,000	0,010	0,851	0,768	0,000	0,000	0,000	0,002	0,769	0,000	0,000	0,000	0,379	0,054	0,000	0,000	0,000	0,542	0,358	0,000	0,000	0,002	0,064	0,666
12	COLTAUCO	0,001	0,000	0,132	0,057	0,392	0,000	0,000	0,031	0,639	0,789	0,000	0,000	0,000	0,023	0,200	0,000	0,000	0,009	0,220	0,034	0,000	0,000	0,000	0,550	0,230	0,000	0,000	0,200	0,063	0,590
13	EL VERGEL	0,000	0,000	0,003	0,323	0,506	0,000	0,000	0,051	0,738	0,557	0,000	0,000	0,000	0,126	0,331	0,000	0,000	0,000	0,358	0,019	0,000	0,000	0,000	0,588	0,053	0,000	0,000	0,007	0,002	0,858
14	LAGUNA ACULEO	0,000	0,000	0,159	0,032	0,813	0,000	0,000	0,038	0,472	0,410	0,000	0,000	0,000	0,049	0,516	0,000	0,000	0,007	0,224	0,048	0,000	0,000	0,000	0,644	0,088	0,000	0,000	0,279	0,012	0,929
15	PUDAHUEL SANTIAGO	0,000	0,000	0,440	0,666	0,602	0,000	0,000	0,209	0,483	0,420	0,000	0,000	0,204	0,136	0,154	0,000	0,000	0,003	0,272	0,061	0,000	0,000	0,001	0,161	0,067	0,000	0,000	0,469	0,007	0,724
16	RANCAGUA (CACHAPOAL-DCP)	0,000	0,000	0,097	0,318	0,129	0,000	0,000	0,009	0,533	0,280	0,000	0,000	0,003	0,008	0,108	0,000	0,000	0,000	0,130	0,016	0,000	0,000	0,000	0,586	0,034	0,000	0,000	0,027	0,017	0,546
17	QUINTA NORMAL SANTIAGO	0,000	0,000	0,071	0,448	0,937	0,000	0,000	0,486	0,973	0,383	0,000	0,000	0,100	0,416	0,376	0,000	0,000	0,010	0,671	0,075	0,000	0,000	0,002	0,242	0,083	0,000	0,000	0,228	0,116	0,513
18	TERRAZAS OFICINA CENTRAL DGA	0,000	0,000	0,128	0,337	0,473	0,000	0,000	0,157	0,346	0,821	0,000	0,000	0,006	0,477	0,470	0,000	0,000	0,000	0,045	0,120	0,000	0,000	0,000	0,380	0,625	0,000	0,000	0,102	0,003	0,802
19	HUECHUN EMBALSE	0,000	0,000	0,161	0,491	0,865	0,000	0,000	0,095	0,944	0,060	0,000	0,000	0,019	0,657	0,377	0,000	0,000	0,008	0,567	0,040	0,000	0,000	0,004	0,859	0,148	0,000	0,000	0,123	0,153	0,852
20	EULOGIO SANCHEZ AD	0,000	0,000	0,475	0,716	0,023	0,000	0,000	0,024	0,632	0,859	0,000	0,000	0,340	0,104	0,197	0,000	0,000	0,002	0,389	0,087	0,000	0,000	0,007	0,706	0,213	0,000	0,000	0,030	0,114	0,655
21	PIRQUE	0,000	0,000	0,470	0,554	0,819	0,000	0,000	0,138	0,932	0,770	0,000	0,000	0,067	0,094	0,566	0,000	0,000	0,012	0,329	0,032	0,000	0,000	0,012	0,910	0,237	0,000	0,000	0,243	0,177	0,773
22	QUEBRADA RAMON	0,000	0,000	0,162	0,813	0,309	0,000	0,000	0,044	0,356	0,204	0,000	0,000	0,281	0,218	0,650	0,000	0,000	0,034	0,119	0,017	0,000	0,000	0,031	0,228	0,332	0,000	0,000	0,022	0,008	0,216
23	LA OBRA RECINTO REMOS	0,000	0,000	0,206	0,432	0,295	0,000	0,000	0,095	0,962	0,518	0,000	0,000	0,060	0,116	0,510	0,000	0,000	0,001	0,344	0,013	0,000	0,000	0,002	0,658	0,208	0,000	0,000	0,026	0,067	0,700
24	CERRO CALAN	0,000	0,000	0,035	0,436	0,231	0,000	0,000	0,022	0,880	0,274	0,000	0,000	0,233	0,323	0,408	0,000	0,000	0,007	0,648	0,027	0,000	0,000	0,001	0,659	0,394	0,000	0,000	0,009	0,034	0,214
25	ANTUPIREN	0,000	0,000	0,003	0,410	0,381	0,000	0,000	0,020	0,970	0,320	0,000	0,000	0,000	0,365	0,451	0,000	0,000	0,000	0,443	0,016	0,000	0,000	0,000	0,505	0,317	0,000	0,000	0,002	0,201	0,723
26	SAN JOSE DE MAIPO RETEN	0,001	0,000	0,008	0,370	0,341	0,000	0,000	0,001	0,405	0,384	0,000	0,000	0,000	0,016	0,236	0,000	0,000	0,057	0,291	0,078	0,000	0,000	0,000	0,067	0,096	0,000	0,000	0,017	0,001	0,805
27	MAITENES BOCATOMA	0,001	0,000	0,032	0,037	0,092	0,000	0,000	0,007	0,255	0,593	0,000	0,000	0,000	0,005	0,317	0,000	0,000	0,008	0,078	0,345	0,000	0,000	0,000	0,294	0,709	0,000	0,000	0,005	0,002	0,691
28	SAN GABRIEL	0,000	0,000	0,016	0,315	0,104	0,000	0,000	0,006	0,893	0,395	0,000	0,000	0,000	0,027	0,642	0,000	0,000	0,027	0,396	0,559	0,000	0,000	0,000	0,502	0,810	0,000	0,000	0,027	0,014	0,616
29	EL YESO EMBALSE	0,000	0,000	0,001	0,133	0,837	0,000	0,000	0,001	0,558	0,584	0,000	0,000	0,000	0,024	0,829	0,000	0,000	0,001	0,070	0,417	0,000	0,000	0,000	0,048	0,428	0,000	0,000	0,072	0,001	0,246

NORMALIDAD DE LOS MODELOS

Tabla 26 P-valores para el test de Shapiro Wilk para verificar la normalidad de los modelos lineales. Fuente: elaboración propia.

Estación pluviométrica		Abril	Mayo	Junio	Julio	Agosto	Sept.
1	RAPEL	0,001	0,000	0,000	0,000	0,000	0,000
2	SANTO DOMINGO AD.	0,000	0,000	0,000	0,000	0,000	0,000
3	ESTERO PUANGUE RUTA 78	0,005	0,000	0,000	0,000	0,000	0,011
4	LOS GUINDOS	0,008	0,000	0,000	0,013	0,000	0,000
5	QUILLOTA	0,054	0,000	0,000	0,000	0,001	0,000
6	BARRERA LONCHA	0,044	0,000	0,000	0,000	0,000	0,000
7	RODELILLO	0,048	0,000	0,002	0,000	0,000	0,000
8	LO ROJAS	0,551	0,000	0,000	0,000	0,062	0,000
9	LOS PANGUILES	0,277	0,000	0,000	0,000	0,000	0,000
10	LAGUNILLAS	0,112	0,000	0,000	0,000	0,000	0,000
11	LITUECHE	0,064	0,000	0,000	0,000	0,000	0,000
12	COLTAUCO	0,000	0,000	0,000	0,000	0,000	0,000
13	EL VERGEL	0,000	0,000	0,000	0,000	0,000	0,000
14	LAGUNA ACULEO	0,000	0,000	0,000	0,000	0,000	0,000
15	PUDAHUEL SANTIAGO	0,000	0,000	0,000	0,000	0,000	0,000
16	RANCAGUA (CACHAPOAL-DCP)	0,000	0,000	0,000	0,000	0,000	0,000
17	QUINTA NORMAL SANTIAGO	0,000	0,000	0,000	0,000	0,000	0,000
18	TERRAZAS OFICINA CENTRAL DGA	0,018	0,000	0,000	0,001	0,000	0,001
19	HUECHUN EMBALSE	0,000	0,000	0,000	0,000	0,000	0,000
20	EULOGIO SANCHEZ AD	0,001	0,000	0,000	0,000	0,000	0,000
21	PIRQUE	0,000	0,000	0,000	0,000	0,000	0,000
22	QUEBRADA RAMON	0,000	0,000	0,000	0,000	0,000	0,000
23	LA OBRA RECINTO REMOS	0,001	0,000	0,000	0,000	0,000	0,000
24	CERRO CALAN	0,000	0,000	0,000	0,000	0,000	0,000
25	ANTUPIREN	0,093	0,011	0,001	0,000	0,000	0,000
26	SAN JOSE DE MAIPO RETEN	0,084	0,001	0,000	0,000	0,000	0,000
27	MAITENES BOCATOMA	0,011	0,008	0,000	0,053	0,000	0,000
28	SAN GABRIEL	0,000	0,001	0,001	0,000	0,000	0,000
29	EL YESO EMBALSE	0,000	0,000	0,000	0,000	0,000	0,000

4. NO-COLINEALIDAD

Tabla 27: Test de colinealidad valores de FIV. Fuente: elaboración propia.

Estación pluviométrica		Abril	Mayo	Junio	Julio	Agosto	Sept.
1	RAPEL	2,8	1,7	1,8	2,5	2,7	1,8
2	SANTO DOMINGO AD.	2,4	1,9	1,4	2,5	2,6	2,1
3	ESTERO PUANGUE RUTA 78	1,9	1,3	1,8	2,2	2,2	2,1
4	LOS GUINDOS	2,0	1,7	1,7	3,1	2,8	1,5
5	QUILLOTA	1,8	1,4	1,5	1,8	2,3	2,6
6	BARRERA LONCHA	1,9	1,7	1,8	2,1	2,5	2,4
7	RODELILLO	1,4	1,2	1,6	1,5	1,9	1,8
8	LO ROJAS	1,6	1,3	1,8	2,4	2,5	2,8
9	LOS PANGUILES	1,4	2,3	2,2	1,8	2,3	3,0
10	LAGUNILLAS	1,9	1,4	1,9	1,6	2,6	2,7
11	LITUECHE	2,0	1,7	1,8	3,2	2,4	1,9
12	COLTAUCO	1,5	1,7	2,3	2,3	2,6	1,4
13	EL VERGEL	1,1	2,1	2,0	2,2	3,4	2,5
14	LAGUNA ACULEO	1,3	1,9	1,9	2,5	2,7	1,2
15	PUDAHUEL SANTIAGO	1,4	1,7	1,8	1,9	2,4	1,7
16	RANCAGUA (CACHAPOAL-DCP)	1,1	2,2	1,8	2,4	3,3	1,9
17	QUINTA NORMAL SANTIAGO	1,5	1,8	1,8	1,9	2,5	1,8
18	TERRAZAS OFICINA CENTRAL DGA	1,9	1,5	1,7	2,3	2,1	2,4
19	HUECHUN EMBALSE	1,6	2,0	2,0	2,0	2,3	2,0
20	EULOGIO SANCHEZ AD	1,3	1,7	2,1	2,1	1,8	1,6
21	PIRQUE	1,1	2,2	2,4	2,7	2,7	1,8
22	QUEBRADA RAMON	1,7	1,6	2,0	2,0	1,8	1,3
23	LA OBRA RECINTO REMOS	1,4	1,7	2,3	2,2	1,8	1,8
24	CERRO CALAN	1,4	1,7	2,1	2,3	2,3	1,4
25	ANTUPIREN	2,1	1,7	2,4	1,7	2,0	1,3
26	SAN JOSE DE MAIPO RETEN	1,4	1,9	1,9	2,3	2,4	1,1
27	MAITENES BOCATOMA	1,1	1,5	1,7	1,7	2,0	1,1
28	SAN GABRIEL	1,3	1,9	1,7	1,5	1,8	1,3
29	EL YESO EMBALSE	1,3	1,6	1,9	2,5	1,8	1,1

E. RELACIÓN ENTRE VARIABLES

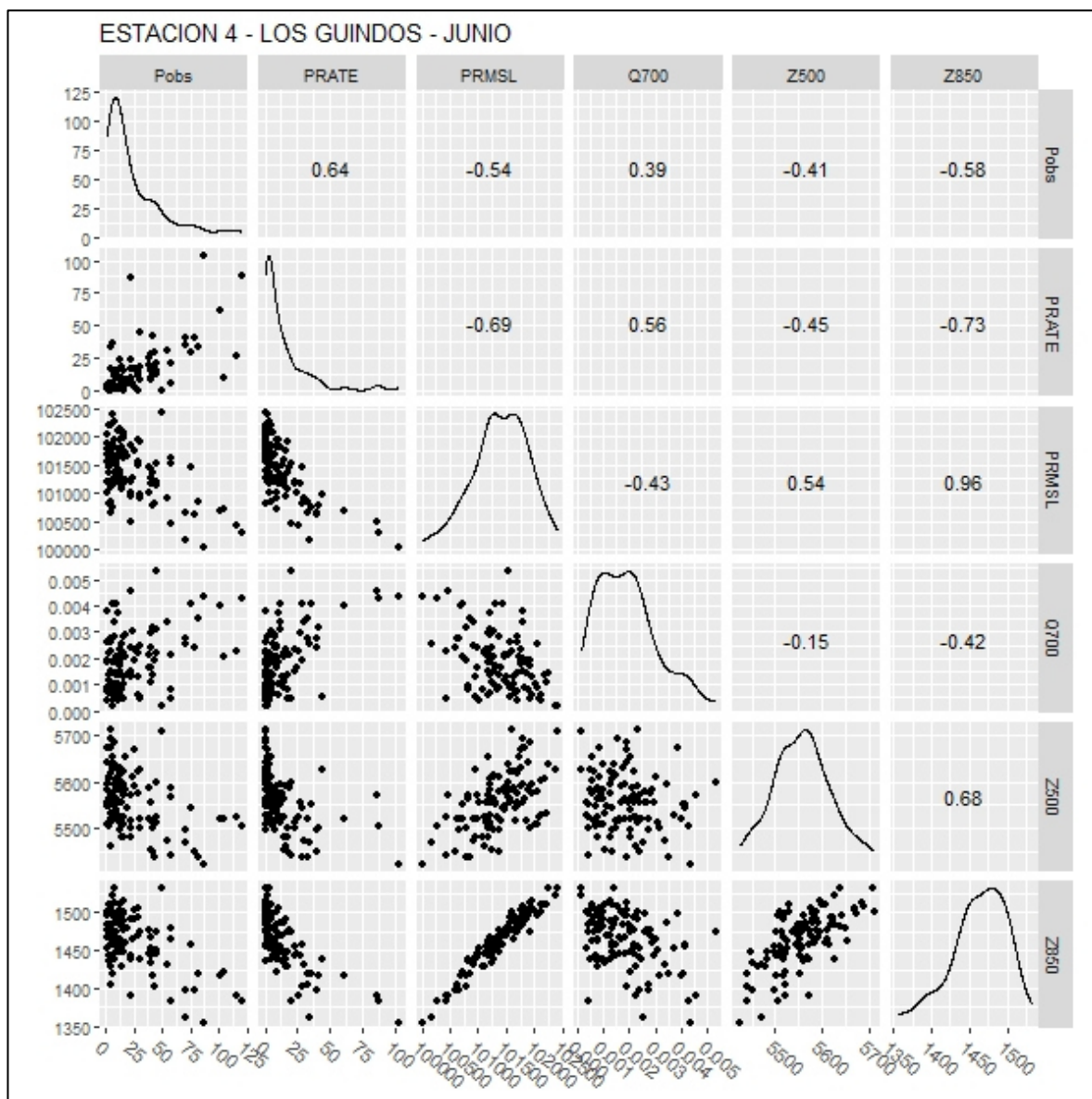


Figura 41: Gráfica y correlación entre variables en estación 4.
Fuente: Elaboración propia

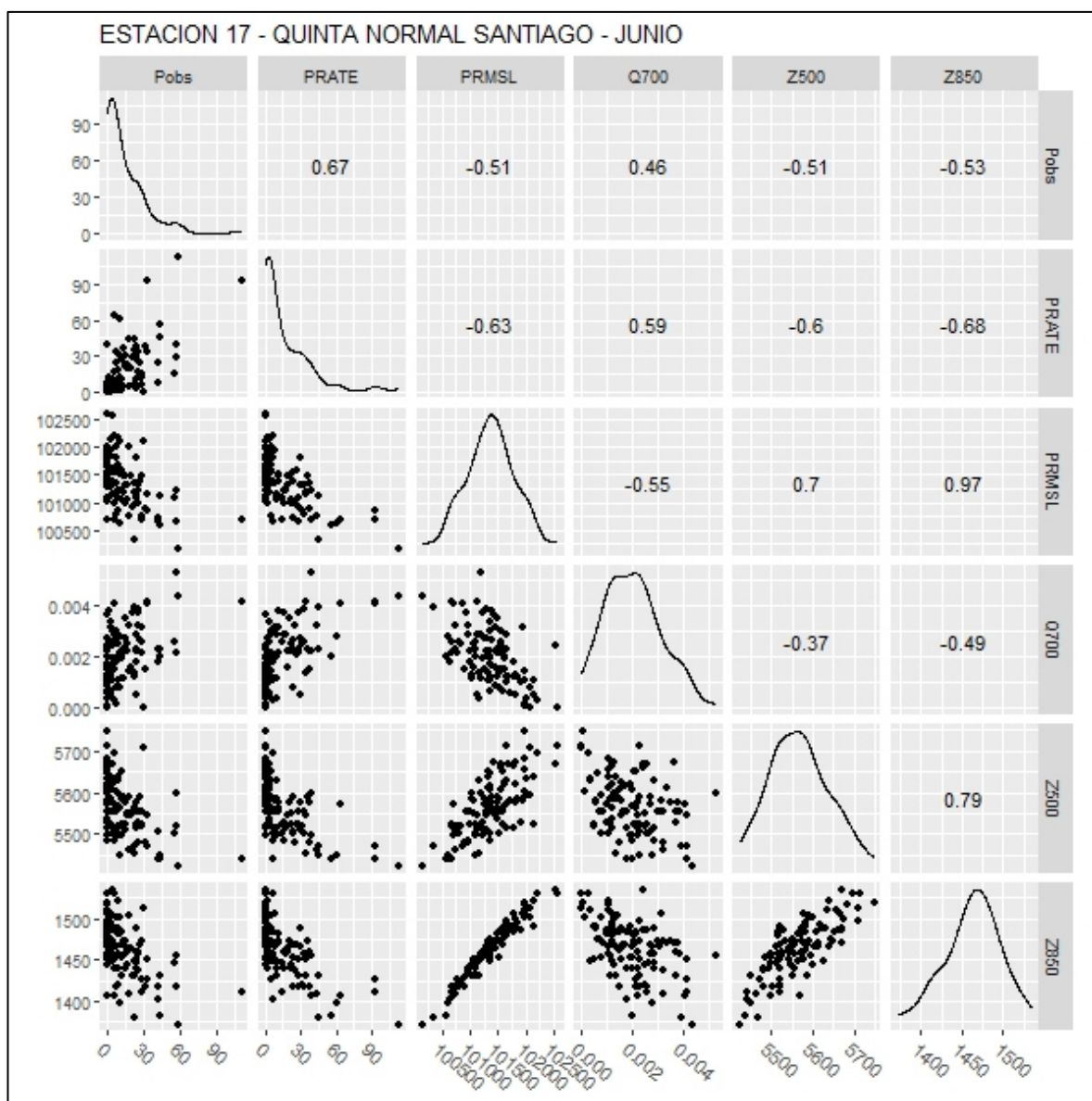


Figura 42: Gráfica y correlación entre variables en estación 19.
Fuente: Elaboración propia

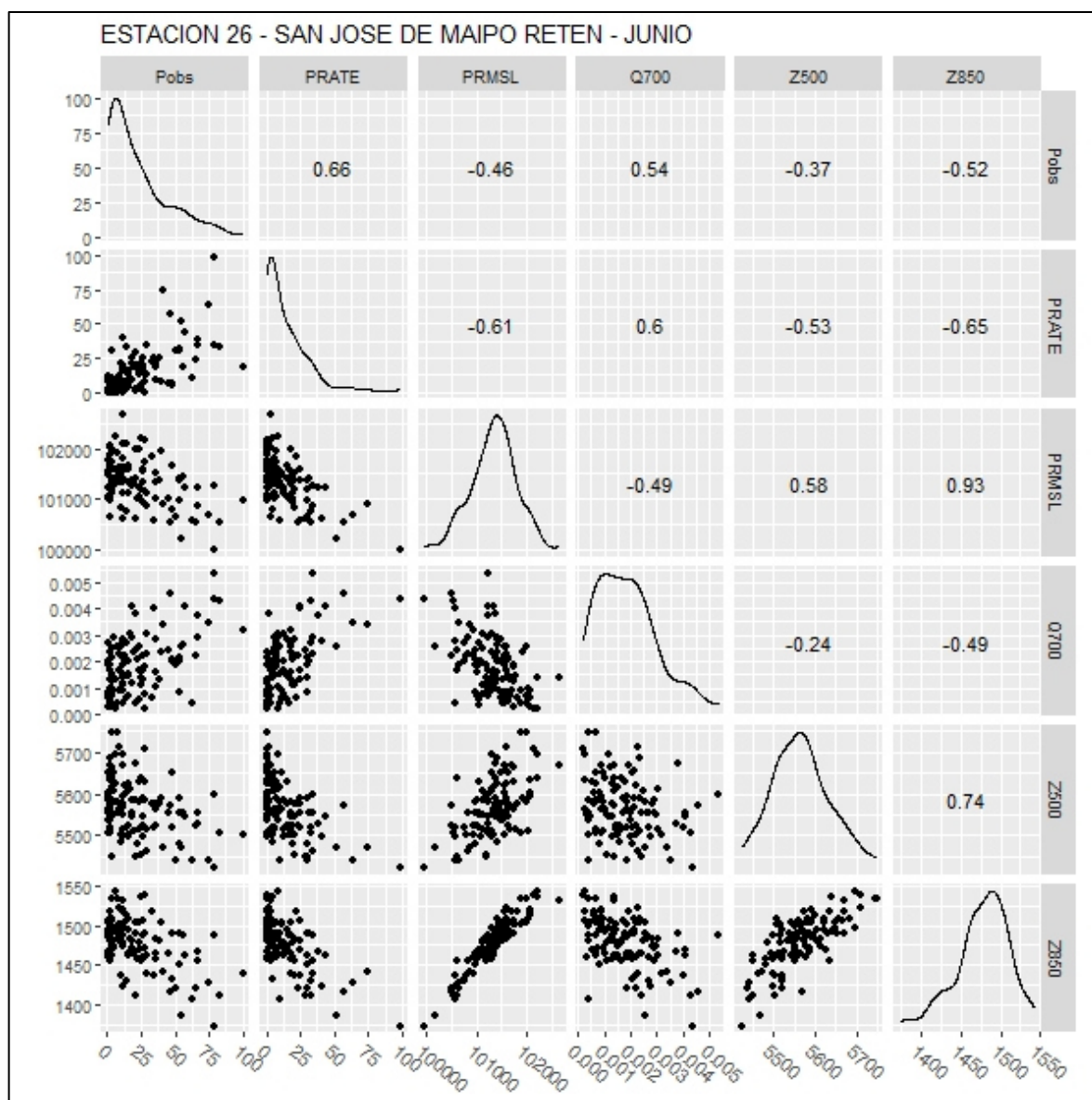


Figura 43: Gráfica y correlación entre variables en estación 26.
Fuente: Elaboración propia