



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA CIVIL

DESARROLLO DE UN PRODUCTO GRILLADO DE PRECIPITACIÓN Y TEMPERATURA DIARIA PARA APLICACIONES HIDROLÓGICAS EN LA PUNA CHILE - ARGENTINA

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL

JUAN EDUARDO GALLEGOS LILLO

PROFESOR GUÍA:

EDUARDO MUÑOZ CASTRO

PROFESORA CO-GUÍA:

MIJAL BARON ZAGAL

MIEMBROS DE LA COMISIÓN:

ALONSO PIZARRO VALDEBENITO

XIMENA VARGAS MESA

Este trabajo ha sido parcialmente financiado por:
Montgomery & Associates Consultores Limitada

SANTIAGO DE CHILE

2024

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO CIVIL
POR: Juan Eduardo Gallegos Lillo
FECHA: 2024
PROFESOR GUÍA: Eduardo Muñoz-Castro

DESARROLLO DE UN PRODUCTO GRILLADO DE PRECIPITACIÓN Y TEMPERATURA DIARIA PARA APLICACIONES HIDROLÓGICAS EN LA PUNA CHILE - ARGENTINA

La Puna Chileno-Argentina se caracteriza por su topografía compleja, la presencia de diversas especies de flora y fauna endémica, y, en términos industriales, la presencia de grandes yacimientos de litio como, por ejemplo, los salares de Atacama (Chile) y Hombre Muerto (Argentina). La actividad industrial en la zona demanda estudio vinculados a la disponibilidad hídrica para efectos de la estimación de la recarga histórica y futura y, asimismo, precipitaciones máximas para el análisis de crecidas pluviales y vulnerabilidad de la infraestructura emplazada en las faenas. No obstante, la escasez de información meteorológica en la zona es un problema que no se ha abordado a partir de un enfoque orientado a caracterizar los patrones de variabilidad climática local.

El presente estudio aborda el desarrollo de un producto grillado diario y una resolución espacial de $0,05^\circ \times 0,05^\circ$, de base observacional de precipitaciones y temperaturas en la Puna Chile-Argentina, denominado RFPUNA, con el objetivo de abordar el problema de escasez de mediciones meteorológicas *in-situ*. Para ello, se aplica el método de *Machine Learning* conocido como *Random Forest*, el cual es configurado con diversos predictores como, por ejemplo, las características fisiográficas del área de estudio. Además, se explora una versión estocástica a través de la implementación del algoritmo estadístico *Bluecat*.

La revisión de los productos meteorológicos grillados disponibles en la zona de estudio muestra que IMERG y ERA5 presentan, en general, el mejor desempeño cuando se contrasta con las observaciones de precipitación y temperatura respectivamente, en el periodo entre 2007 y 2017. Lo anterior se refleja en métricas como KGE' (0,05 y 0,62), RMSE (2,13 [mm] y 2,59[°C]), coeficiente de correlación de Pearson (0,18 y 0,89) aplicadas a IMERG y ERA5 a escala diaria respectivamente.

Los resultados muestran que el producto grillado propuesto logra buenos valores de precipitación y temperatura para KGE' (0,25 y 0,78), RMSE (1,60 [mm] y 2,68 [°C]) y correlación (0,44 y 0,93) a escala diaria respectivamente. Los resultados indican una mejora notable de la precisión en las estimaciones de precipitación, lo que es vital para la planificación y gestión estratégica de los recursos naturales y la mitigación del impacto ambiental de la minería del litio. No obstante, las estimaciones de temperatura no han logrado superar a las predicciones de ERA5 ni CHELSA, abriendo la discusión sobre la aplicación de *Random Forest* para esta variable climática. De igual forma, la aplicación de *Bluecat* entrega la posibilidad de elaborar un inédito producto grillado probabilístico en la zona de la Puna Chile - Argentina.

Finalmente, el presente estudio desarrolla un producto de precipitación con un mayor rendimiento en la estimación de precipitaciones respecto de los productos globales estudiados. Esto permitiría una mayor comprensión de la distribución espacial y temporal de las precipitaciones en el área de estudio, pudiendo estimar con mayor precisión los caudales superficiales y subterráneos, lo que abre múltiples oportunidades de mejoras en aplicaciones hidrológicas relacionadas a precipitación y caudales, como la configuración de diversos modelos hidrológicos que capturen con mayor precisión la disponibilidad y la distribución espacial y temporal del recurso hídrico.

A mis padres, hermana, familia, amigos y vecinos...

Y hay que acudir corriendo pues se cae...

el porvenir

AGRADECIMIENTOS

Luego de seis duros, pero buenos años, se logra uno de los primeros objetivos propuestos al comenzar mi vida adulta, ser Ingeniero Civil Hidráulico. Fue un largo camino para decidir y cumplir esta primera meta, donde conté con el apoyo y el consejo oportuno de toda mi familia y amigos, quienes siempre brindaron una mano amiga para mi formación profesional. Por supuesto, es necesario mencionar a todos esos profesores que confirmaron mi decisión y me motivaron aún más para alcanzar el objetivo desde lo académico; profesores Miguel Lagos, Franco Ricchetti, Pablo Mendoza, Marcelo Olivares, Juvenal Letelier, y la profesora Katherine Lizama; y desde lo profesional, agradecer a Montgomery & Associates y a mis colegas por la oportunidad y la confianza de investigar para el desarrollo de la empresa.

Sin embargo, la recta final implica disciplina, orden y mucha sabiduría, elementos que solo un grupo de mentores puede otorgar. En ese sentido, le debo muchísimo a Edu, el futuro académico Eduardo Muñoz Castro, quien explotó un potencial desconocido y supo guiarme cuando pude haber perdido el rumbo; no tengo dudas de que serás el mejor en lo que te propongas. También, expresar mi eterna gratitud a mi profesora co-guía Mijal Baron Zagal, la persona que confió en mí desde esa primera llamada telefónica hasta el día de hoy, con la memoria lista y a punto de comenzar mi vida profesional, brindándome un consejo amigo en cada paso. Por último, pero no menos importante, agradecer a la profesora Ximena Vargas Mesa por su inmensa sensatez y amabilidad, y al profesor Alonso Pizarro Valdebenito y sus cafecitos, por abrirme los ojos en este camino de la investigación y motivarme a alcanzar cualquier meta.

El objetivo sigue lejos de estar completo, y ahora viene la parte más complicada, no perderse en el camino. Por suerte, sé que mis seres queridos sabrán reprenderme cuando las dudas nublen el horizonte. Ya me imagino a Alonso, el pollo, retándome mientras nos tomamos unas cervezas; a Nicolás y Felipe, los guatones, echándome la talla pa' tirar pa' arriba; a Diego y Sofia, los primitos, soltando esas cosas que se aferran e inquietan al alma con un tecito; Donoso, Luciano y Mathew, luego de alguna pichanga arreglando el mundo; a Claudio, Ignacio y Seba, sentados en la Ebria poniéndonos al día y riéndonos a más no poder; a los Gabo, Cereceda y Chávez, alegrando el día con cada anécdota; a Ignacia, mi amiga Civil, apañándome en todas como siempre; a Ricardo, mandándome algún meme para alegrar la vida, y tantos otros más, tan únicos y especiales a su modo, que llenan mi vida y alma con su presencia.

Pero, si el consejo amigo no es suficiente, sé que contaré con las personas más indispensables en mi vida, mi familia. Sé que mi hermana, Elisa, cambiará su tono de voz y me retará para hacerme entender; que mis abuelos sabrán darme su sabio consejo; que mi padre, Eduardo, se sentará conmigo y charlaremos; y que mi madre, Elisa, con su eterna comprensión y seriedad, me dirá las cosas tal como son y me conducirá para lograr lo más importante, ser una persona íntegra.

Y si llega a faltar todo lo anterior, siempre tendré a mi mejor amigo, mi perro Chopin, un corazón sincero con cuatro patas y un plumero.

TABLA DE CONTENIDO

1. INTRODUCCIÓN	1
1.1 MOTIVACIÓN	1
1.2 OBJETIVOS.....	3
1.3 ESTRUCTURA DEL DOCUMENTO.....	3
2. MARCO TEÓRICO	4
2.1 CLIMA EN LA PUNA CHILE-ARGENTINA	5
2.2 PRODUCTOS METEOROLÓGICOS GRILLADOS	6
2.3 MÉTODOS PARA LA GENERACIÓN DE PRODUCTOS GRILLADOS METEOROLÓGICOS.....	9
2.3.1 Metodologías convencionales de observación	9
2.3.2 Productos de base satelital y percepción remota	10
2.3.3 Reanálisis climático.....	10
2.3.4 Aprendizaje Automático (Machine Learning – ML).....	11
2.3.5 Random Forest.....	12
2.4 ENFOQUES ESTOCÁSTICOS.....	13
2.4.1 Algoritmo Bluecat.....	13
2.5 SÍNTESIS DE LA REVISIÓN E HIPÓTESIS DE INVESTIGACIÓN	14
3. ZONA DE ESTUDIO	15
3.1 ANTECEDENTES GENERALES.....	15
3.2 CARACTERIZACIÓN GEOMORFOLÓGICA.....	17
3.3 HIDROCLIMATOLOGÍA.....	20
4. DATOS DISPONIBLES	23
4.1 INFORMACIÓN METEOROLÓGICA.....	23
4.1.1 Estaciones meteorológicas	24
4.1.2 Revisión y filtro de estaciones según calidad.....	27
4.2 DISPONIBILIDAD DE PRODUCTOS GRILLADOS.....	29
4.2.1 Precipitación.....	29
4.2.2 Temperatura.....	31
5. METODOLOGÍA	33
5.1 METODOLOGÍA GENERAL	33
5.2 PRODUCTO RFPUNA: PRECIPITACIÓN	33
5.3 PRODUCTO RFPUNA: TEMPERATURA.....	36
5.4 APLICACIÓN DE ALGORITMO <i>BLUECAT</i>	37
5.5 DIAGNÓSTICO Y AJUSTE DE LOS PRODUCTOS	38
5.5.1 Análisis por estación (puntual)	38
5.5.2 Análisis de distribución espacial.....	40
5.5.3 Análisis de importancia de predictores	41
6. RESULTADOS	42

6.1	ANÁLISIS DE VARIABLES Y CONFIGURACIÓN DEL MODELO	42
6.1.1	<i>Precipitación</i>	43
6.1.2	<i>Temperatura</i>	44
6.2	EVALUACIÓN Y VERIFICACIÓN DE LOS PRODUCTOS DE PRECIPITACIÓN Y TEMPERATURA	45
6.2.1	<i>Precipitación</i>	45
6.2.2	<i>Temperatura</i>	54
6.3	RESUMEN	57
7.	DISCUSIONES Y COMENTARIOS	60
7.1	ALCANCES Y LIMITACIONES DEL PRODUCTO	60
7.2	SOBRE EL POTENCIAL DE APLICACIONES HIDROLÓGICAS.....	61
7.3	TRABAJO FUTURO	61
8.	CONCLUSIONES	63
	BIBLIOGRAFÍA	66
	ANEXO A: RESULTADOS DE LA CALIBRACIÓN Y VERIFICACIÓN DEL MODELO	74
	ANEXO B: REGISTRO DE ESTACIONES.....	86
	ANEXO C: DATOS DE ENTRADA AL MODELO.....	98
	ANEXO D: CURVAS DE VARIACIÓN ESTACIONAL.....	100

ÍNDICE DE TABLAS

Tabla 2.1: Productos grillados de interés en la zona de estudio y sus características.	8
Tabla 3.1: Características geomorfológicas del área de estudio.....	17
Tabla 4.1: Distribución de estaciones meteorológicas en una red básica. Datos extraídos de López (2009).	23
Tabla 4.2: Número de estaciones con una cantidad determinada de datos diarios para un número de años seguidos.	28
Tabla 4.3: Mediana de los resultados de los índices de KGE' para los distintos productos globales de precipitación.....	30
Tabla 4.4: Mediana de los resultados de los índices de KGE' para los distintos productos globales de temperatura.	31
Tabla 5.1: Métricas de la tabla de contingencia.	39
Tabla 5.2: Clasificación de eventos basados en intensidad de lluvia (preliminar).....	40
Tabla 6.1: Descripción de los productos grillados obtenidos.....	46
<i>Tabla 6.2: Mediana de los resultados de los índices de KGE' para los distintos productos de precipitación.</i>	<i>48</i>
Tabla 6.3: Mediana de los resultados de los índices de KGE' para los distintos productos de temperatura.	55
Tabla 6.4: Autocorrelación espacial según índice de Morán. Prec y Temp hacen referencia a las variables de precipitación y temperatura respectivamente.	58

ÍNDICE DE FIGURAS

Figura 1.1: Mapas de la meseta Andina y los salares. En a) se muestra el Triángulo del Litio (rosado), la Medialuna del Litio (verde) y todos los salares altiplánicos. En b), se observa la ubicación continental.....	6
Figura 2.1: Ocean Niño Index (ONI) entre 1948 a la actualidad. Datos extraídos de Climate Prediction Center - ONI, 2023.....	6
Figura 2.2: Diagrama base de Random Forest. Fuente: Khan et al., 2021.....	12
Figura 3.1: Mapa de la extensión de la alta meseta de los Andes Centrales. Fuente: Allmendinger et al., 1997	15
Figura 3.2: Zona de estudio con sus principales características morfológicas y climáticas. (a) Modelo de elevación digital SRTM; (b) Clasificación climática de Köppen-Geiger; (c) Vista general de la zona de estudio respecto a Sudamérica.	16
Figura 3.3: Cuencas hidrográficas en el área de estudio.	18
Figura 3.4: Curvas hipsométricas de las cuencas hidrográficas en el área de estudio.	19
Figura 3.5: Curvas de distribución de probabilidad para los atributos de elevación, área, pendiente y compartición de las cuencas hidrográficas del área de estudio.....	20
Figura 3.6: Perfiles de los valores medianos de precipitación anual según longitud, latitud y elevación.....	21
Figura 3.7: Perfiles de los valores medianos de temperatura media anual según longitud, latitud y elevación.....	21
Figura 4.1: Cantidad de estaciones de precipitación (izquierda) y temperatura (derecha) según distancias a la redonda de cada estación.....	24
Figura 4.2: Área de estudio con elevación y estaciones meteorológicas con información de precipitaciones.....	25
Figura 4.3: Área de estudio con elevación y estaciones meteorológicas con información de temperatura.....	26
Figura 4.4: Distribución de estaciones según país de origen y administración de la información.....	27
Figura 4.5: a,b: Gráfico de selección de estaciones según criterio escogido de años de data completa. En a) se considera el total de años de información completa, mientras que en b) el año es completo si tiene al menos el 85% de los días con información.....	28
Figura 4.6: Valor de KGE' para los productos grillados de precipitación ERA5, PERSIANN-CDR, IMERG, CHIRPSv2 y CHELSA en la zona de estudio para el periodo 2007-2016 a escala diaria. En línea segmentada color rojo se presenta el valor óptimo de cada indicador.....	30
Figura 4.7: Valor de KGE' para los productos grillados de temperatura ERA5 y CHELSA en la zona de estudio para el periodo 2007-2016 a escala diaria. En línea segmentada color rojo se presenta el valor óptimo de cada indicador.....	31
Figura 5.1: Metodología para configuración de las precipitaciones del producto grillado RFPUNA. En rojo se encierran los pasos que presentan similitudes con la configuración de RFMEP.....	35

Figura 5.2: Metodología para configuración del producto grillado RFPUNA de temperatura.	37
Figura 6.1: Correlación de las variables en estudio a escala anual en toda el área de estudio. Se consideraron los periodos anuales entre 2007-2016.....	42
Figura 6.2: Índice de Gini de los predictores de Random Forest para precipitación. Se consideró el periodo entre 2007-2017.....	44
Figura 6.3: Índice de Gini de los predictores de Random Forest para temperatura. Se consideró el periodo entre 2007-2016.....	45
Figura 6.4: Gráficos de caja asociados a productos grillados de precipitación y su KGE' con los parámetros r , β y γ para el periodo 2007-2017 en toda el área de estudio.	47
Figura 6.5: Métricas de los productos grillados de precipitación para el periodo 2007-2017.	50
Figura 6.6: Coeficientes de variación estacional y distribución espacial de cada producto junto a la clasificación Köppen-Geiger del área de estudio.	51
Figura 6.7: Distribución espacial de las precipitaciones simuladas por los distintos productos grillados.	52
Figura 6.8: Distribución espacial de los valores KGE' para distintos productos grillados en el área de estudio.....	53
Figura 6.9: Curvas de variación estacional para las estaciones Ayquina y Ollague para 4 distintas probabilidades de excedencia para el periodo 2007-2017.....	54
Figura 6.10: Gráficos de caja de productos grillados de temperatura y su KGE' con los parámetros r , β y γ para el periodo 2007-2016 en toda el área de estudio.	55
Figura 6.11: Distribución espacial de las temperaturas simuladas por los distintos productos grillados.	57
Figura 6.12: Comparación de KGE' determinístico contra el KGE' de la media estocástica según la longitud. Como referencia, la longitud -67.5^{a}W divide Chile de Argentina. La línea entrecortada roja representa la recta 1:1.	58
Figura 6.13: Distribución de precipitaciones y temperaturas medias anuales observadas y simuladas por el producto RFPUNA y ERA5 en el periodo 2007-2016.....	59

1. INTRODUCCIÓN

1.1 Motivación

La variabilidad en la precipitación y temperatura desempeñan un papel fundamental en la toma de decisiones estratégicas para la gestión de recursos naturales y diversos sectores, como la agricultura y la minería. La creación de un producto grillado que analice estos elementos permitirá no solo comprender mejor la dinámica climática en una determinada zona de interés, sino que también proporcionar información de utilidad para la planificación y adaptación ante posibles cambios futuros.

En los últimos años, la puna chileno-argentina ha generado un gran interés, principalmente en la emergente industria minera del Litio, debido a las grandes reservas de este mineral en la zona, donde los territorios chilenos y argentinos concentran el 41 y 10% de las reservas mundiales del mineral a 2021 (U.S. Geological Survey, 2023). La extracción del litio es una actividad económica de gran importancia para el desarrollo económico de los países en cuestión, puesto que este mineral presenta una alta demanda tecnológica, representando el 8,2 y el 0,8% de las exportaciones de las economías de Chile y Argentina respectivamente, traduciéndose en un 3,4% de aporte al Producto Interno Bruto (PIB) de la economía chilena en 2022 (Banco Central de Chile, 2023; Gobierno de Argentina, 2023). Este dato no está disponible de forma directa para el territorio argentino, pero, tomando en cuenta que al año 2019 el 4.5% del PIB lo componía el sector minero, y que en 2023 el litio representó cerca del 21% de las exportaciones mineras totales, se puede estimar un aporte de entre el 0.5 y el 1% al PIB argentino (Gobierno de Argentina, 2023).

Sin embargo, la explotación del litio también tiene un gran impacto en el medio ambiente y en los recursos hídricos de la zona (Marconi et al., 2022), por lo que es necesario contar con información meteorológica e hidrológica de calidad y extensa data que permita la gestión estratégica de estos recursos a través de una mejor comprensión en los procesos y mecanismos que determinan su disponibilidad y variabilidad espaciotemporal.

En la zona denominada como el “Triángulo del litio” (ver **¡Error! No se encuentra el origen de la referencia.**), las aguas subterráneas son fundamentales para el consumo humano, la ganadería, el riego de cultivos, las especies endémicas y el propio desarrollo de la industria del litio. Esto realza la importancia de caracterizar el clima e hidrología superficial en la zona a modo de estimar de manera adecuada las tasas de recarga, los balances hídricos y, por otro lado, estudiar el grado de afectación de los ecosistemas. Por ejemplo, Marazuela et al. (2018) observaron una disminución de la tasa de evaporación después de las operaciones mineras en el Salar de Atacama, lo que se atribuye a una disminución del nivel freático y el agotamiento de la humedad del suelo debido, potencialmente, a las actividades mineras.

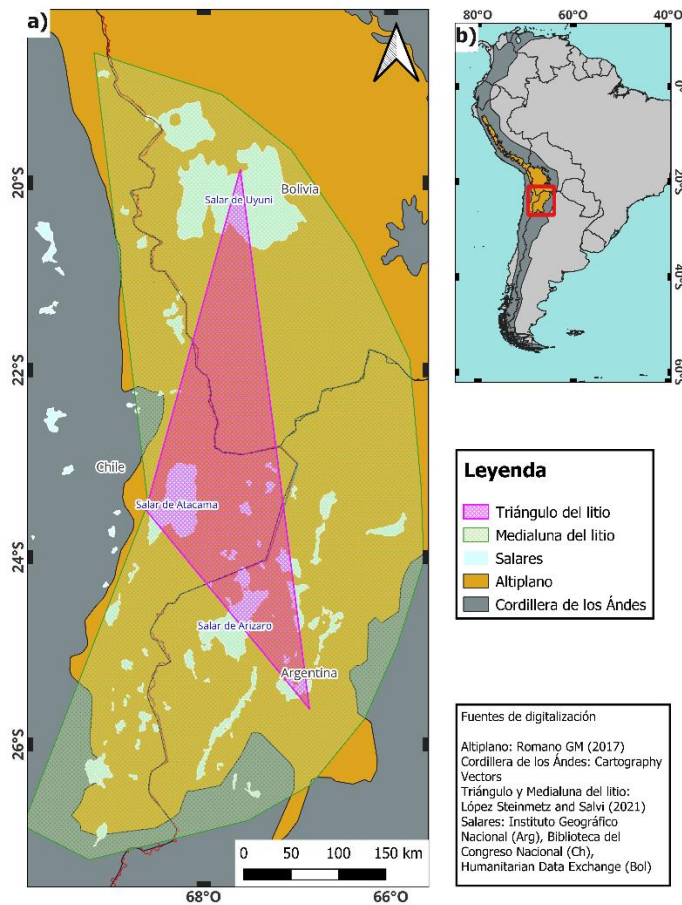


Figura 1.1: Mapas de la meseta Andina y los salares. En a) se muestra el Triángulo del Litio (rosado), la Medialuna del Litio (verde)¹ y todos los salares altiplánicos. En b), se observa la ubicación continental.

La limitada disponibilidad de registros meteorológicos en la Puna sugiere un desafío directo en la caracterización de los patrones espaciotemporales de precipitaciones y temperatura. La falta de datos precisos dificulta la implementación de modelos hidrológicos y puede resultar en estimaciones que no reflejan adecuadamente las características climáticas en resoluciones espaciotemporales finas. Asimismo, distribuir espacialmente variables meteorológicas basados únicamente en mediciones in situ y su relación con descriptores de posición (e.g., latitud, longitud y elevación) está sujeto a supuestos que dependen de la densidad de la red meteorológica y del grado de correlación espacial, ambos aspectos muy variables a escala mundial (Baez-Villanueva *et al.*, 2020; Beck *et al.*, 2017). No obstante, el creciente desarrollo de tecnologías de percepción remota abiertas al público, como los Productos de Precipitación y Temperatura basados en Satélites (SPPs y STPs por sus siglas en inglés; e.g., IMERG, MOD11), productos de reanálisis atmosféricos (e.g., MERRA, ERA5), productos estadísticos de base observacional (e.g., CR2MET, PISCO), surgen como una potente alternativa para obtener estimaciones globales (semi) continuas de estas variables climáticas (Zhang *et al.*, 2021).

¹ Propuesto por López Steinmetz & Salvi (2021)

De los productos globales anteriores, solo CHIRPS y ERA5 han sido puestos a prueba en las mesetas sudamericanas, específicamente en el altiplano (Birkel *et al.*, 2022a; Torres-Batló & Martí-Cardona, 2020). En ambos productos se destacan los resultados a escala mensual, con leves sesgos a la sobreestimación de precipitaciones. En el caso de ERA5, se presenta un correcto desempeño estimando temperaturas. Localmente, en la región altiplánica del Perú, los productos de precipitación diaria PISCO y RAIN4PE han mostrado buen desempeño en contextos similares a la Puna (Aybar *et al.*, 2020; Fernandez-Palomino *et al.*, 2022). Este último, caracteriza los patrones espaciotemporales de precipitación diaria a partir de la elaboración de un producto grillado de base observacional combinado con diferentes predictores (*e.g.*, descriptores fisiográficos, productos grillados no corregidos como ERA5, entre otros) mediante la implementación de métodos de *Random Forest*. Usando bases similares a RAIN4PE, se estimará la precipitación y temperatura diarias en la zona de la Puna Chile-Argentina.

1.2 Objetivos

La investigación tiene por objetivo estudiar oportunidades de mejora en la estimación de las precipitaciones y las temperaturas diarias en la Puna Chile-Argentina, mediante la elaboración de un producto grillado diario. A partir de dicho objetivo general, se definen los siguientes objetivos específicos:

- Sintetizar los registros *in-situ* de precipitación y temperatura en la zona de estudio, así como la información de los productos meteorológicos grillados disponibles y descriptores fisiográficos (*e.g.*, elevación, pendiente, orientación).
- Evaluar la calidad de las observaciones meteorológicas y aplicar filtros de estaciones según disponibilidad y coherencia en los registros.
- Elaborar un producto grillado de base observacional de precipitación y temperatura diaria para la zona de estudio.
- Cuantificar y comparar las estimaciones del producto desarrollado respecto a otros productos grillados disponibles.

1.3 Estructura del Documento

El documento se organiza en 6 capítulos. En el Capítulo **¡Error! No se encuentra el origen de la referencia.** se presenta la revisión bibliográfica del área de estudio y las metodologías empleadas para el desarrollo de la investigación; en el Capítulo 3 se aborda la descripción de la zona de estudio; en el Capítulo 4 se reúnen los datos disponibles para llevar a cabo el trabajo; en el Capítulo 5, se resume la metodología empleada en la investigación. Los resultados del estudio se incluyen en el Capítulo 6 y estos juntos a las oportunidades de desarrollo a futuro se discuten en el Capítulo 7. Finalmente, en el Capítulo 8, se resumen las principales conclusiones de la investigación.

2. MARCO TEÓRICO

Caracterizar espaciotemporalmente patrones climáticos globales apoyado en modelos numéricos y/o estadísticos requiere, a lo menos, comprender a nivel conceptual procesos y mecanismos asociados a la física atmosférica e interacciones atmósfera-superficie. Los modelos deben apuntar a capturar, ya sea de manera conceptual o explícita, estas interacciones para proporcionar estimaciones que permitan representar los patrones espaciotemporales observados para una determinada variable.

Existen diferentes enfoques para efecto de caracterizar los patrones de variabilidad climática en un determinado sector de interés. Entre ellos, se encuentra la utilización directa de estaciones meteorológicas, técnicas de interpolación espacial para efecto de distribuir observaciones puntuales (e.g., Inverse Distance Weighting; métodos de Kriging (Lucas et al., 2022)), productos derivados de imágenes satelitales mediante técnicas de percepción remota (e.g. TRMM, IMERG, PERSIANN-CDR (Ashouri et al., 2015; Hong et al., 2018; Huffman et al., 2010; C. Wang et al., 2018)), modelos estadísticos y de aprendizaje automático (*Machine Learning*) (e.g., RFMEP, RAIN4PE (Baez-Villanueva et al., 2020; Fernandez-Palomino et al., 2022)), reanálisis atmosféricos (e.g., ERA5 (Hersbach et al., 2020)), entre otros.

La estimación de precipitaciones y temperaturas presenta grandes desafíos cuando a ello se suma la escasez de información. Factores orográficos por la presencia de la Cordillera de los Andes afectan los patrones climáticos de la Puna Chile-Argentina, junto con la influencia del clima altiplánico y los fenómenos climáticos de gran escala que condicionan el tiempo y clima local, los cuales se ven acentuados por el cambio climático (Bennett et al., 2016a). Todo esto convierte a esta área en un lugar especialmente desafiante para la generación de productos climáticos de alta resolución.

Considerando las dificultades para estimar la precipitación en la zona, se debe desestimar el uso individual de las estaciones meteorológicas y cualquier técnica de interpolación espacial como un método de estimación confiable, debido a la baja densidad, gran variabilidad y sesgos presentes en estas en zonas montañosas como en el área de estudio (Woldemeskel et al., 2013). De hecho, en zonas con este tipo de topografía, la precipitación es típicamente subestimada en altas elevaciones porque la mayoría de las estaciones meteorológicas, por temas de costo y accesibilidad, se encuentran en zonas bajas (Derin & Yilmaz, 2014). Lo mismo ocurre con los enfoques satelitales, que, si bien son capaces de observar grandes áreas con una gran resolución, sus métodos son bastante susceptibles a sesgos sistemáticos (Laviola et al., 2013).

Los productos grillados globales y regionales se presentan como alternativas para estimar estas variables climáticas. Sin embargo, aun los productos basados en mediciones satelitales de mejor desempeño a nivel global reportan errores y sesgos en el dominio de los Andes (Erazo et al., 2018; Chavez & Takahashi, 2017; Baez-Villanueva et al., 2018).

Para abordar efectivamente la variabilidad climática en la Puna Chile-Argentina, es fundamental comprender los mecanismos que condicionan el clima local para evitar cualquier sobreajuste en el producto.

2.1 Clima en la Puna Chile-Argentina

La meseta Puna-Altiplano, ubicada en la zona elevada sobre los 3000 [m.s.n.m.] al norte de los 27°S (Jordan et al., 2010), es parte integral de un sistema de baja presión atmosférica a bajos niveles y un anticiclón en la alta atmósfera. Este sistema se desarrolla en verano y atrae aire húmedo proveniente de la Amazonia y del Atlántico, creando así pronunciados gradientes de precipitación (Strecker et al., 2006).

El clima regional se encuentra determinado por el efecto de enfriamiento por elevación combinado con efectos longitudinales y latitudinales. Desde el noreste llega la influencia húmeda de los trópicos, produciendo más precipitaciones, y desde el suroeste llega la influencia de la subsidencia subtropical del Alto Pacífico Sur, que genera, principalmente, fuertes condiciones áridas del desierto de Atacama (Bennett et al., 2016b). Esta tendencia refuerza las condiciones de aridez asociadas a la "Diagonal Árida" que atraviesa los Andes desde el desierto de Atacama hacia Argentina entre los 24°-25°S, intensificando la barrera biogeográfica existente para muchas especies vegetales y animales (Arroyo & Cavieres, 2013).

Se pueden identificar 3 distintos climas en la zona según sus montos de precipitación anual. En el flanco occidental del orógeno se encuentra el desierto de Atacama, con alrededor de 20 mm/año de lluvia y bajas tasas de erosión. Luego, las cuencas entre las montañas en el centro-este de la puna se tiene un clima de hiperaridez para un contexto de alta montaña, con menos de 200 mm/año. Finalmente, y en contraste de lo anterior, por efectos de los vientos húmedos del Monzón Sudamericano en el margen oriental de la meseta, se registran precipitaciones entre los 1000 y 3000 mm/año (Strecker et al., 2006).

El impacto de las fases positivas (+) y negativas (-) de los modos climáticos (i) Oscilación Decadal del Pacífico (PDO), (ii) El Niño-Oscilación del Sur (ENSO) con eventos El Niño (EN) y La Niña (LN), y (iii) Oscilación Antártica (AAO) fueron evaluados por Seiler et al. (2013). Allí, se reportó que la temperatura se eleva durante eventos de PDO(+), EN y AAO(+). Por otra parte, las precipitaciones y las crecidas aumentan con PDO(+), EN, y LN en las zonas de menor altura. Además, los autores concluyen que existe una tendencia al alza de las temperaturas en 0,1°C por década y un decaimiento de las precipitaciones desde el 1984 aproximadamente (-4% en DJF² y -10% en JJA) (Seiler et al., 2013). En la misma línea, Falvey & Garreaud (2009) mencionan que, en el sector chileno de la Puna, la tendencia al alza puede ser aún mayor, cercana a los 0,25°C por década.

En el siglo XXI, observando datos del *Ocean Niño Index* (ONI) de la Oficina Nacional de Administración Oceánica y Atmosférica (NOAA por sus siglas en inglés), se identifica que el fenómeno del ENSO se ha hecho presente con distintas intensidades, destacándose un evento de El Niño muy intenso (ONI sobre 2) entre mayo de 2014 y junio de 2016, y eventos fuertes de La Niña (ONI bajo -1,5) en dos periodos, el primero entre los meses de julio de 2007 y 2008 y el último entre junio de 2010 y julio de 2011. En la Figura 2.1 se observan los principales eventos del ENSO según la categorización del ONI desde 1950 a la actualidad.

² DJF hace referencia a los meses de diciembre, enero y febrero por sus siglas en inglés.

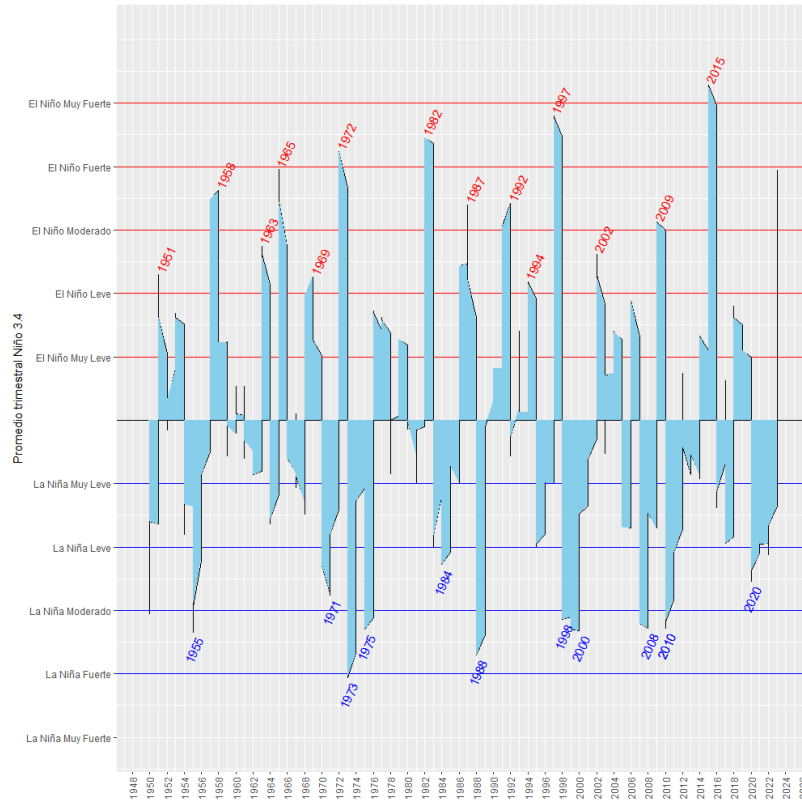


Figura 2.1: Ocean Niño Index (ONI) entre 1948 a la actualidad. Datos extraídos de Climate Prediction Center - ONI, 2023

Adicionalmente, autores como Chen et al. (2002) y Bennett et al. (2016b) concuerdan con estudios que sugieren que se está fortaleciendo la circulación atmosférica Hadley tropical y aumentando la convergencia en la baja troposfera y la divergencia en la alta troposfera, lo que está haciendo que los trópicos internos sean más húmedos y los trópicos externos más secos.

2.2 Productos Meteorológicos Grillados

En el estado del arte de los productos grillados meteorológicos, los conjuntos de datos globales de precipitación y temperatura destacan por su importancia y aplicabilidad en diversos campos de estudio, como la hidrología y la climatología. Entre los más relevantes, el producto de reanálisis ERA5 del European Centre for Medium-Range Weather Forecast (ECMWF) sobresale por su alta resolución temporal (horaria) y aceptable resolución espacial ($0,25^\circ \times 0,25^\circ$), proporcionando datos consistentes de temperatura y precipitación respecto a observaciones, además de otras variables climáticas. En la Tabla 2.1 se observan las características de los productos en mayor detalle.

A pesar de su utilidad, los productos grillados y de reanálisis globales pueden no capturar algunas componentes que definen la climatología en regiones específicas y complejas, como por

ejemplo, la Puna chileno-argentina. Del mismo modo, modelos regionales tampoco logran estimar con precisión las precipitaciones en la zona de estudio. Uno de los productos más usados en Chile es CR2MET, con cobertura en todo Chile Continental, pero sin mucha precisión a escala diaria en el norte grande chileno, además de no representar completamente el área delimitado de la Puna en el lado argentino (Baez-Villanueva et al., 2021). Por otro lado, CHIRPS y ERA5 funcionan relativamente bien en las alturas de la Cordillera de los Andes, CHIRPS con calificaciones muy buenas en sesgo y desviación estándar, y ERA5 con una Raíz del Error Cuadrado Medio (RMSE por sus siglas en inglés) de 13,92 y 1,655 para precipitación y temperatura respectivamente para la zona de La Quiaca, región incluida en el presente estudio (Birkel et al., 2022a; Torres-Batló & Martí-Cardona, 2020).

En el altiplano y zonas de altura de los Andes, las estimaciones de precipitación de IMERG presentan un KGE' mejor que otros productos ($KGE' = 0,79$) y es el método que entrega el mejor coeficiente de correlación de Pearson (R) ($R=0,89$) en la zona (Satgé et al., 2021). ERA5 por su parte, presenta tendencias consistentes al calentamiento de las temperaturas globales, y, a pesar de sobreestimar las precipitaciones, se concluye que es la mejor alternativa para lugares remotos con escasez de datos (Birkel et al., 2022b). Por su parte, CHELSA aún no ha sido probado en la zona de la Puna, lo cual presenta una oportunidad de prueba al estar validado en otras regiones de características similares (Karger et al., 2023).

Tabla 2.1: Productos grillados de interés en la zona de estudio y sus características.

Producto	Res. ⁵ espacial	Res. temporal	Periodo	Cobertura espacial	VARIABLES ⁶	Referencia
CHELSA ¹	0,0083°	Diaria	1979-2016	Global	PP/T	Karger et al. (2021)
CHIRPSv2 ²	0,05°	Diaria	1981-presente	50°N-50°S	PP	Peterson et al. (2015)
CR2MET ¹	0,05°	Diaria	1979-2020	Chile Continental	PP/T	Boisier et al. (2018)
RF-MEP ¹	0,05°	Diaria	2000-2018	Chile Continental	PP	Baez-Villanueva (2020)
IMERG ³	0,10°	30 min	2000-presente	Global	PP	Mahmoud et al. (2021)
MSWEPv2.2 ²	0,10°	3 horas	1979-presente	Global	PP	Beck et al. (2017a, 2019)
PISCOpv2.1 ³	0,10°	Diaria	1981-presente	Perú y Ecuador	PP	Aybar et al. (2020)
RAIN4PE ¹	0,10°	Diaria	1981-presente	Perú y Ecuador	PP	Fernández-Palomino et al. (2020)
CMORPHv1 ³	0,25°	30 min	1998-presente	60°N-60°S	PP	Xie et al. (2017)
ERA5 ⁴	0,25°	1 hora	1940-presente	Global	PP/T	(Hersbach et al., 2020)
TRMM 3B42v7 ³	0,25°	3 horas	1998-presente	50°N-50°S	PP	Huffman et al. (2010, 2007)
PERSIANN-CDR ³	0,25°	6 horas	1983-2017	60°N-60°S	PP	Ashouri et al. (2015)

1. Mediciones satelitales, observaciones, reanálisis y características topográficas. 2. Mediciones satelitales, observaciones y reanálisis. 3. Mediciones satelitales y observaciones. 4. Reanálisis global. 5. Res = Resolución. 6. PP y T hacen referencia a las variables climáticas de precipitación y temperatura respectivamente.

Además de CR2MET existen otros productos locales, como RF-MEP, orientados a estimar precipitaciones en el territorio chileno. Dicho producto combina mediciones satelitales, observaciones in situ, reanálisis y características topográficas del territorio mediante algoritmos de *Random Forest* (RF; Baez-Villanueva et al., 2020). Del mismo modo, el modelo RAIN4PE es el producto enfocado en la estimación de precipitaciones más actual en los Andes de Perú y Ecuador (Fernández-Palomino et al., 2022), el cual utiliza un método similar basado en RF. Este tipo de enfoque basado en RF, a diferencia de otros métodos basados en modelos lineales y/o herramientas geoestadísticas (Chávez & Takahashi, 2017), permite capturar dependencias e interacciones no lineales de variables de interés respecto a diferentes predictores. Los productos locales de RAIN4PE y PISCOv2.1 logran resultados de KGE' de 0,86 y 0,77 en las alturas del Lago Titicaca (3.819 m.s.n.m. en promedio) respectivamente.

Por su parte, RF-MEP fue aplicado en Chile continental en el periodo de 2000 a 2016, elaborando dos productos distintos: RF-MEP_{3P} y RF-MEP_{5P}, con tal de definir su configuración óptima en términos de rendimiento, parsimonia y cálculo computacional. La diferencia entre estos

productos radicaba en la cantidad de productos grillados de precipitación usados como entrada para la configuración del producto; mientras RF-MEP_{3P} utiliza PERSIANN-CDR, CHIRPSv2 y ERA-Interim, RF-MEP_{5P} agrega los productos TRMM 3B42v7 y CMORPHv1 a los ya mencionados (Baez-Villanueva et al., 2020). Para ambos productos, solo se considera la altitud como componente fisiográfica. El desempeño del producto ha motivado a otros investigadores a implementar dicha metodología en otras latitudes. Así, se destacan algunos estudios de evaluación de la metodología del producto RF-MEP_{3P} en China (Y. Shi et al., 2023), la cuenca del Río Nilo (Mohammed et al., 2023) y Korea del Sur (Nguyen et al., 2021), enfatizando este último por la elección de productos usada para la configuración del producto grillado local basado en RF-MEP: CHIRPS, PERSIANN-CCS, IMERG y ERA5.

2.3 Métodos para la Generación de Productos Grillados Meteorológicos

Los productos meteorológicos grillados constituyen una herramienta analítica útil en la climatología y la meteorología, pues facilitan la representación espacial de variables como la temperatura, precipitación, presión atmosférica, entre otros. Asimismo, proporcionan una representación estructurada de variables climáticas para la modelación de sistemas meteorológicos complejos (Newman et al., 2019).

Existen distintas técnicas y metodologías para la elaboración de productos grillados meteorológicos, las cuales se pueden combinar para lograr resultados más refinados y precisos. Entre ellas destacan la Interpolación por Distancia Inversa Ponderada (IDW por sus siglas en inglés), el *Kriging*, métodos basados en modelos numéricos, el aprendizaje automático (*Machine Learning*) y el reanálisis, que integra datos históricos y actuales para reconstruir condiciones pasadas. A continuación, se presenta una breve revisión de los enfoques metodológicos enunciados anteriormente.

2.3.1 Metodologías convencionales de observación

Una de las técnicas más utilizadas para estimar la distribución espacial y temporal de las variables climáticas son los métodos de interpolación, los que utilizan de manera directa las observaciones in situ de estaciones meteorológicas (Haylock et al., 2008), distribuyendo bajo supuestos de correlación espaciotemporal las variables. Sin embargo, la posibilidad de una baja densidad de mediciones y una limitada extensión geográfica plantean un potencial problema para obtener una adecuada representación espacial (Xie & Arkin, 1996).

De estos, los más destacados son el IDW y el método de *Kriging*. El último, cuantifica la estructura espacial de los datos mediante el uso de variogramas. Este supone que los datos más cercanos a un punto conocido tienen mayor peso o influencia sobre la interpolación, la cual disminuye en medida que se aleja del punto de interés. La medición de la probabilidad, efectuada por el método de *Kriging*, hace la diferencia con respecto a los métodos determinísticos para

interpolaciones espaciales (Burrough & McDonnell, 1998). En ambos modelos se considera que las predicciones son una combinación lineal de los datos (Li et al., 2018)

La configuración de modelos hidrológicos son una buena parte de la aplicación de estos métodos. Una prueba de esto fue la investigación de Liu et al. (2018), realizando simulaciones de precipitación por medio de métodos estadísticos IDW, *Nearest Neighbor* (NN) y *Linear Spline* (LN), y geoestadísticos como el *Kriging*, verificando a través de un modelo hidrológico derivado del modelo GR4H *Hourly Rainfall Runoff Model*. El estudio entrega resultados favorables para el índice *Nash-Sutcliffe efficiency* (NSE) en interpolaciones de tipo IDW y sesgos favorables para simulaciones con métodos de *Kriging*.

2.3.2 Productos de base satelital y percepción remota

En las últimas décadas, el avance tecnológico ha permitido el desarrollo de técnicas de percepción remota, potenciando las estimaciones de variables climáticas. El progreso en la tecnología satelital y el desarrollo de métodos avanzados, como drones y radares, han permitido un análisis más profundo y extenso del clima a escalas locales y globales.

Por lo anterior, se incorpora el uso de tecnologías de observación remota, como satélites, radares o drones para abarcar mayores áreas de observación. No obstante, distintas investigaciones concluyen que las metodologías satelitales no son precisas en los montos de precipitación por sí solas (Fleming et al., 2011) cuando se validan con observaciones in situ. Así, actualmente se han adoptado técnicas que combinan observaciones in situ y productos satelitales para dar origen a una amplia variedad de productos grillados, en los que destacan IMERG (C. Wang et al., 2018), TRMM 3B42v7 (Huffman et al., 2010), PERSIANN-CDR (Ashouri et al., 2015), entre otros (ver Tabla 2.1).

Estos productos han servido de entrada para modelos hidrológicos. Wang et al. (2017) realizó simulaciones hidrológicas con los productos IMERG y TRMM 3B42v7 en un modelo Variable Infiltration Capacity (VIC), logrando rendimientos notables del modelo en ambos, con resultados de 0,742 y 0,696 para NSE, y 8,35 y 10% de sesgo respectivamente.

Cabe destacar que la comparación en el desempeño de los métodos geoestadísticos y de base satelital depende del objetivo del estudio y las características fisiográficas e hidroclimáticas que definen la zona de interés. En general, si existe una alta densidad de observaciones, los métodos de interpolación son más precisos en magnitud; en caso contrario, se destacan los métodos que combinan observaciones en terreno y productos satelitales (Fleming et al., 2011; Shi et al., 2022).

2.3.3 Reanálisis climático

Un reanálisis climático describe numéricamente el clima mediante la combinación de modelos climáticos con observaciones combinados mediante técnicas de asimilación de datos. Contiene estimaciones de variables atmosféricas como la temperatura del aire, la presión, el viento

a diferentes altitudes y parámetros de superficie, como las precipitaciones, el contenido de humedad del suelo, la altura de las olas oceánicas y la temperatura de la superficie del mar. Las estimaciones se elaboran para todos los lugares de la Tierra y abarcan un largo periodo de tiempo que puede remontarse a varias décadas (Climate Reanalysis | ECMWF, 2023.).

Productos como ERA5 del ECMWF se basan en complejos modelos de circulación general de base física y una integración de datos a lo largo de extensos periodos temporales, produciendo conjuntos de datos consistentes y de alta calidad que son fundamentales para la investigación climática y las aplicaciones operativas (Hersbach et al., 2020).

Este producto ha servido como entrada de muchos modelos hidrológicos. En particular, Sun et al. (2021) evaluó el rendimiento del modelo VIC con ERA5 como entrada de precipitación para simular caudales en cuencas de alta montaña de Asia, con una topografía similar al área de estudio. Según el autor, ERA5 sobreestima las precipitaciones en la zona, sesgando las mediciones de caudal en las cuencas, y concluyendo en la necesidad de corregir los sesgos propios de ERA5 antes de incorporarlo a un modelo hidrológico en cuencas de alta montaña.

2.3.4 Aprendizaje Automático (Machine Learning – ML)

El aprendizaje automático (*Machine Learning* (ML) en inglés) es el estudio científico de algoritmos y modelos estadísticos que los sistemas informáticos utilizan para realizar una tarea sin estar explícitamente programados para ello. La principal ventaja de utilizar ML es que, una vez que un algoritmo aprende qué hacer con los datos, este puede realizar su trabajo automáticamente (Mahesh, 2018).

En términos de la representación de variables climáticas como la precipitación y temperatura, diversos algoritmos se han posicionado como herramientas útiles para dichos fines. Entre ellos se encuentran:

- **Regresión Lineal y Logística:** La regresión lineal busca establecer una relación lineal entre variables continuas. La regresión logística, por otro lado, se emplea para la clasificación, prediciendo la probabilidad de que un evento pertenezca a una categoría. Esta metodología se caracteriza por su rapidez y facilidad de interpretación, siendo ideal para establecer relaciones entre variables meteorológicas. El algoritmo considera una relación lineal entre las variables, sin reconocer otro tipo de relaciones, siendo una de las principales limitaciones de este algoritmo.
- **Máquinas de Vectores de Soporte (SVM):** Algoritmo de aprendizaje supervisado que encuentra un límite de decisión óptimo entre las clases, incluso en conjuntos de datos complejos, utilizando un enfoque basado en la maximización del margen entre clases. Ideal para trabajar con datos no lineales, pero puede ser ineficiente para grandes conjuntos de datos y complejo de interpretar.
- **Redes Neuronales Artificiales (ANN):** Inspiradas en la estructura del cerebro humano, las redes neuronales son sistemas interconectados de nodos que aprenden patrones complejos y relaciones no lineales en los datos. Se adapta bien a grandes cantidades de datos y la

modelación de relaciones complejas entre variables. Así mismo, las redes son propensas al sobreajuste y a requerir grandes esfuerzos computacionales.

- **Árboles de Decisión:** Basado en la estructura de un árbol, representa decisiones y reglas. En cada nodo del árbol, se toma una decisión binaria basada en una característica particular, dividiendo el conjunto de datos en ramas que conducen a decisiones finales. Este algoritmo es útil para la selección de características y el manejo de datos no lineales, pero presentan tendencia al sobreajuste, no siendo aptos para extrapolar estimaciones.

Los árboles de decisión son la base fundamental del algoritmo de Random Forest, el cual combina múltiples árboles de decisión para crear un “bosque”, donde cada árbol se entrena con un subconjunto de datos y características.

2.3.5 Random Forest

Random Forest es un método de *Machine Learning*, en el cual se combinan predictores de árboles tal que cada árbol depende de los valores de un vector aleatorio muestreado independientemente y con la misma distribución para todos los árboles del bosque (Breiman, 2001). Cada árbol se entrena con una muestra aleatoria de datos y un subconjunto aleatorio de características con el propósito de reducir al máximo los sesgos en la calibración.

La Figura 2.2 muestra la metodología de *Random Forest*. Primero, el conjunto de datos de entrenamiento (*Dataset*) crea N árboles de decisión; cada uno de estos produce un resultado basado en una muestra del conjunto de datos; después, estos resultados individuales son combinados a través de una decisión por mayoría o promedio para obtener un resultado final.

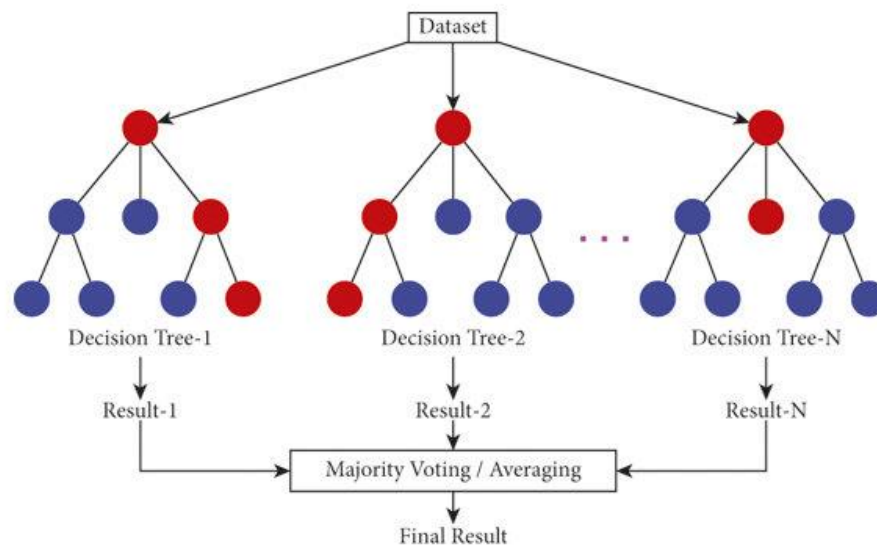


Figura 2.2: Diagrama base de Random Forest. Fuente: Khan et al., 2021.

Una de las principales ventajas de *Random Forest* radica en su capacidad para mitigar el sobreajuste y la varianza inherente a los árboles de decisión individuales, al tiempo que mantiene un alto rendimiento predictivo. Además, es capaz de manejar conjuntos de datos grandes con múltiples características, identificar la importancia de las variables y proporcionar estimaciones de la relevancia de las características para las predicciones, lo que lo convierte en una herramienta valiosa para la generación de productos meteorológicos precisos y confiables (Roy & Larocque, 2012).

2.4 Enfoques Estocásticos

En el campo de la meteorología y climatología, los enfoques estocásticos se han vuelto indispensables ante la no posibilidad de resolver todos los procesos y escalas necesarias en modelos numéricos completos de predicción meteorológica y climática. Los modelos estocásticos proporcionan una distribución de probabilidades que ofrece una visión más completa y matizada de los posibles valores que puede adoptar una determinada variable y su incertidumbre. La incorporación de este enfoque puede mejorar significativamente la simulación y la estimación de incertidumbre en pronósticos meteorológicos y climáticos a corto plazo. (Franzke et al., 2015).

Por otro lado, Yürekli et al. (2007) aplica un enfoque estocástico con técnicas basadas en el modelo *Autoregressive Integrated Moving Average* (ARIMA) para generar estimaciones de temperatura, radiación solar y humedad relativa, obteniendo coeficientes de determinación sobre los 0,9 en las dos primeras variables y cercano a los 0,6 en humedad relativa, verificando el uso de modelos ARIMA para la simulación de estas variables climáticas.

Además de su aplicación en la modelación y predicción de las variables mencionadas, los enfoques estocásticos han sido fundamentales en el desarrollo de productos grillados de precipitación y temperatura, como el producto de reanálisis ERA5.

Sin embargo, existe otra clase de metodologías estadísticas para cuantificar la incertidumbre de estimaciones determinísticas. Uno de estos métodos propuestos por Sharifi et al. (2019) analiza las incertidumbres de productos de precipitación satelital basado en un enfoque de dependencia entre variables con diferentes tipos de distribuciones, mediante teoría de cópulas. Otro de estos enfoques corresponde al algoritmo *Bluecat* (Koutsoyiannis & Montanari, 2022a), el cual no requiere de un preprocesamiento de datos, siendo posible integrar modelos determinísticos sencillos como entrada (Jorquera & Pizarro, 2023). Además, la aplicación del método *Bluecat* no implica mayores costos computacionales, lo cual facilita su implementación.

2.4.1 Algoritmo Bluecat

Bluecat (Brisk local uncertainty estimator for generic simulations and predictions; Koutsoyiannis & Montanari, 2022a) es una metodología propuesta para extender un modelo determinístico (modelo D) a uno estocástico (modelo S). Se basa en la hipótesis de que la información contenida en una serie temporal de observaciones y las predicciones concurrentes

realizadas por el modelo D son suficientes para apoyar esta extensión. El paso a un modelo S desde un modelo D ofrece dos funciones importantes:

- Corrige apropiadamente el sesgo del modelo D con respecto a las observaciones.
- Representa la incertidumbre del modelo en base a bandas de confianza.

El algoritmo destaca por su simplicidad, permitiendo su fácil implementación dado el bajo costo computacional. Además, *Bluecat* tiene la capacidad de inferir la distribución de probabilidad de la predicción sin exigir hipótesis sobre las propiedades estadísticas del error de predicción (e.g., normalidad, homocedasticidad).

Inicialmente, *Bluecat* se propuso como una herramienta para complementar la modelación hidrológica de caudales, sin embargo, estudios más recientes indican que la metodología puede extenderse a otras aplicaciones. Ejemplo de ello son las aplicaciones a variables climáticas, como la precipitación, mejorando la variabilidad en comparación al modelo D, y la temperatura media, corrigiendo tendencias sesgadas al aumento. (Koutsoyiannis & Montanari, 2022b)-

2.5 Síntesis de la Revisión e Hipótesis de Investigación

A modo de resumir los aspectos más importantes de la revisión bibliográfica, se menciona:

- La Puna Chile-Argentina cuenta con una amplia variedad de climas, donde destacan la influencia de diferentes modos de variabilidad climática de gran escala.
- Las condiciones climáticas, topográficas y la falta de observaciones meteorológicas dificultan el desempeño de los productos grillados a nivel global y local.
- La literatura muestra que se pueden obtener mejores resultados combinando las distintas metodologías para la predicción de variables climáticas respecto a la utilización de un único enfoque. En particular, se destaca *Random Forest* dada su habilidad en caracterizar la distribución y temporalidad de las precipitaciones en Los Andes de Ecuador, Perú y Chile.
- Existe una oportunidad de innovación y mejora mediante la integración de la incertidumbre a un producto grillado meteorológico determinístico, convirtiéndolo en uno estocástico. En esa línea, el algoritmo *Bluecat* se posiciona como una potente alternativa dada su sencillez de implementación y bajo costo computacional.

A partir de los antecedentes recopilados en la revisión de literatura, se plantea como hipótesis que un producto de base observacional elaborado en base a técnicas de *Random Forest* tenga un mejor rendimiento que los productos globales disponibles en la zona de interés, en particular, los productos CHIRPSv2, PERSIANN-CDR, ERA5, IMERG y la base metodológica de RFMEP aplicada en la zona. Además, se plantea como hipótesis que la implementación del algoritmo *Bluecat* permitirá mejorar el producto determinístico a través de la generación de un producto estocástico de las variables meteorológicas de interés.

3. ZONA DE ESTUDIO

3.1 Antecedentes generales

La Puna chileno-argentina es una región de alta montaña con un clima árido y extremo (Sarricolea et al. 2023) perteneciente a la meseta Altiplano Puna de los Andes Centrales, la segunda con mayor altura y extensión después del Tíbet (Figura 3.1, Allmendinger et al., 1997). La Puna comprende las regiones sobre los 3.000 [m.s.n.m.]. al sur del 22°S, extendida desde el sur de Bolivia hasta el norte de Chile y Argentina hasta los 27°S aproximadamente (zonas de color azul en la Figura 3.1).

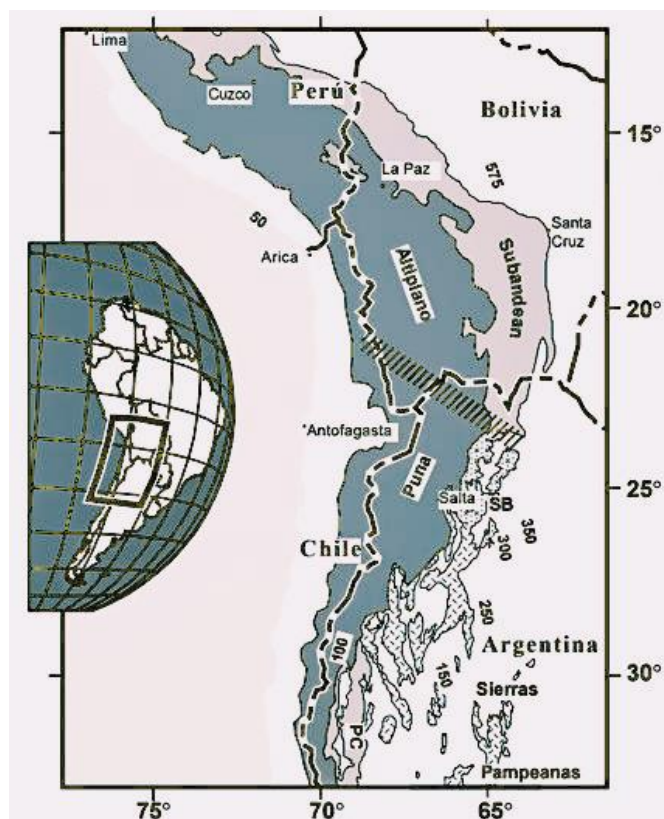


Figura 3.1: Mapa de la extensión de la alta meseta de los Andes Centrales. Fuente: Allmendinger et al., 1997

Al este del dominio (a la izquierda de la zona azul en la Figura 3.1), se observan cinturones tectónicos de piel fina (*thin-skinned thrust belts*) en las cordilleras sub-andinas de Bolivia y la precordillera en Argentina. Además, se observa una zona achurada con tendencia NW-SE en las latitudes 21°S y 23°S, cerca del límite entre Bolivia, Argentina y Chile, representando un límite geológico entre el Altiplano y la Puna. Estas áreas poseen drenajes internos similares, pero presentan diferencias importantes entre la corteza, la litósfera, la astenosfera, la topografía y la historia magmática (Whitman et al., 1996; S. L. Beck & Zandt, 2002; Allmendinger et al., 1997; Trumbull et al., 2006).

Como zona de estudio, se selecciona un dominio mayor al identificado geográficamente como Puna, con el fin de mejorar la caracterización regional del clima de la zona, incluir observaciones adyacentes a la zona de interés y evitar problemas en la representación de los bordes del dominio. Asimismo, se adopta dicha decisión metodológica a modo de complementar la base de datos de observación e incluir registros adicionales. En consecuencia, el dominio de estudio queda definido por el polígono que se muestra en la Figura 3.2, delimitado por las latitudes 21°S y 27°S y las longitudes 64°W y 69.5°W.

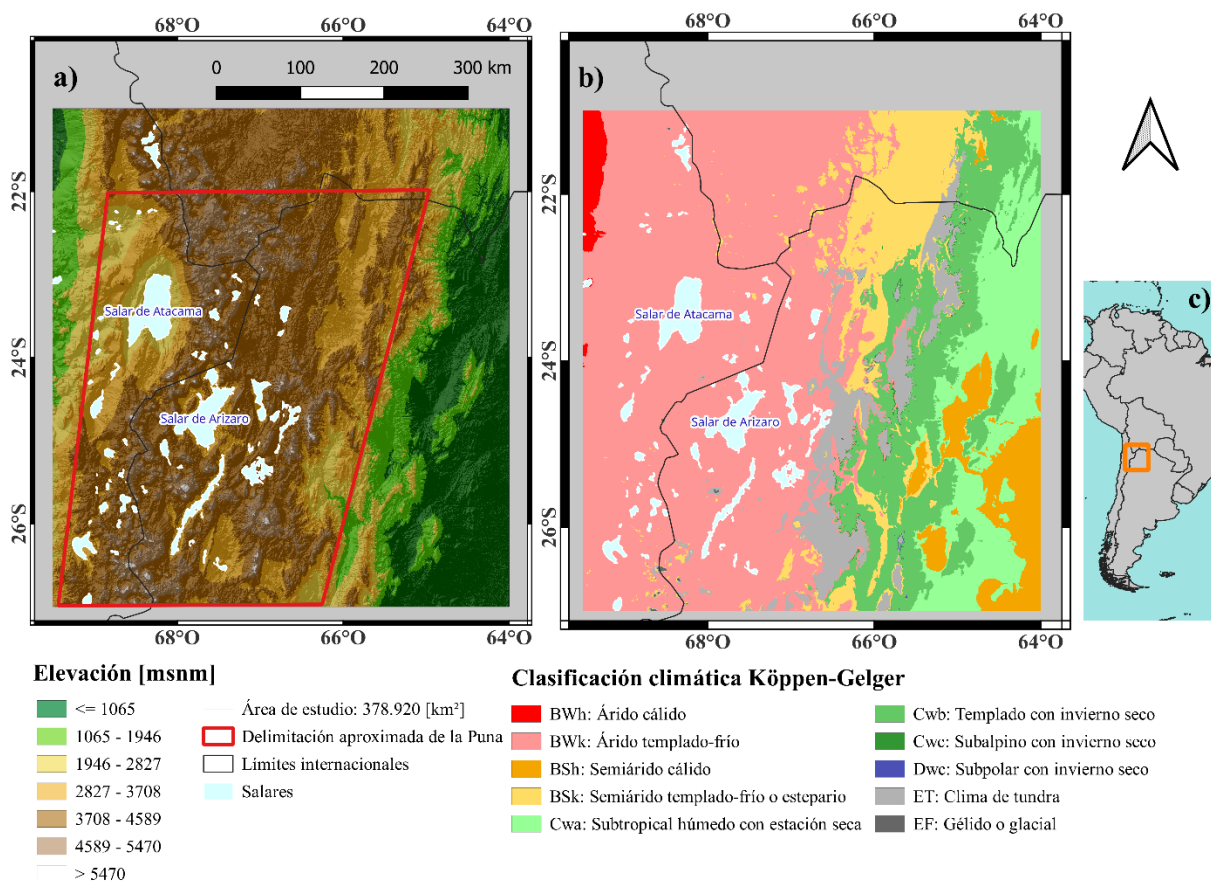


Figura 3.2: Zona de estudio con sus principales características morfológicas y climáticas. (a) Modelo de elevación digital SRTM; (b) Clasificación climática de Köppen-Geiger; (c) Vista general de la zona de estudio respecto a Sudamérica.

El área geográfica de interés involucra en términos político-administrativos a Chile, Bolivia y Argentina. En Chile se considera toda la cordillera y precordillera en la región de Antofagasta, y parcialmente en las regiones de Tarapacá y Atacama; en Bolivia, mayoritariamente los departamentos de Potosí, Tarija y parcialmente Chuquisaca; y en Argentina la totalidad de la provincia de Jujuy, la mayoría de Salta, Catamarca y Tucumán, y parcialmente Santiago del Estero.

El primer mapa de la Figura 3.2 se realizó utilizando los DEM del proyecto internacional *Shuttle Radar Topography Mission* (SRTM). Mientras que la clasificación climática del segundo

mapa se obtuvo con el archivo ráster de alta resolución (1 km) publicado por (H. E. Beck et al., 2023).

3.2 Caracterización Geomorfológica

La Puna presenta una elevación media aproximada de 4400 m.s.n.m., junto con una actividad volcánica plio-pleistocénica bastante extendida, sumado a una corteza que varía localmente entre los 40 y 70 km de espesor, y un alto grado de atenuación de las ondas sísmicas (Q) por debajo del nivel del mar, que sugiere que la litósfera del manto es delgada o inexistente (Whitman et al., 1996; Mcglashan et al., 2008). Otras características destacables de la Puna son su escasa vegetación y la presencia de salares y lagunas de origen volcánico (Allmendinger et al., 1997).

El área de estudio abarca más allá de la puna, tal como se observa en la Figura 3.2. Esta delimitación tiene una superficie de 378.820 [km²] y una elevación media de 3.069 [m.s.n.m.], menor a la elevación media de la puna de 3.905 [m.s.n.m.] por incluir zonas aledañas de menor altura, incorporando pixeles de alturas medias hasta 184 [m.s.n.m.]. El resto de las características geomorfológicas se encuentran en la Tabla 3.1.

Tabla 3.1: Características geomorfológicas del área de estudio.

Parámetro	Unidad	Valor
Área	km ²	378.820
Elevación mínima	m.s.n.m.	184
Elevación máxima	m.s.n.m.	6.351
Elevación media	m.s.n.m.	3.069
Rango de elevación	m	6.167

También, se realiza un análisis geomorfológico a partir de las cuencas hidrográficas de la zona de estudio. A través del Ministerio de Economía y Finanzas Públicas de Argentina (2024), el Ministerio de Desarrollo Sostenible (1994) y el Ministerio de Obras Públicas de Chile (2024), se recopilaron un total de 89 cuencas en el área, distribuidas como se observa en la Figura 3.3.

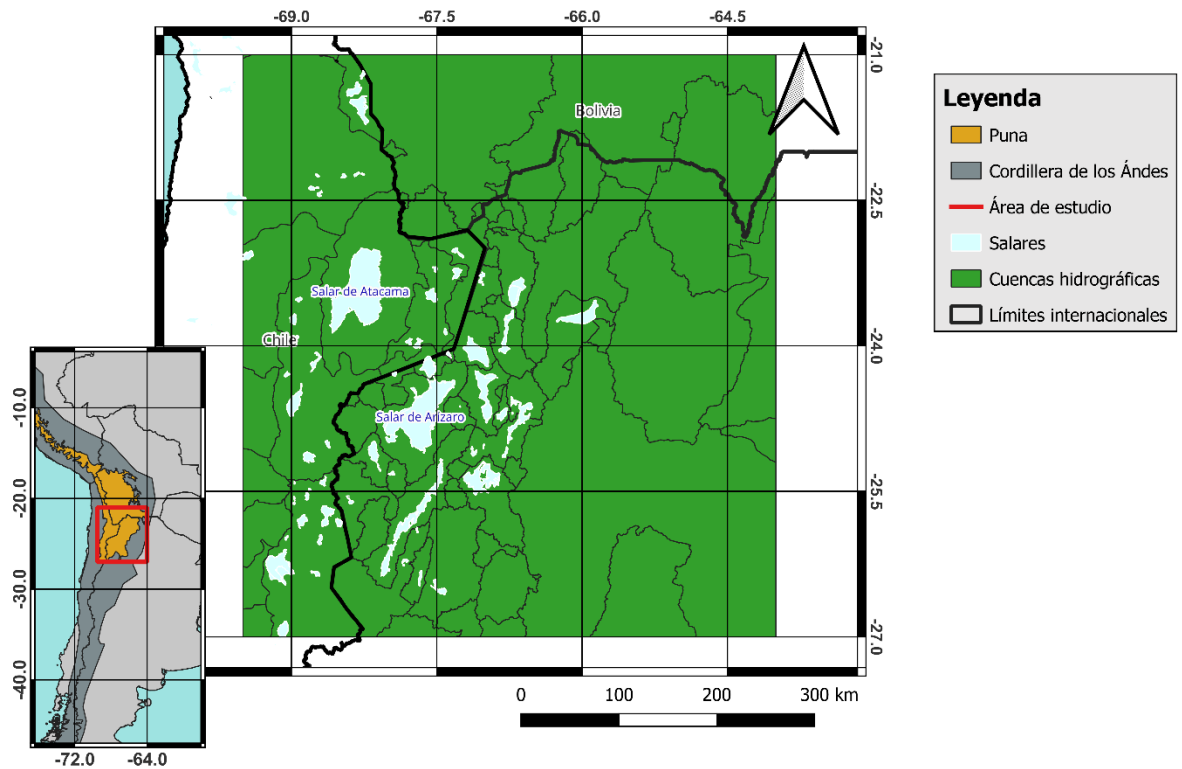


Figura 3.3: Cuencas hidrográficas en el área de estudio.

A partir de esta distribución, se pueden reconocer cuencas hidrográficas de distintas áreas y elevaciones, así como se distinguen algunas más relevantes en el análisis de los cuerpos salinos, como las cuencas que contienen los salares de Atacama y Arizaro en Chile y Argentina respectivamente. De esta manera, se grafican las curvas hipsométricas de todas las cuencas en la Figura 3.4, segmentando por países y con especial énfasis en las cuencas del Salar de Atacama y Arizaro.

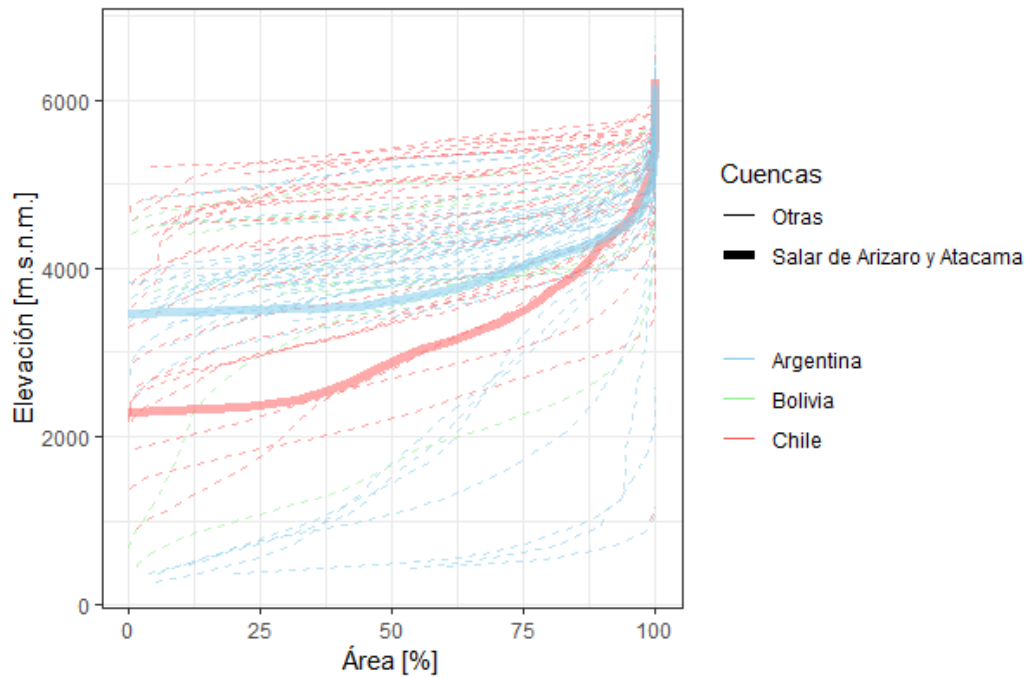


Figura 3.4: Curvas hipsométricas de las cuencas hidrográficas en el área de estudio.

Por último, se analiza las curvas de distribución de probabilidad para distintas variables de interés de las cuencas, como la elevación [m.s.n.m.], el área [km²], la pendiente [m km⁻¹] y la compartidad [m]. Estas curvas se muestran en la Figura 3.5, separando en dos gráficos por temas de escala y visualización, usando una escala logarítmica para las variables de elevación y área por ser de magnitudes mayores, y una escala lineal para el resto de las características.

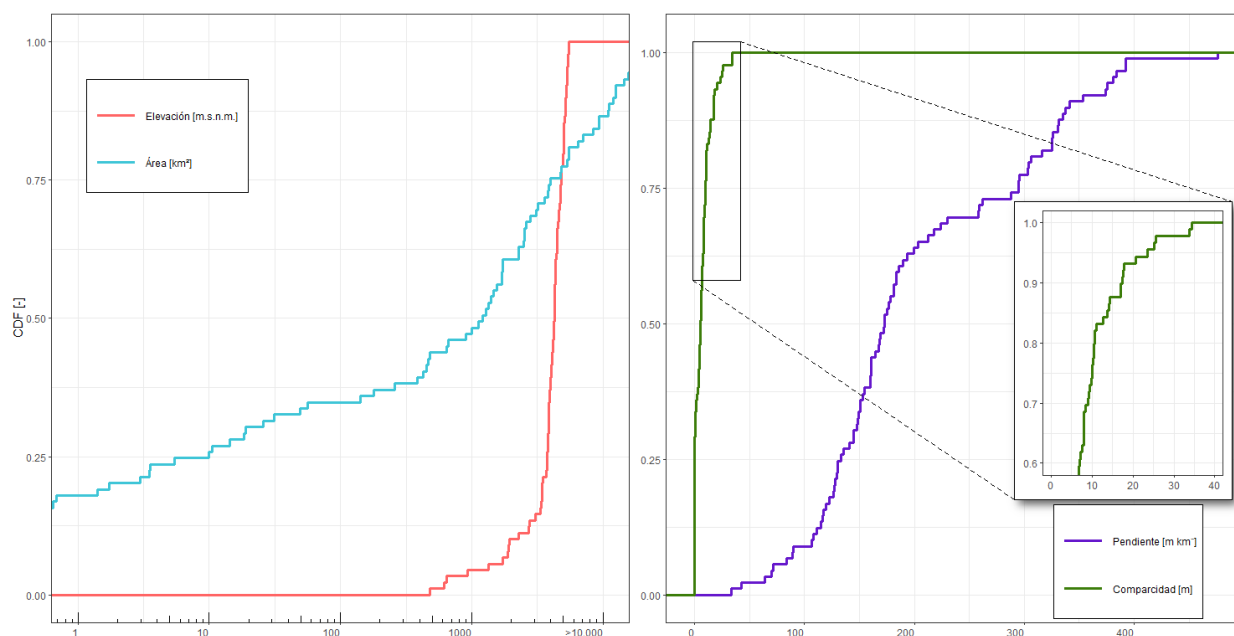


Figura 3.5: Curvas de distribución de probabilidad para los atributos de elevación, área, pendiente y comparidad de las cuencas hidrográficas del área de estudio.

3.3 Hidroclimatología

Como se indica en la sección 2.1, en el área de estudio se desarrollan a los menos tres tipos de zonas climáticas: el desierto chileno, entre los 68 y 69°W y los 22 a 25°S, con precipitaciones menores a 20 [mm] al año; la zona más elevada de la puna, en toda la latitud estudiada y los 66 a 68°W, con precipitaciones hasta los 200 [mm] anuales; y el margen oriental de la meseta, con un gradiente importante de elevaciones (rango entre los 184 y 3.000 [m.s.n.m.]), en el área restante del polígono estudiado, alcanzando magnitudes de precipitación sobre los 1.000 [mm] y hasta los 3.000 [mm] al año.

Con la clasificación climática Köppen-Geiger en la Figura 3.2.b, se puede observar que predomina un clima árido templado-frío (BWk), localizado entre los 66,5 y 69°W a lo largo de toda la latitud. Esto se contradice con lo mencionado anteriormente, puesto que engloba en una misma zona climática a dos sectores con precipitaciones medias anuales de magnitudes muy distintas.

El resto de la zona de estudio presenta un clima árido cálido (BWh) al noroeste; un clima semiárido templado-frío (BSk) en la zona de mayor pendiente al este de la cordillera (alrededor de los 66°W); clima de tundra (ET), Subalpino con invierno seco (Cwc), subtropical húmedo con estación seca (Cwa) y semiárido cálido (BSh) al este de los 66°W, donde se esperan las mayores precipitaciones y temperaturas del área de estudio.

A modo de ilustrar los patrones de variabilidad latitudinal y longitudinal de la precipitación anual promedio y de temperatura media promedio anual en el periodo 2007-2017, se presentan las Figura 3.6 y Figura 3.7 que, en forma complementaria, relacionan estas variables climáticas con la elevación de la zona de estudio. Estos gráficos fueron elaborados con 89 estaciones meteorológicas

distribuidas en una superficie de 378.820 [km²] (i.e., >> 1 estación/km²), lo que dificulta el análisis de los patrones espaciales de las variables meteorológicas de interés.

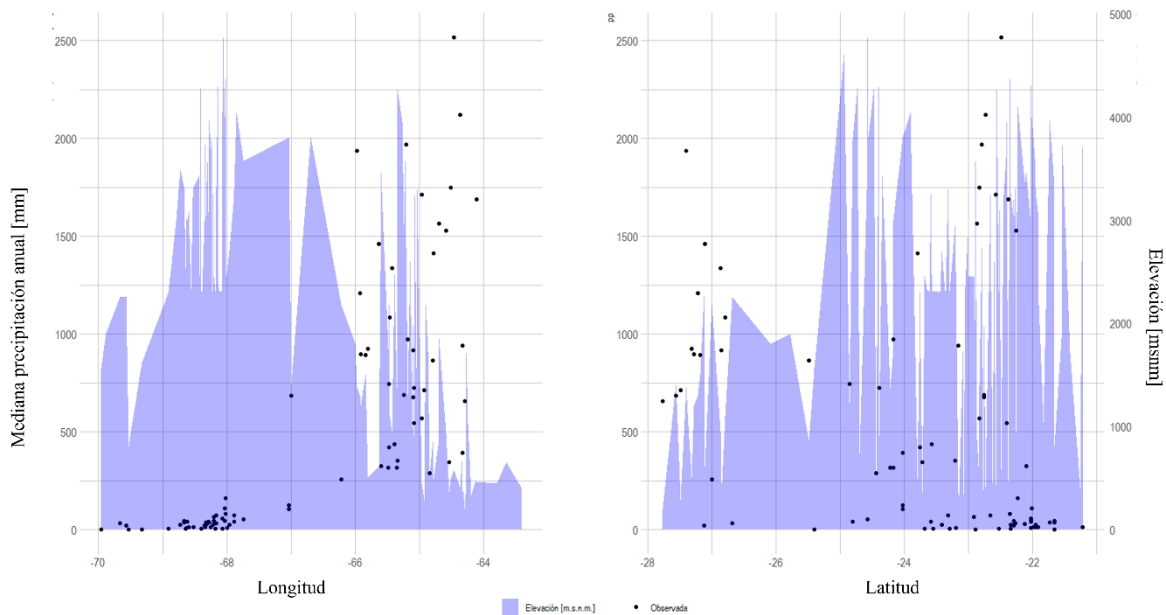


Figura 3.6: Perfiles de los valores medianos de precipitación anual según longitud, latitud y elevación.

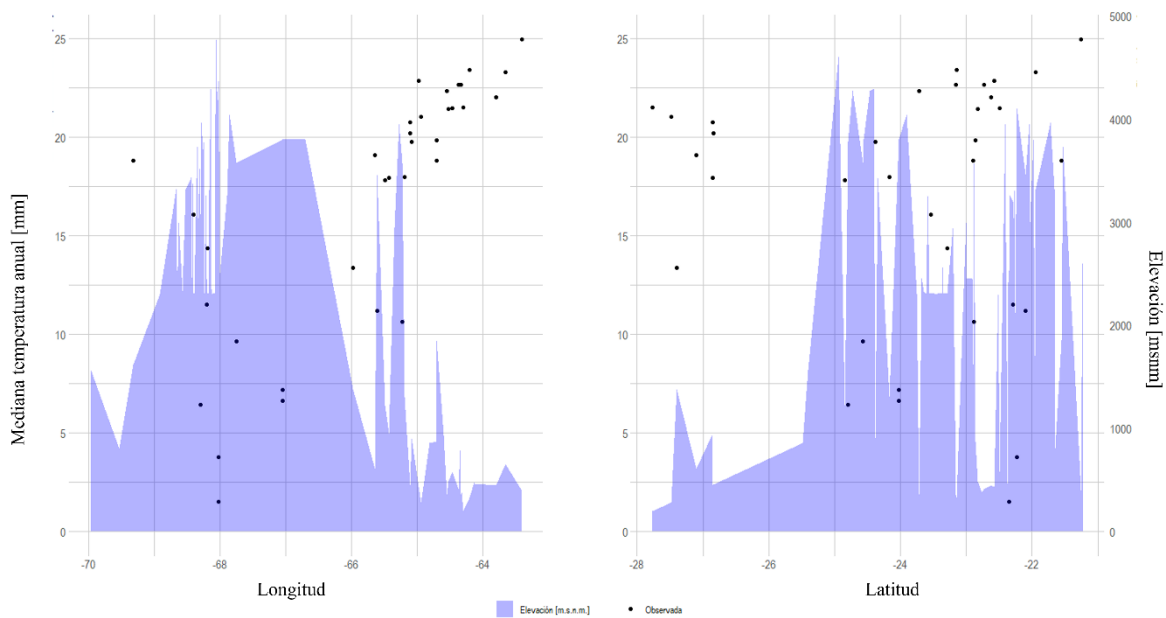


Figura 3.7: Perfiles de los valores medianos de temperatura media anual según longitud, latitud y elevación.

De la Figura 3.6Figura **3.6** se desprende el notorio gradiente de precipitación de oeste a este, relacionado con la topografía del lugar; además, se aprecia una diferencia en las magnitudes de precipitación en el tramo 68-69°W con el tramo 66-68°W, en dirección a lo indicado en la clasificación climática de Köppen-Geiger.

Con respecto a la temperatura en la Figura 3.7, también se observa una marcada tendencia con la longitud entre los -66 y 64°W por causa de la presencia de la Cordillera de los Andes, llegando a valores muy cercanos a 0 a grandes elevaciones (>4000 [m.s.n.m]).

4. DATOS DISPONIBLES

Las variables climáticas de precipitación y temperatura son parámetros que definen en buena medida el clima de una región (López Jiménez et al., 2019). Por este motivo, es importante hacer una revisión de los datos disponibles de precipitación y temperatura para abordar la presente investigación, considerando la cantidad y la calidad de estos.

En esta sección, se analizará la información disponible in situ y en internet, principalmente de productos grillados de variables climáticas como precipitación y temperatura.

4.1 Información Meteorológica

Una de las características de la zona de estudio es la baja densidad de estaciones meteorológicas, las cuales no logran representar los patrones climáticos locales. En la zona se identifican 89 estaciones meteorológicas de precipitación y temperatura (ver listado completo en ANEXO B: REGISTRO DE ESTACIONES, separadas por la cordillera, largas distancias y distintas elevaciones. Sin embargo, según la Organización Meteorológica Mundial (WMO por sus siglas en inglés), las estaciones se pueden dividir en grupos de acción para áreas de topografía quebrada o condiciones montañosas. De acuerdo con esta clasificación, la Tabla 4.1 indica que en una red meteorológica básica existen tres grupos de estaciones para las observaciones meteorológicas.

Tabla 4.1: Distribución de estaciones meteorológicas en una red básica. Datos extraídos de López (2009).

GRUPO	TIPO ESTACIÓN	DISTANCIA ENTRE ESTACIONES [KM]	RADIO ACCIÓN PROMEDIO [km]	SUGERENCIAS TÉCNICAS OMM
1-R	REFERENCIA	300 - 400	150 - 200	Una por cada 4 o 5 del primer grupo
RED BÁSICA	1	SP - CP- AM	150 - 200	Aprox. Una por cada 10 estaciones del grupo 2
	2	CO - ME	50 - 60	Aprox. 10 estaciones por cada una del primer grupo
	3	PM - PG	25 - 30	La mitad de la distancia en áreas montañosas

Las estaciones Agrometeorológicas (AM), Climatológica Principal (CP) y Sinóptica Principal (SP), forman parte del primer grupo en una red básica. Luego, en el segundo grupo, se ubican las estaciones Climatológicas Ordinarias (CO) y las Meteorológicas (ME). Por último, se encuentran las estaciones Pluviométricas (PM) y Pluviográficas (PG), en las cuales solamente se hace la observación de fenómenos atmosféricos relacionados con la precipitación.

En la Figura 4.1 se realiza un análisis de distribución de estaciones según distancia, considerando los radios de acción de la Tabla 4.1.

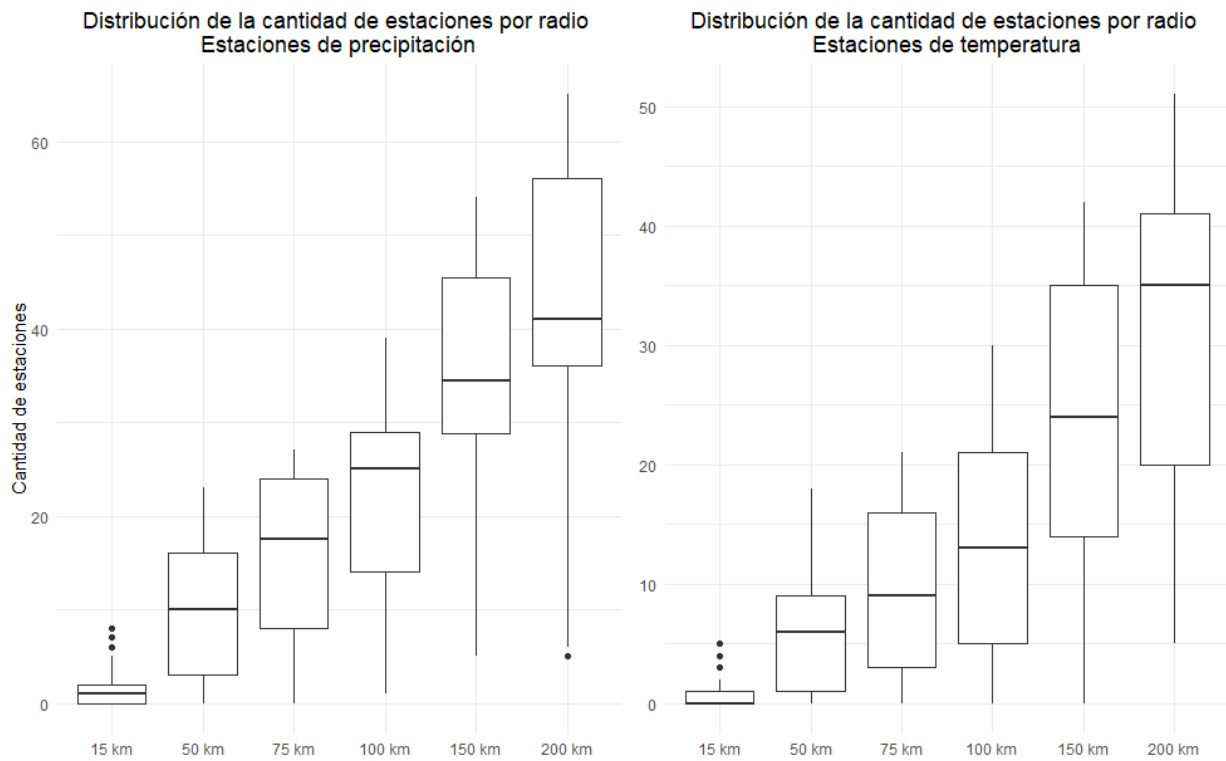


Figura 4.1: Cantidad de estaciones de precipitación (izquierda) y temperatura (derecha) según distancias a la redonda de cada estación.

Además de la baja densidad de estaciones meteorológicas en la zona de interés, caracterizada por la escasez de estaciones de los grupos 1 y 2, así como los periodos de registro limitados y las discontinuidades presentes, se suma la problemática de la baja calidad de algunos registros. Esta baja calidad se manifiesta, por ejemplo, en la falta de coherencia espaciotemporal y en rellenos arbitrarios.

A continuación, se presenta un análisis de las estaciones con registros de precipitación y temperatura disponibles en la zona de interés. Asimismo, se analiza la extensión y calidad de sus registros de manera tal de guiar la selección de estaciones previo a su utilización.

4.1.1 Estaciones meteorológicas

En total, se recopilaron datos de 141 estaciones meteorológicas con registros de precipitación en el área de estudio, pero solo 89 cuentan con información de temperatura. En la Figura 4.2 se muestra la distribución espacial de todas las estaciones disponibles en la zona de

estudio, coloreadas según disponibilidad de información completa consecutiva. Del mismo modo, en la Figura 4.3 se grafica distribución de las estaciones con temperatura en la zona de estudio.

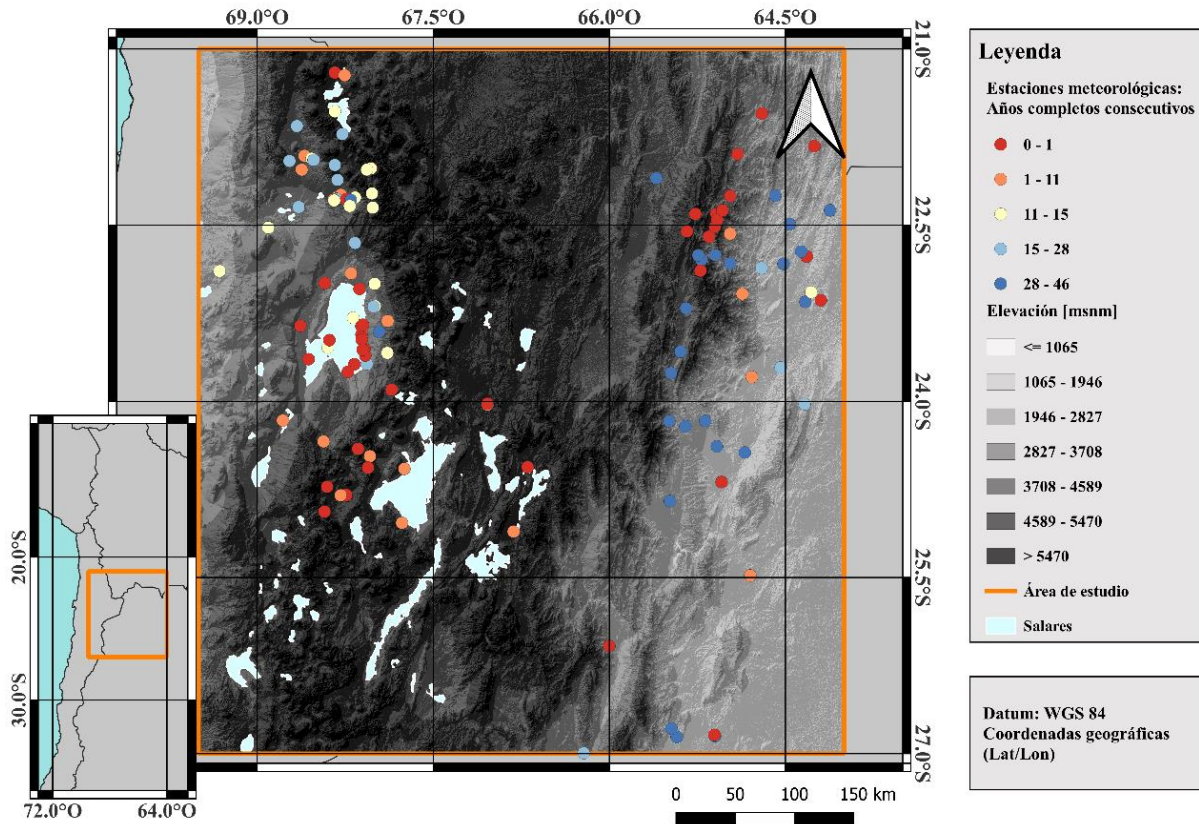


Figura 4.2: Área de estudio con elevación y estaciones meteorológicas con información de precipitaciones.

Se evidencia la distribución dispar de las estaciones de precipitación, guardando relación con las características fisiográficas del terreno y los límites políticos internacionales. En la Figura 4.4, se resume la distribución de las estaciones según su carácter público o privada y su administración política.

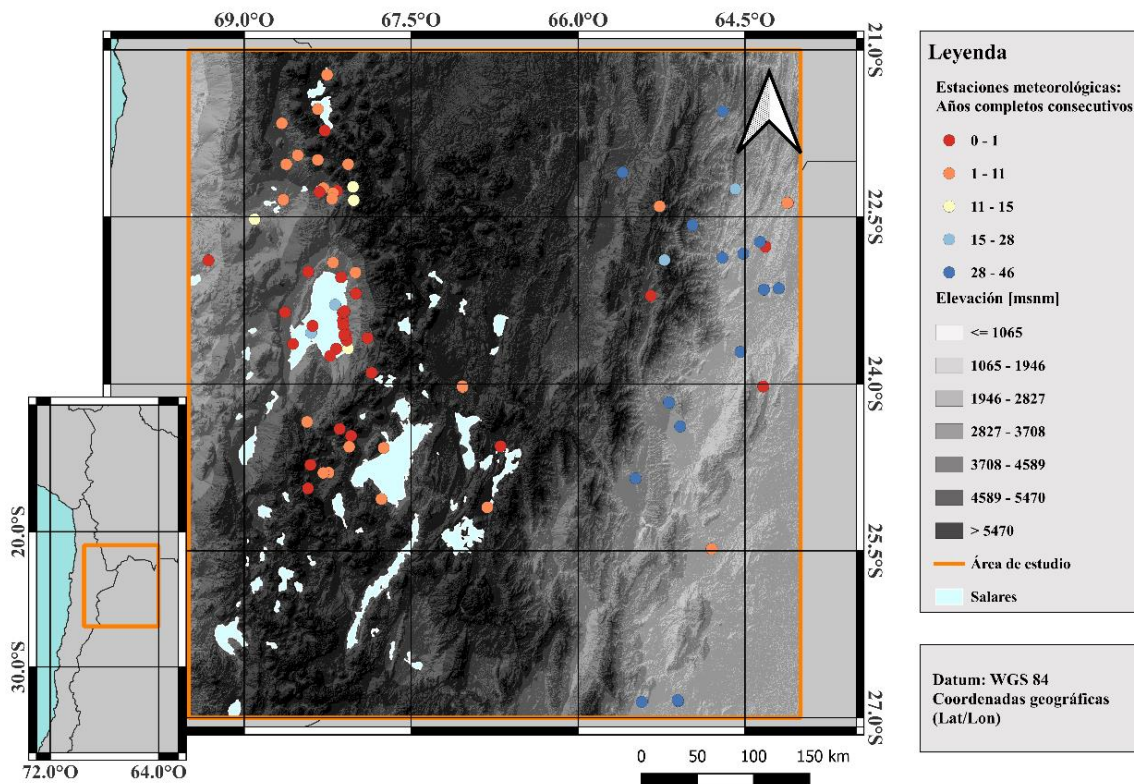


Figura 4.3: Área de estudio con elevación y estaciones meteorológicas con información de temperatura.

De la Figura 4.4, se desprende que la mayoría de las estaciones con registros de precipitación se concentran en Argentina, correspondiente a la mitad del total de las estaciones en el área de estudio. Por otra parte, Bolivia solo concentra el 5% del total, debido al reducido área de territorio contemplado en la investigación y la dificultad de conseguir información privada en ese país. Adicionalmente, se observa que respecto a la distribución de estaciones según fuente de información predominan las estaciones públicas por sobre las privadas. En el presente estudio se incorporan estaciones privadas disponibles en la base de datos de Montgomery & Associates Consultores Ltda.

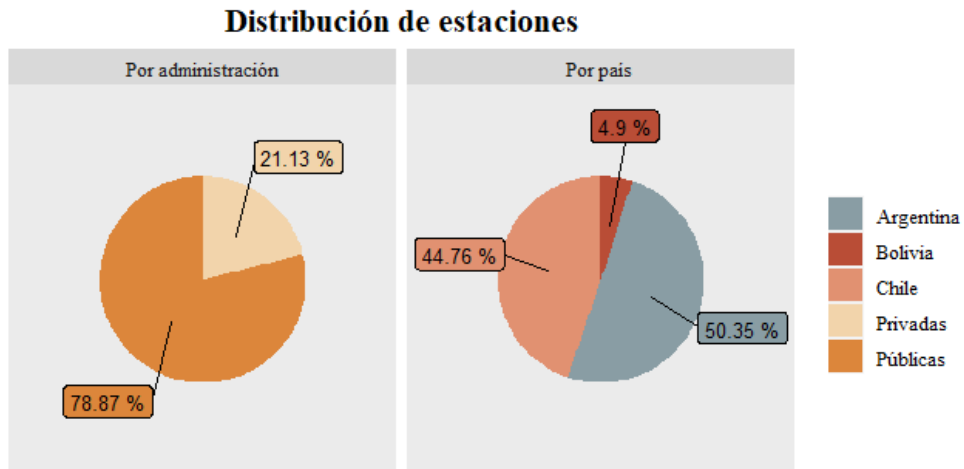


Figura 4.4: Distribución de estaciones según país de origen y administración de la información.

4.1.2 Revisión y filtro de estaciones según calidad

El control de calidad consta de un análisis cuantitativo, calculando la bondad de información según la cantidad de años seguidos de información completa, y un análisis cualitativo, enfocado en un control de malas prácticas (Abaurrea et al., 2004). Para efectos del análisis cualitativo se consideran los siguientes enfoques:

- Gráficos de cantidad de información semanal vs fracción de días húmedos (WD frac)
- Búsqueda de abuso en redondeos, con gráficas de distribución de valores decimales.

En la sección 4.1.2.2, se describen los análisis realizados y sus resultados.

4.1.2.1 Análisis cuantitativo

El uso de la data de una estación meteorológica se encuentra estrechamente ligada a la calidad y extensión de sus registros. Para el análisis, se usaron las directrices de Estévez et al. (2022). En primer lugar, mediante métodos de detección, se deben eliminar los valores “imposibles” de precipitación y temperatura, como registros diarios continuos de precipitaciones sobre los 100 [mm] en zonas con medias anuales por debajo de ese monto. Luego, a través de un ponderador de completitud α , se obtiene un umbral para clasificar como año “aprobado” o “reprobado” según la superación de este, tal como se muestra en la ecuación (4.1).

$$\text{SerieDiaria}_{año} \geq 365 * \alpha \rightarrow \text{Año aprobado} \quad (4.1)$$

$$\text{SerieDiaria}_{año} < 365 * \alpha \rightarrow \text{Año reprobado}$$

Estévez et al. (2022) define que una estación con al menos un 90% de la data diaria es considerado aceptable, con un 95% una buena calidad y sobre un 99% una calidad excelente. En la Figura 4.5 se presenta el números de estaciones según los años consecutivos, considerando los años aprobados según dos umbrales: a) 100% de la información, y b) con un mínimo de 85%.

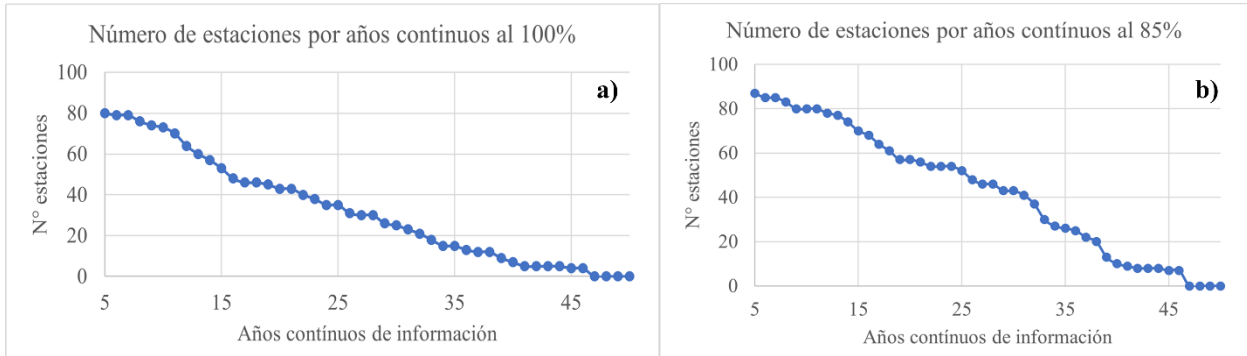


Figura 4.5: a,b: Gráfico de selección de estaciones según criterio escogido de años de data completa. En a) se considera el total de años de información completa, mientras que en b) el año es completo si tiene al menos el 85% de los días con información.

Tomando en cuenta que la cantidad de años seguidos a considerar dependerá de la función que se le quiera dar a la información climática, de manera complementaria a la Figura 4.5, la Tabla 4.2 resume el número de estaciones para distintos porcentajes de cumplimiento.

Tabla 4.2: Número de estaciones con una cantidad determinada de datos diarios para un número de años seguidos.

Cumplimiento anual	Número de años según cumplimiento								
	5	10	15	20	25	30	35	40	45
50%	94	84	78	69	61	52	39	23	14
75%	89	80	73	63	54	44	30	18	10
80%	87	80	72	61	54	44	29	17	10
85%	87	80	70	57	52	43	26	10	7
90%	85	79	70	57	52	43	26	10	7
95%	84	78	64	47	43	30	17	9	6
100%	80	73	53	43	35	25	15	7	4

Las 4 estaciones que cumplen con 45 años seguidos de información completa se encuentran en Argentina: Oran Aero (Ar_2_pub), Santiago del Estero Aero (Ar_7_pub), Las Mesadas (Ar_457_pub) y Astillero (Ar_638_pub).

4.1.2.2 Análisis cualitativo o de malas prácticas

El análisis de malas prácticas permite identificar errores sistemáticos en la medición y/o inscripción de datos por parte de los operadores de las estaciones. Entre los análisis más reveladores está la distribución semanal de datos y el WD frac asociado a cada uno de los días de la semana, permitiendo reconocer la existencia de tendencias en el registro de datos, como la ausencia de información en días no hábiles. Otra imagen interesante de observar es la distribución del último dígito de los datos, identificando abusos en el redondeo de los datos en contra de la precisión de estos, con resultados disponibles en el ANEXO B: REGISTRO DE ESTACIONES .

4.2 Disponibilidad de Productos Grillados

A continuación, se analizan los productos de precipitación (Sección 4.2.1) y temperatura (Sección 4.2.2) con cobertura en la zona de interés respecto a su desempeño a escala temporal diaria, mensual y anual.

4.2.1 Precipitación

A partir de la revisión de antecedentes, se seleccionan cinco productos de precipitación para evaluar su desempeño en el área de estudio. Estos corresponden a CHIRPSv2, PERSIANN-CDR, ERA5, IMERG y CHELSA, este último en una versión preliminar de resolución de 0.5° (aprox. 50 km). En la Figura 4.6 se grafican los valores de KGE' con sus demás índices para estos productos.

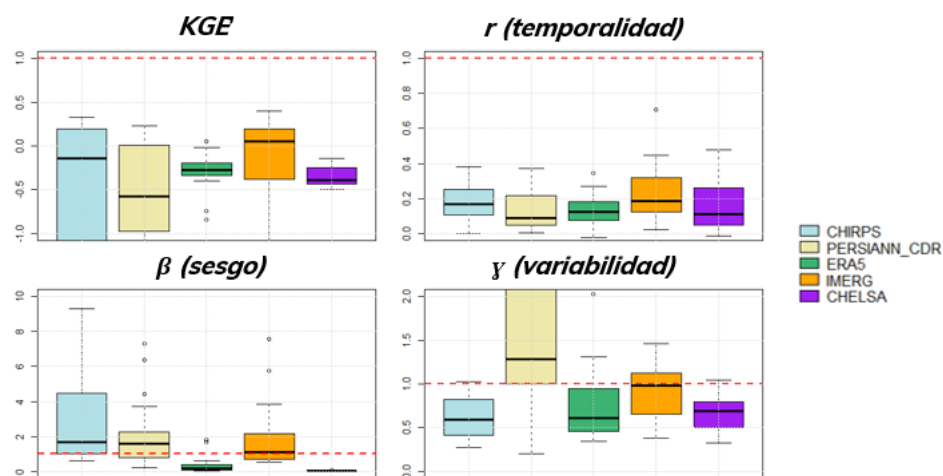


Figura 4.6: Valor de KGE' para los productos grillados de precipitación ERA5, PERSIANN-CDR, IMERG, CHIRPSv2 y CHELSA en la zona de estudio para el periodo 2007-2016 a escala diaria. La línea segmentada color rojo presenta el valor óptimo de cada indicador.

En la Figura 4.6 se puede notar que sólo IMERG presenta una mediana mayor a 0 para el KGE' , destacándose en temporalidad y variabilidad. Aun así, ninguno de los productos logra estimaciones lo suficientemente precisas como para confiar el desarrollo de estudios hidrológicos en el área de estudio. De manera complementaria, en la Tabla 4.3 se resumen otras métricas aplicadas para efecto del diagnóstico del desempeño de los diferentes productos.

Tabla 4.3: Mediana de los resultados de los índices de KGE' para los distintos productos globales de precipitación.

		CHIRPS	PERSIANN	ERA5	IMERG	CHELSA
Escala diaria	KGE'	-0,39	-0,40	-0,24	0,05	-0,39
	r	0,18	0,11	0,13	0,18	0,11
	β	1,80	1,64	0,16	1,05	0,01
	γ	0,55	1,08	0,59	0,88	0,69
	RMSE [mm]	4,63	5,78	3,82	4,84	3,84
Escala mensual	KGE'	-0,27	-0,39	0,03	0,34	-0,31
	r	0,61	0,36	0,61	0,67	0,26
	β	2,03	1,67	0,16	1,07	0,01
	γ	0,55	1,09	0,69	0,85	0,73
	RMSE [mm]	29,76	42,50	30,87	28,49	31,65
Escala anual	KGE'	-0,19	-0,75	-0,06	0,37	-0,49
	r	0,59	-0,21	0,65	0,75	-0,1
	β	1,77	1,76	0,17	1,09	0,01
	γ	0,29	0,84	0,65	0,66	0,71
	RMSE [mm]	11,88	16,53	22,32	10,09	23,92

4.2.2 Temperatura

Para estimar la temperatura en la zona de estudio, solo se recopilaron los productos grillados de ERA5 y CHELSA, con resoluciones de 0,25° y 0,5° respectivamente. En la Figura 4.7 se realiza la evaluación del rendimiento individual de los productos en el área de estudio.

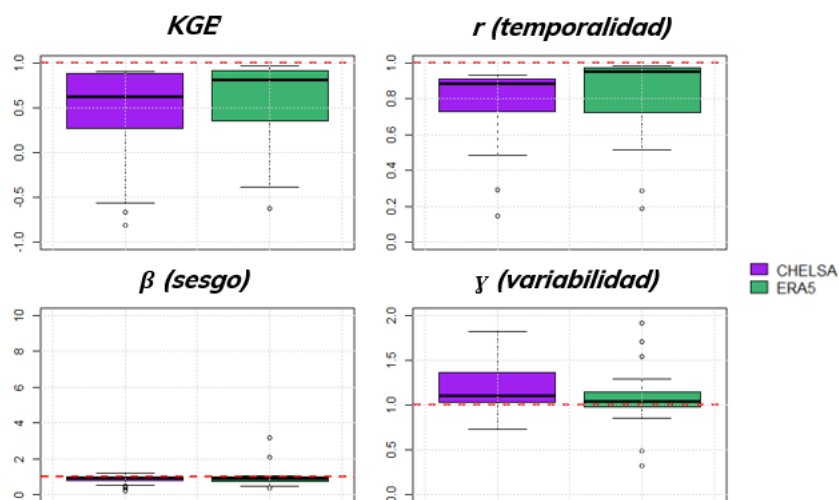


Figura 4.7: Valor de KGE' para los productos grillados de temperatura ERA5 y CHELSA en la zona de estudio para el periodo 2007-2016 a escala diaria. En línea segmentada color rojo se presenta el valor óptimo de cada indicador.

Al contrario de lo que ocurre con los productos de precipitación, las estimaciones de temperatura se ajustan bien a los datos observados, principalmente ERA5, con valores de mediana de KGE' cercanos a 0,8, con una buena temporalidad (en general valores de $r > 0,4$), sesgo cercano al óptimo (i.e., 1) y una variabilidad que, en términos de la mediana, se aproxima al óptimo del indicador. Esto se analiza en mayor detalle en la Tabla 4.4.

Tabla 4.4: Mediana de los resultados de los índices de KGE' para los distintos productos globales de temperatura.

		CHELSA	ERA5
Escala diaria	KGE'	0,62	0,81
	r	0,89	0,95
	β	0,95	0,86
	γ	1,11	1,04
	RMSE [°C]	4,37	4,85
Escala mensual	KGE'	0,81	0,66
	r	0,97	0,97
	β	0,86	0,95
	γ	1,03	1,08
	RMSE [°C]	4,06	4,15
Escala anual	KGE'	0,57	0,77
	r	0,84	0,91

	β	0,86	0,95
	γ	0,87	1,09
	RMSE [°C]	3,40	3,34

En conclusión, estos productos disponibles permiten capturar algunos de los patrones de variabilidad espaciotemporales de las variables de interés, entre los cuales puede destacar el sesgo acotado y la variabilidad para precipitación. De todas maneras, están muy lejos de llegar a valores óptimos que permitan calcular la temporalidad de los eventos. Con respecto a la temperatura, los productos destacan en todas las métricas, sobre todo el producto ERA5 a escala diaria y anual, mientras que CHELSA muestra un mayor rendimiento a escala mensual.

5. METODOLOGÍA

5.1 Metodología General

En este capítulo se describe la metodología empleada para la configuración del producto grillado de base observacional de precipitación y temperatura media diaria, denominado RFPUNA. El producto se construye en base a una grilla espacialmente distribuida de 0,05° latitud-longitud y utiliza como base la metodología propuesta para la construcción del producto RF-MEP (Baez-Villanueva et al., 2020).

RF-MEP es un producto grillado de precipitación diaria que combina mediante técnicas de *Random Forest* observaciones in-situ, productos grillados de precipitación y características fisiográficas del área de estudio. RF-MEP divide aleatoriamente las observaciones en bases de dato de entrenamiento y validación, dándole la alternativa al usuario para elegir la proporción entre ellos. La implementación de la metodología de RF-MEP se encuentra disponible en el paquete de *RFmerge* (Zambrano-Bigiarini et al., 2020) en la plataforma de libre acceso y código abierto R. Para mayor detalle sobre la metodología de RF-MEP se refiere al lector al trabajo realizado por Baez-Villanueva et al. (2020) y Zambrano-Bigiarini et al. (2020).

Tomando como base lo anterior, se presenta la configuración metodológica de RFPUNA para la construcción de los productos de precipitación y de temperatura, y las modificaciones realizadas al enfoque base definido por RF-MEP.

5.2 Producto RFPUNA: Precipitación

La construcción del producto RFPUNA de precipitación considera la base metodológica del paquete *RFmerge* para implementar el algoritmo de Random Forest (Mohammed et al., 2023; Nguyen et al., 2021; Y. Shi et al., 2023, Zambrano-Bigiarini et al. 2020). Sin embargo, el producto RFPUNA de precipitación se diferencia en varios aspectos a la configuración original de RF-MEP.

En la Figura 5.1 se observa la estructura del producto de precipitación RFPUNA. En la parte superior derecha del diagrama se ubica una leyenda de cada conjunto de pasos categorizados por colores. Allí se identifican los siguientes pasos metodológicos considerados para la implementación de RFPUNA de precipitación:

1. **Productos grillados:** De manera análoga a lo realizado por Nguyen et al. (2021), se introducen como productos de precipitación CHIRPSv2, PERSIANN-CDR, ERA5 e IMERG
2. **Datos de entrada:** Además de los productos grillados, ingresa al modelo la elevación, las características de orientación y pendiente del área de estudio para complementar la caracterización de las componentes fisiográficas, conjunto a los set de observaciones.

3. **Procesos:** Los procesos del modelo, los cuales siguen algunas directrices de RF-MEP, como el remuestreo y el uso de la distancia euclidiana entre las estaciones y los centroides de las grillas. No obstante, y a diferencia de RF-MEP, se incorpora una matriz binaria de intermitencia para clasificar eventos de precipitación y no precipitación mediante un umbral que dicte cuándo ocurre un evento de precipitación y cuándo no. El objetivo de incorporar esta matriz es reducir las falsas alarmas, forzando los valores bajos de precipitación simulada a ser considerados nulos, convirtiendo un supuesto evento a un día sin lluvia. El umbral se define preliminarmente como 0,1 [mm] según lo recomendado por la organización National Weather Service en Handbook, F. (2005)
4. **Random Forest:** En rojo están los procesos de *Random Forest*, basados en la configuración de RF-MEP, siguiendo los pasos del paquete *RFmerge* (Zambrano-Bigiarini et al., 2020). De cualquier modo, RFPUNA realiza otro *Random Forest* a raíz de la matriz de intermitencia, con el fin de distribuir las clasificaciones observadas. Ambas matrices de *Random Forest*, una con valores simulados de precipitación y otra binaria para clasificar eventos, se multiplican entre sí para eliminar falsos eventos, obteniendo el “*Random Forest* corregido por intermitencia” al medio de la Figura 5.1.
5. **Verificaciones:** Inmediatamente después del *Random Forest* de precipitación, se realiza la validación mediante la metodología *Leave One Out Cross Validation* (LOOCV). Esta técnica se aparta del enfoque habitual de clasificación aleatoria de los conjuntos de datos para entrenamiento y validación, como se hace en RF-MEP.
6. **Calibración:** Siguiendo el diagrama, existe una segunda validación según resultados de probabilidad de detección y de radio de falsa alarma (POD y FAR por sus siglas en inglés respectivamente), la cual buscará maximizar y minimizar los parámetros respectivamente mediante un método de calibración del umbral de incertidumbre, considerando 20 valores equidistantes con valores de umbral entre 0,01 y 1 [mm]. Esta segunda validación sí se realizará mediante clasificación aleatoria de *datasets* de entrenamiento y validación por ser un método más rápido. Una vez que se obtenga el valor que optimice los índices se obtiene el producto determinístico de RFPUNA.
7. **Bluecat:** Calculada la versión determinística de RFPUNA, es posible ingresar la serie simulada conjunto a las series observadas y obtener una serie estocástica asociada a distintas bandas de incertidumbre según se defina.

8. **Resultados:** Si bien el primer resultado es el producto determinístico de RFPUNA, se espera configurar un producto estocástico distribuyendo los promedios de las series sintéticas de *Bluecat*. Se presume que los métodos de interpolación del tipo IDW o *Kriging* puedan ser útiles para distribuir la incertidumbre en toda el área de estudio.

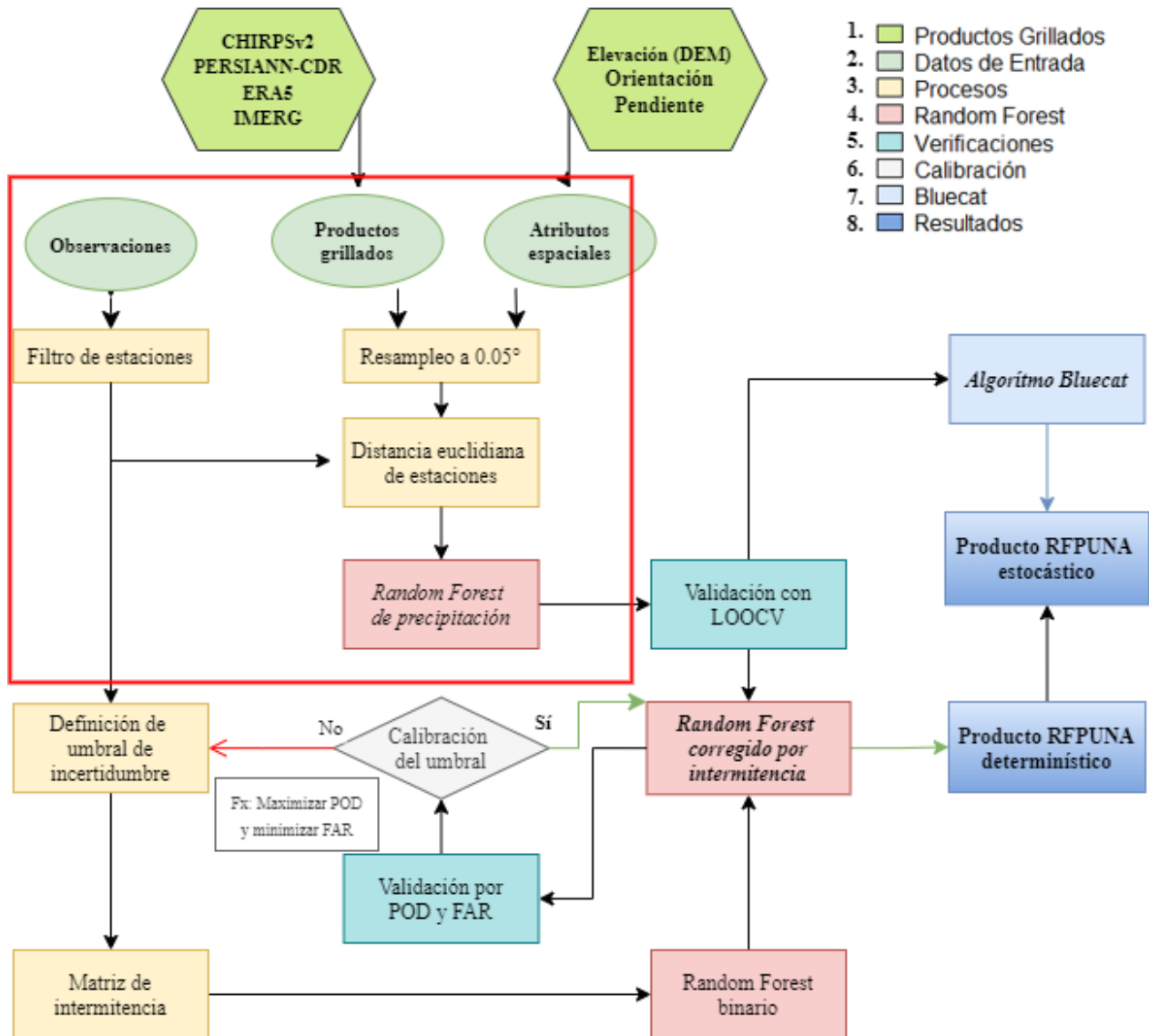


Figura 5.1: Metodología para configuración de las precipitaciones del producto grillado RFPUNA. En rojo se encierran los pasos que presentan similitudes con la configuración de RFMEP.

5.3 Producto RFPUNA: Temperatura

Para las estimaciones de temperatura, se usará el mismo método anteriormente mencionado, variando los datos de entrada y eliminando la corrección por intermitencia, ya que no existen eventos binarios de temperatura. El diagrama metodológico para la elaboración del producto grillado de temperatura RFPUNA está en la Figura 5.2, siguiendo un orden similar al producto de precipitación:

1. **Productos grillados:** La elección de los productos grillados se hizo, en gran parte, por correlaciones básicas entre variables climáticas, y las restricciones propias de cada variable. Por esto, se descarta el uso del producto CHELSA, por su gran correlación con las demás variables de temperatura, acotar el periodo de estudio hasta 2016, y tener una resolución espacial o temporal muy gruesa. También ingresan otros productos de ERA5 relevantes para el estudio, como la precipitación, presión atmosférica y evaporación potencial.
2. **Datos de entrada:** Además de los productos grillados, ingresa al modelo la elevación, las características de orientación y pendiente del área de estudio para complementar la caracterización de las componentes fisiográficas, conjunto a los set de observaciones.
3. **Procesos:** Los procesos del modelo de temperaturas son análogos a la configuración básica del producto de precipitación de RFPUNA, como el resampleo y la obtención de grillas con la distancia euclidiana entre las estaciones. La diferencia radica en eliminar los procesos de RF binario.
4. **Random Forest:** En rojo están los procesos de *Random Forest*, basados en la configuración de RF-MEP, siguiendo los pasos del paquete *RFmerge* (Zambrano-Bigiarini et al., 2020).
5. **Verificaciones:** Inmediatamente después del *Random Forest* de temperatura, se realiza la validación mediante la metodología *Leave One Out Cross Validation* (LOOCV). Esta técnica se aparta del enfoque habitual de clasificación aleatoria de los conjuntos de datos para entrenamiento y validación, como se hace en RF-MEP.
6. **Bluecat:** Calculada la versión determinística de RFPUNA de temperatura, se ingresan las series simulada y observadas para obtener una serie estocástica asociada a distintas bandas de incertidumbre, según se defina.
7. **Resultados:** Si bien el primer resultado es el producto determinístico de RFPUNA, se espera configurar un producto estocástico distribuyendo los promedios de las series sintéticas de *Bluecat*. Se presume que los métodos de interpolación del tipo

IDW o *Kriging* puedan ser útiles para distribuir la incertidumbre en toda el área de estudio.

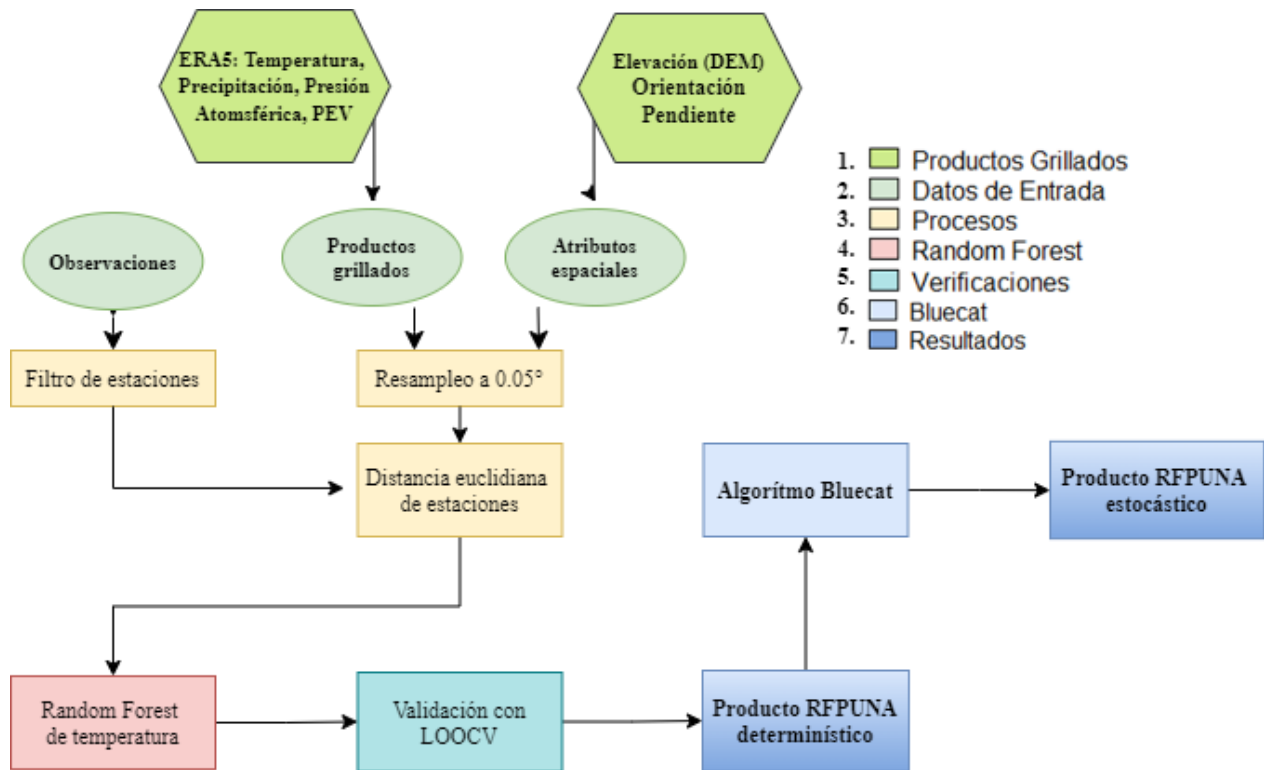


Figura 5.2: Metodología para configuración del producto grillado RFPUNA de temperatura.

Al contrario de la precipitación, no se encontraron estudios que relacionen metodologías de *Random Forest* con la elaboración de productos de temperatura. Por ende, la elección de los datos de entrada se hizo por correlaciones básicas entre variables climáticas, y las restricciones propias de cada variable.

5.4 Aplicación de algoritmo *Bluecat*

Validados los resultados de precipitación y temperatura del *Random Forest* mediante LOOCV, se ingresan los valores del modelo RFPUNA determinístico al algoritmo *Bluecat* (Koutsoyiannis & Montanari, 2022a) para la grilla de cada estación. Para ello, además de los resultados determinísticos de RFPUNA y las observaciones disponibles, se definen dos parámetros que definen los límites de confianza de las bandas de incertidumbre. El algoritmo entrega 5 series

diarias de nuevos valores asociados al producto determinístico ingresado. La primera serie diaria corresponde al promedio de los valores estocásticos, la segunda corresponde a la serie determinística ingresada sin cambios, mientras que las últimas tres series están asociadas a los valores límites de las probabilidades de la banda y la mediana. Los límites de las bandas se calculan por la ecuación (5.1), donde l_1 y l_2 corresponden a los límites de las bandas de incertidumbre a partir de los parámetros m_1 y m_2 .

$$l_1 = \left\lceil \frac{100}{m_1+m_2} \right\rceil ; l_2 = 100 - l_1 \quad (5.1)$$

A modo de ejemplo, si se ingresan las entradas pp_{sim} , pp_{obs} , $m_1 = 10$, $m_2 = 10$, como resultado de la implementación se tendrán las series diarias de pp_{sim} , junto a la serie del promedio y mediana estocástica, y dos series diarias asociadas a la probabilidad de ocurrencia del 5 y 95%, que corresponden a los valores límites de la banda de incertidumbre ingresada (*i.e.*, m_1 y m_2).

Se debe precisar que los resultados obtenidos con *Bluecat* se asocian a cada estación dentro del dominio y no se cuenta, en forma directa, con una distribución espacial de la incertidumbre. En este contexto, para la distribución espacial de los resultados obtenidos con *Bluecat* se debe adoptar algún enfoque complementario (*e.g.*, métodos de interpolación geoestadística, *Machine Learning*, etc.).

5.5 Diagnóstico y ajuste de los Productos

5.5.1 Análisis por estación (puntual)

Para evaluar el desempeño del modelo se definen distintos criterios, entre ellos, el KGE' y el RMSE, definidos por las ecuaciones (5.2) y (5.3); **Error! No se encuentra el origen de la referencia.** respectivamente.

$$KGE' = 1 - \sqrt{(r - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} ; \text{ con } \beta = \frac{\sigma_s}{\sigma_o} \text{ y} \quad (5.2)$$

$$\gamma = \frac{\bar{S}}{\bar{O}}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (S_i - O_i)^2} \quad (5.3)$$

Donde:

r : Coeficiente de correlación de Pearson.

\bar{S} : Promedio asociado a la serie de tiempo de la variable simulada.

\bar{O} : Promedio asociado a la serie de tiempo de la variable observada.

σ_s : Desviación estándar asociada a la serie de tiempo de la variable simulada.

σ_o : Desviación estándar asociada a la serie de tiempo de la variable observada.

N: Número de días evaluados.

El KGE', además de entregar su valor compuesto, se puede desagregar en sus diferentes componentes. En primer lugar, el coeficiente de correlación de Pearson r mide la temporalidad de las estimaciones respecto a la observación; luego, el parámetro β calcula los sesgos asociados y el índice γ representa la variabilidad de la estimación. Todos estos valores tienen en la unidad su valor óptimo, al igual que el KGE' integrado. Por su parte, el RMSE calcula la raíz cuadrada del error cuadrático medio, cuantificando el error asociado a cada simulación. Su valor óptimo es 0.

A modo de complementar el diagnóstico del desempeño del modelo, los resultados se analizan según diferentes agregaciones temporales. Así, los resultados se analizan según periodos trimestrales, considerando las estaciones de otoño (MAM), invierno (JJA), primavera (SON) y verano (DJF por sus siglas en inglés), y semestrales (octubre a marzo y abril a septiembre). Cabe mencionar que se define el primer semestre de acuerdo con el mayor registro de precipitaciones en la zona.

Para evaluar el rendimiento del producto de precipitación, se incluyen métricas discretas recomendadas en la literatura (Nurmi & Department, 2003) para evaluar la habilidad en la captura de eventos para diferentes intensidades de precipitación (ver Tabla 5.2 y Tabla 5.2). A continuación, se presentan las ecuaciones de los índices de Probabilidad de Detección, Frecuencia de Sesgos, Radio de Falsa Alarma y el Índice Crítico de Éxito (POD, f_{bias} , FAR, CSI por sus siglas en inglés respectivamente).

Tabla 5.1: Métricas de la tabla de contingencia.

Métrica	Fórmula	Rango	Valor óptimo	
POD	$\frac{H}{H + M}$	[0, 1]	1	(5.4)
Fbias	$\frac{H + M}{H + F}$	[0, ∞^+)	1	(5.5)
FAR	$\frac{H + M}{F}$	[0, 1]	0	(5.6)
CSI	$\frac{H + F}{H + M + F}$	[0, 1]	1	(5.7)

Donde:

H: Evento registrado por las observaciones y la simulación (*Hit*).

F: Evento registrado solo por la simulación (*False alarm*).

M: Evento registrado solo por la observación (*Miss*).

Tabla 5.2: Clasificación de eventos basados en intensidad de lluvia (preliminar).

Evento de precipitación	Intensidad [mm/d]
Sin lluvia	[0,1)
Lluvia ligera	[1,5)
Lluvia moderada	[5,10)
Lluvia fuerte	[10,30)
Lluvia violenta	≥ 30

El POD calcula que tan seguido la simulación estima correctamente la intensidad de precipitación observada, con valor óptimo 1. Por su parte, f_{bias} compara el número de eventos simulados y observados; si es mayor a 1, la ocurrencia está sobrestimada por la simulación; si es menor, la subestima. El índice FAR mide la fracción de eventos que no son correctamente definidos por la simulación, este tiene su valor óptimo en 0. Por último, el CSI pondera los éxitos (H) con relación a todos los eventos registrados, encontrando su valor óptimo en la unidad.

5.5.2 Análisis de distribución espacial

Con el fin de evaluar la representación espacial de las variables de estudio y cómo se ajustan los valores modelados respecto a la observación, se analiza la correlación espacial a través del índice de Morán. Este índice tiene por objeto evaluar si un patrón observado es aleatorio o no (Getis, 2010; Ghalhari et al., 2016a). La hipótesis nula del índice dice que los valores de las variables estudiadas están distribuidos aleatoriamente en el espacio, es decir, que no existe autocorrelación espacial. Con el p_{valor} , asociado a una significancia del 5%, se puede comprobar la hipótesis nula.

La ecuación (5.8) presenta la composición del índice de Moran, donde y_i e \bar{y} representan las observaciones y las medias de las observaciones respectivamente; n es el número de observaciones; $w_{i,j}$ es el elemento de la matriz de pesos espaciales entre las ubicaciones i y j ; y S_0 es la suma total de todos los pesos espaciales.

$$\text{Índice de Morán} = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{i,j} (y_i - \bar{y})(y_j - \bar{y})}{S_0 \sum_{i=1}^n (y_i - \bar{y})^2} \quad (5.8)$$

En la fórmula presentada anteriormente, el numerador representa la covarianza ponderada espacialmente de los valores en relación con la media y el denominador la varianza total de los valores. El índice de Moran puede variar entre -1 (indicando autocorrelación espacial negativa) y +1 (autocorrelación espacial positiva). Un valor cercano a 0 sugiere un patrón espacial aleatorio, sin una autocorrelación significativa. Este índice es comúnmente usado en geografía (Musakwa &

Van Niekerk, 2014), ecología (Karbalaeei Saleh et al., 2021) y otras disciplinas que estudian la distribución espacial de fenómenos (Ghalhari et al., 2016b).

5.5.3 Análisis de importancia de predictores

A modo de analizar la coherencia de los predictores seleccionados en la modelación con *Random Forest* y la conceptualización que se tiene sobre los procesos y mecanismos que condicionan el clima en la zona de estudio, se analiza la importancia relativa de ellos. Para ello, se utiliza el Índice de Gini. Este índice es una medida estadística utilizada para evaluar la pureza de un conjunto de datos, y se define según la ecuación (5.9), donde p_i corresponde a la proporción de elementos de clase i en el conjunto.

$$\text{Índice de Gini} = 1 - \sum p_i^2 \quad (5.9)$$

Los valores que puede adoptar el índice de Gini son entre 0 y 1, donde un valor bajo del índice indica que el nodo (zona de ramificación del árbol) es muy homogéneo, es decir, que la mayoría de las instancias en ese nodo pertenecen a una sola clase. Por el contrario, un valor cercano a 1 representa heterogeneidad, lo que significa que en ese nodo se presentan diferentes clases para la toma de decisión, estos valores son considerados impuros. Un buen predictor en el modelo es el que mejor reduce estas impurezas.

6. RESULTADOS

6.1 Análisis de variables y configuración del modelo

Previo al desarrollo del modelo, se analiza la correlación entre las variables observadas y potenciales predictores. Para ello, se calcula el coeficiente de correlación de Spearman asociado a cada una de las variables seleccionadas para el análisis. En la Figura 6.1 se presentan los resultados asociados al coeficiente de correlación de Spearman para todas las variables en estudio. Cabe mencionar que, para efectos del análisis, se considera una agregación temporal anual y espacial interpolada al punto definido por las coordenadas de las estaciones disponibles en el dominio. Los números y los colores representan la correlación, mientras que la 'x' indica que la correlación no es significativa por superar el umbral del p-valor definido en 0,05.

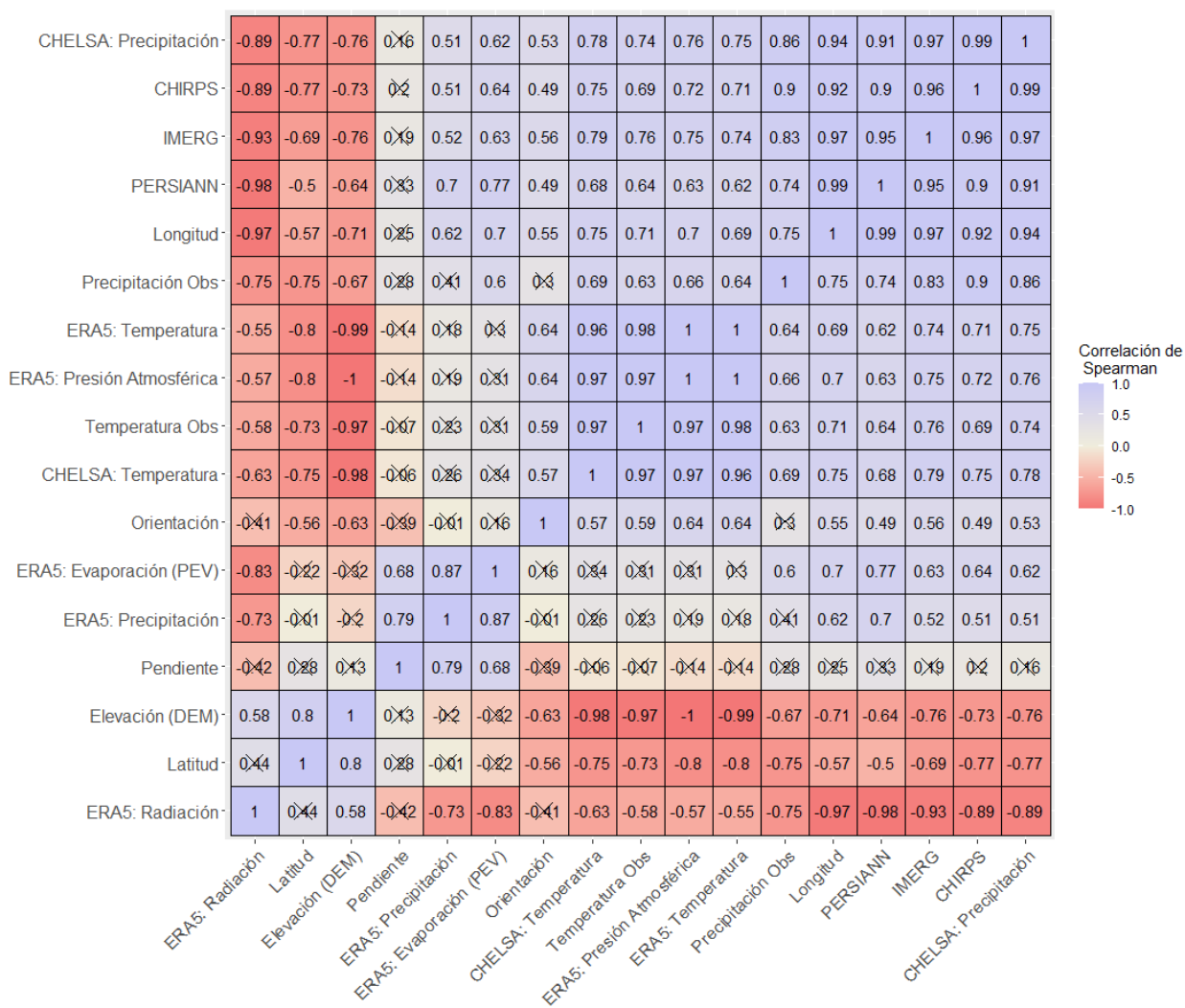


Figura 6.1: Correlación de las variables en estudio a escala anual en toda el área de estudio. Se consideraron los periodos anuales entre 2007-2016.

Al revisar las correlaciones presentadas en la Figura 6.1, se observa una correlación fuerte ($\rho > 0,8$) entre los productos de precipitación IMERG, PERSIANN-CDR, CHIRPS, CHELSA con la longitud del área de estudio, mientras que ERA5 presenta una correlación no tan fuerte con los otros productos. No obstante, los productos antes mencionados no presentan una correlación significativa con las precipitaciones observadas. Por su parte, los productos de temperatura ERA5 y CHELSA presentan una fuerte correlación entre ellos ($\rho = 0,96$), junto al producto de presión atmosférica de ERA5 y las observaciones de temperatura.

A modo de complementar el análisis de correlación, se repite el procedimiento diferenciando las estaciones en Chile y Argentina (Bolivia no es considerado ya que solo se tienen 3 estaciones de ese país). De allí se observan correlaciones distintas a distintos lados de la cordillera de los Andes para variables de precipitación, temperatura y evaporación. Estos resultados son presentados en detalle en el ANEXO C: DATOS DE ENTRADA AL MODELO.

6.1.1 Precipitación

Si bien una de las características de las herramientas de *ML* es su independencia para ponderar la importancia de las variables, por razones de parsimonia se descartan variables con alta correlación y físicamente ligadas (i.e., redundancia de variables) si es que su incorporación implica una limitante temporal o espacial. Lo anterior a modo de reducir potenciales problemas de sobreajuste del modelo. En consecuencia, se seleccionan los productos de precipitación PERSIANN-CDR, IMERG, CHIRPSv2 y ERA5. Se excluye CHELSA considerando que la versión recuperada sólo se extiende hasta 2016 y su resolución espacial gruesa en comparación a los otros productos ($0,50^\circ$ respecto a $0,05^\circ - 0,10^\circ - 0,25^\circ$).

A partir del criterio de selección mencionado anteriormente se pudieron haber eliminado más de un producto, sin embargo, se prefirió mantener algunas de las recomendaciones de Baez-Villanueva et al. (2020) e incluir dos de los tres productos usados en la configuración de RFMEP (i.e., PERSIANN-CDR y CHIRPSv2), conjunto a la versión más reciente de ERA, agregando ERA5 en vez de ERA INTERIM. Finalmente, observando el desempeño de los productos en el área de estudio (ver Figura 4.6), se integra IMERG al modelo.

De manera complementaria a los productos de precipitación seleccionados, se agrega elevación, orientación y pendiente como parte de los predictores empleados para la modelación. Adicionalmente, se incluye el producto de temperatura de ERA5.

Debido a la extensión temporal de los productos seleccionados y la extensión de los registros de las estaciones disponibles, se utiliza el periodo 2007-2017 para la configuración y análisis del producto de precipitación de RFPUNA. Se debe precisar que sólo 39 estaciones meteorológicas cumplen con el criterio antes mencionado, el cual a su vez, se define para maximizar dicho número.

Luego de implementar el modelo de *RF*, se estima el índice de Gini para los predictores empleados. Dicho índice permite evaluar la importancia relativa de cada variable en la toma de decisiones del modelo. Los resultados obtenidos se resumen en la Figura 6.2

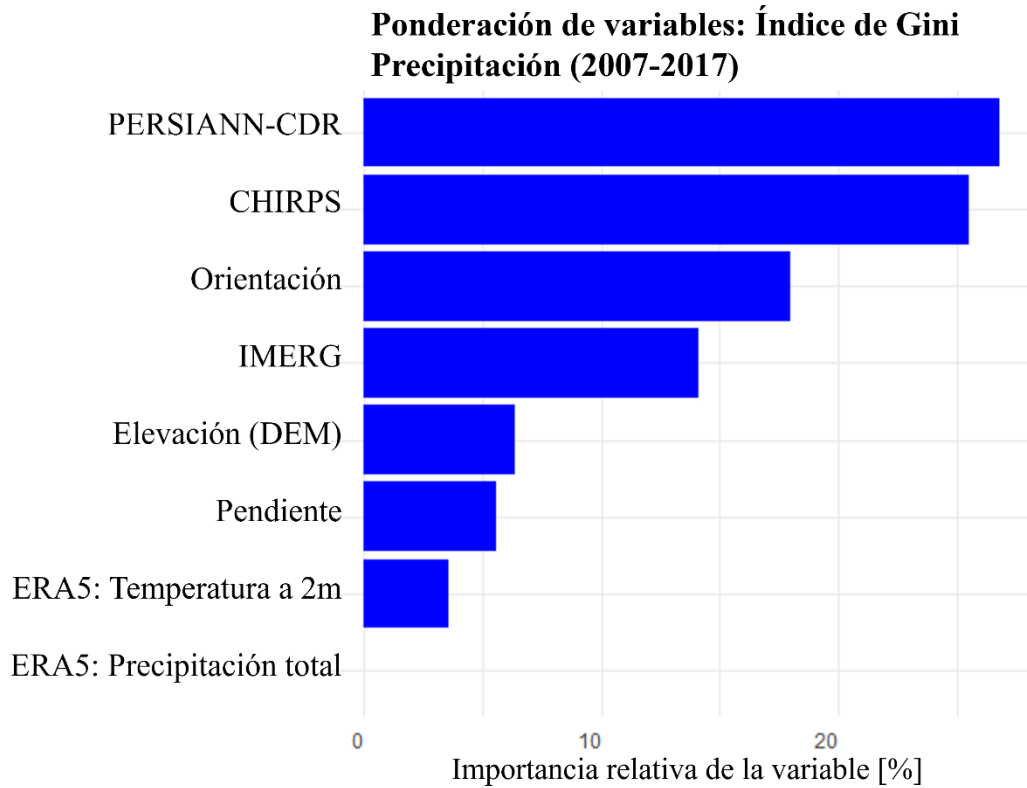


Figura 6.2: Índice de Gini de los predictores de Random Forest para precipitación. Se consideró el periodo entre 2007-2017.

Del análisis, se puede concluir que el producto de precipitación ERA5 no aporta al modelo, mientras que la temperatura aporta en al menos 3% de los nodos de decisión. Por el contrario, el análisis revela que la orientación juega un papel importante en las predicciones del modelo

Adicionalmente, se ajusta el valor umbral óptimo de intermitencia (i.e., ocurrencia o no ocurrencia de precipitación). Para ello se evalúa el rendimiento del producto según diferentes configuraciones de umbral (20 diferentes) entre 0,01 y 1 (mm/d). Los experimentos realizados se detallan en el ANEXO A: RESULTADOS DE LA CALIBRACIÓN Y VERIFICACIÓN DEL MODELO, de los cuales se desprende que un umbral equivalente a 0,5 [mm] permite maximizar el rendimiento del modelo según diferentes métricas de evaluación.

6.1.2 Temperatura

Para la configuración del producto de temperatura RFPUNA, se eligen las variables asociadas a ERA5 por su amplia extensión temporal y la cantidad de elementos de interés que posee, como variables de precipitación, temperatura a 2 metros, evapotranspiración potencial, presión atmosférica. Debido a la extensión temporal de los productos seleccionados y la extensión de los registros de las estaciones disponibles, se utiliza el periodo 2007-2016 para la configuración y análisis del producto de temperatura de RFPUNA. A partir de dicho criterio, se cuenta con 42 estaciones meteorológicas con datos de temperatura para el análisis.

La aplicación del índice de Gini (Figura 6.3) revela que la elevación y los productos de ERA5 son las variables más relevantes en las estimaciones, a excepción del producto de precipitación. Las tres variables que más reducen las impurezas están fuertemente correlacionadas, tanto positiva como negativamente.

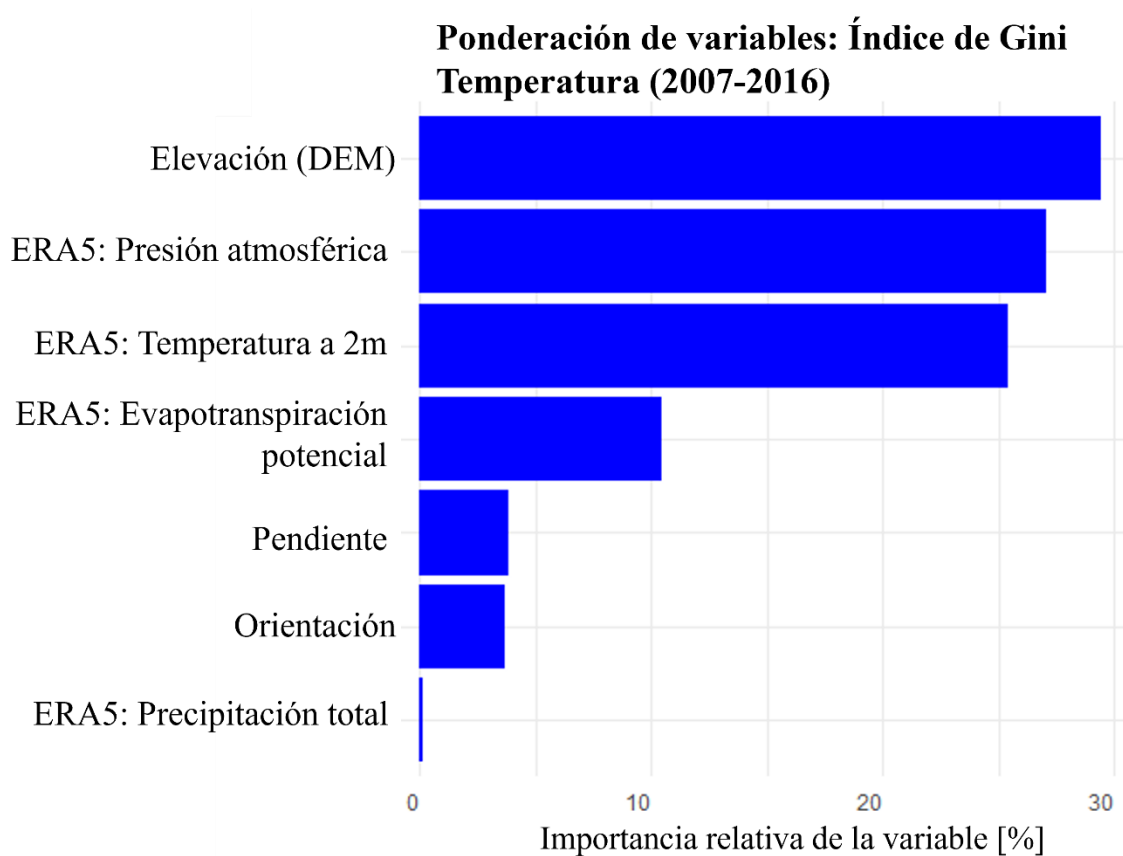


Figura 6.3: Índice de Gini de los predictores de Random Forest para temperatura. Se consideró el periodo entre 2007-2016.

6.2 Evaluación y Verificación de los Productos de Precipitación y Temperatura

6.2.1 Precipitación

A partir del producto de precipitación RFPUNA se generan 4 versiones diferentes, las cuales se describen en la Tabla 6.1. Adicionalmente, se incluye para efectos de comparación una versión del producto RF-MEP configurada con las estaciones empleadas para la construcción de RFPUNA.

Tabla 6.1: Descripción de los productos grillados obtenidos.

ID	Tipo	Descripción
RFMEP	Determinístico	Configuración descrita por (Baez-Villanueva et al., 2020) aplicada al área de estudio. Para su elaboración se consideraron los productos CHIRPSv2, PERSIANN-CDR y ERA5, además del componente de elevación.
RFPUNA	Determinístico	Configuración original del producto
RFPUNA1	Probabilístico	RFPUNA con <i>Bluecat</i> aplicado con un rango de confianza del 97.5 y 2.5%
RFPUNA2	Probabilístico	RFPUNA con <i>Bluecat</i> aplicado con un rango de confianza del 95 y 5%
RFPUNA3	Probabilístico	RFPUNA con <i>Bluecat</i> aplicado con un rango de confianza del 90 y 10%

En la Figura 6.4 se muestran los resultados de KGE' de los productos para el periodo de 2007 – 2017 y la desagregación de sus tres componentes evaluadas según tres distintas agregaciones temporales: diaria, mensual y anual. A modo de complementar la comparación, se incluyen los productos grillados analizados anteriormente (*i.e.*, CHIRPSv2, PERSIANN-CDR, ERA5 e IMERG) y RF-MEP (Baez-Villanueva, et al., 2020) implementado en toda la zona de estudio. La línea roja marca el valor óptimo de cada parámetro y la línea negra de los gráficos de caja indica el valor mediano de los resultados.

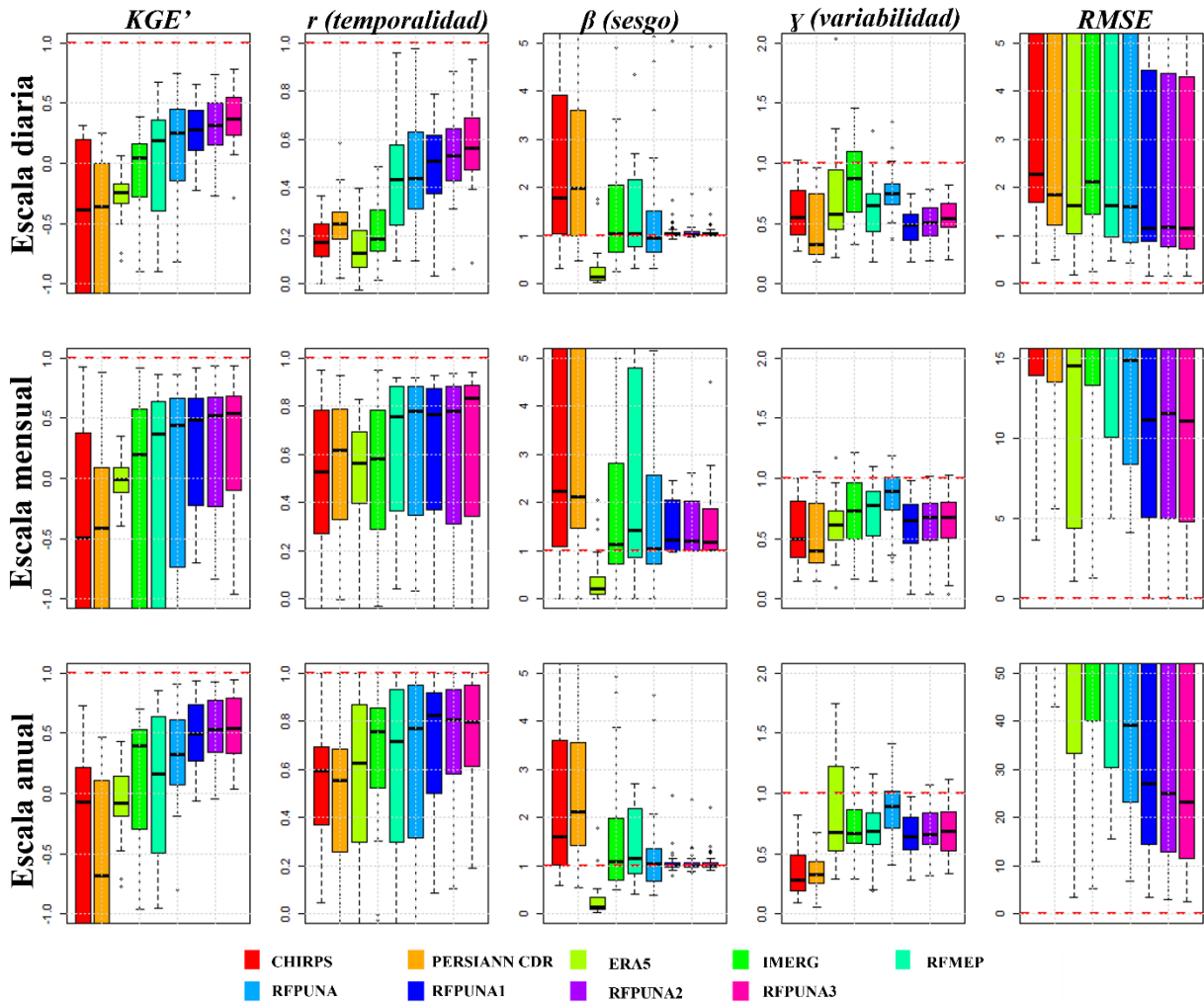


Figura 6.4: Gráficos de caja asociados a productos grillados de precipitación y su KGE' con los parámetros r , β y γ para el periodo 2007-2017 en toda el área de estudio.

De los gráficos de caja se destaca el rendimiento de los productos RFPUNA y, adicionalmente, de IMERG a escalas temporales grandes, con valores medianos mejores que los productos RFMPEP y RFPUNA determinístico a escala anual. De manera complementaria a la Figura 6.4, en la **¡Error! No se encuentra el origen de la referencia.** se resumen las medianas asociadas a cada producto, componente del KGE' y agregación temporal, para distintas zonas de evaluación, considerando toda el área de estudio (Figura 6.4), el territorio chileno y las estaciones en Argentina, con sus respectivos gráficos de caja en el ANEXO A: RESULTADOS DE LA CALIBRACIÓN Y VERIFICACIÓN DEL MODELO .

Tabla 6.2: Mediana de los resultados de los índices de KGE' para los distintos productos de precipitación.³

		CHIRPS	PERS ⁴	ERA5	IMERG	RFMEP	RFP ⁵	RFP1	RFP2	RFP3	
Todo	Día	KGE'	-0.39	-0.36	-0.24	0.05	0.19	0.25	0.28	0.31	0.37
		r	0.17	0.25	0.13	0.19	0.43	0.44	0.51	0.53	0.56
		β	1.79	1.99	0.15	1.04	1.04	0.94	1.01	1.02	1.02
		γ	0.55	0.33	0.58	0.88	0.65	0.75	0.49	0.51	0.55
		RMSE [mm]	2.27	1.84	1.63	2.13	1.62	1.60	1.15	1.17	1.15
	Mes	KGE'	-0.49	-0.41	-0.01	0.20	0.37	0.44	0.49	0.52	0.54
		r	0.53	0.62	0.56	0.58	0.76	0.78	0.76	0.78	0.83
		β	2.23	2.11	0.20	1.13	1.43	1.04	1.22	1.20	1.17
		γ	0.49	0.40	0.61	0.73	0.78	0.89	0.65	0.68	0.67
		RMSE [mm]	23.36	18.21	14.50	19.87	16.49	14.88	11.13	11.55	11.10
	Año	KGE'	-0.07	-0.68	-0.08	0.40	0.17	0.32	0.49	0.53	0.54
		r	0.59	0.55	0.63	0.76	0.72	0.77	0.82	0.80	0.80
		β	1.61	2.11	0.15	1.09	1.15	1.03	1.00	1.00	1.00
		γ	0.28	0.33	0.68	0.67	0.69	0.89	0.64	0.66	0.69
		RMSE [mm]	88.46	107.50	89.78	69.13	53.63	39.19	26.97	25.05	23.31
Chile	Día	KGE'	-0.52	-1.49	-0.24	-0.03	0.15	0.21	0.34	0.38	0.50
		r	0.14	0.21	0.15	0.17	0.43	0.43	0.57	0.59	0.67
		β	2.11	3.25	0.28	1.10	1.31	1.01	1.02	1.02	1.02
		γ	0.43	0.25	0.48	0.66	0.61	0.74	0.52	0.57	0.63
		RMSE [mm]	1.81	1.39	1.26	1.82	1.07	0.98	0.95	0.96	0.93
	Mes	KGE'	-0.53	-1.78	0.08	0.08	0.36	0.52	0.47	0.51	0.54
		r	0.44	0.56	0.61	0.55	0.76	0.82	0.76	0.78	0.83
		β	2.32	3.64	0.31	1.13	1.45	1.13	1.22	1.20	1.17
		γ	0.46	0.37	0.59	0.62	0.65	0.92	0.66	0.68	0.67
		RMSE [mm]	13.62	14.48	11.44	13.62	9.93	7.30	7.78	7.30	7.00
	Año	KGE'	-0.73	-0.98	0.00	0.34	0.54	0.60	0.71	0.72	0.76
		r	0.58	0.57	0.79	0.76	0.91	0.90	0.90	0.90	0.91
		β	2.23	2.70	0.28	1.19	1.28	1.04	1.02	1.00	1.01
		γ	0.24	0.26	0.57	0.62	0.77	0.93	0.73	0.75	0.80
		RMSE [mm]	71.41	78.21	48.68	47.39	31.88	25.67	21.42	20.85	18.19
Argentina	Día	KGE'	0.21	0.02	-0.30	0.16	0.20	0.26	0.25	0.29	0.33
		r	0.25	0.28	0.09	0.29	0.44	0.44	0.48	0.49	0.53
		β	1.04	1.00	0.07	1.03	0.92	0.86	1.00	1.00	1.00
		γ	0.86	0.80	0.97	1.10	0.66	0.76	0.48	0.50	0.53
		RMSE [mm]	8.33	7.63	7.07	8.86	6.05	6.01	7.09	7.10	7.11
	Mes	KGE'	0.56	-0.02	-0.05	0.58	0.39	0.37	0.54	0.55	0.56

³ Se destacan en negrita los mejores valores por parámetro.

⁴ PERS hace referencia al producto grillado PERSIANN-CDR.

⁵ RFP hace referencia al producto grillado RFPUNA.

Año	r	0.86	0.83	0.50	0.77	0.75	0.74	0.80	0.80	0.81
	β	1.06	1.49	0.07	1.15	1.03	0.99	1.13	1.13	1.14
	γ	0.84	0.83	0.68	0.97	0.80	0.88	0.58	0.60	0.62
	RMSE [mm]	30.73	43.44	33.99	30.30	36.57	35.72	27.84	27.18	26.11
	KGE'	0.29	0.15	-0.18	0.57	0.03	0.20	0.33	0.40	0.35
	r	0.61	0.44	0.39	0.75	0.34	0.37	0.56	0.59	0.70
	β	1.02	0.93	0.07	1.04	0.96	0.90	1.00	1.00	1.00
	γ	0.55	0.50	1.26	0.87	0.63	0.73	0.53	0.59	0.54
	RMSE [mm]	179.74	291.40	718.69	201.39	255.63	251.90	86.08	84.65	85.03

Dentro de los productos determinísticos, RFPUNA resalta en todas las desagregaciones del KGE', sobre todo en el cálculo del sesgo y la variabilidad. El mejor desempeño lo alcanza a escalas mensuales, donde se reportan valores de mediana cercanos al 0,5, posicionándose por sobre los demás productos determinísticos analizados.

Siguiendo la evaluación, se obtienen las métricas de registros de eventos de precipitación a distintas intensidades en la Figura 6.5, dados por las fórmulas e indicaciones de las tablas Tabla 5.1 y Tabla 5.2.

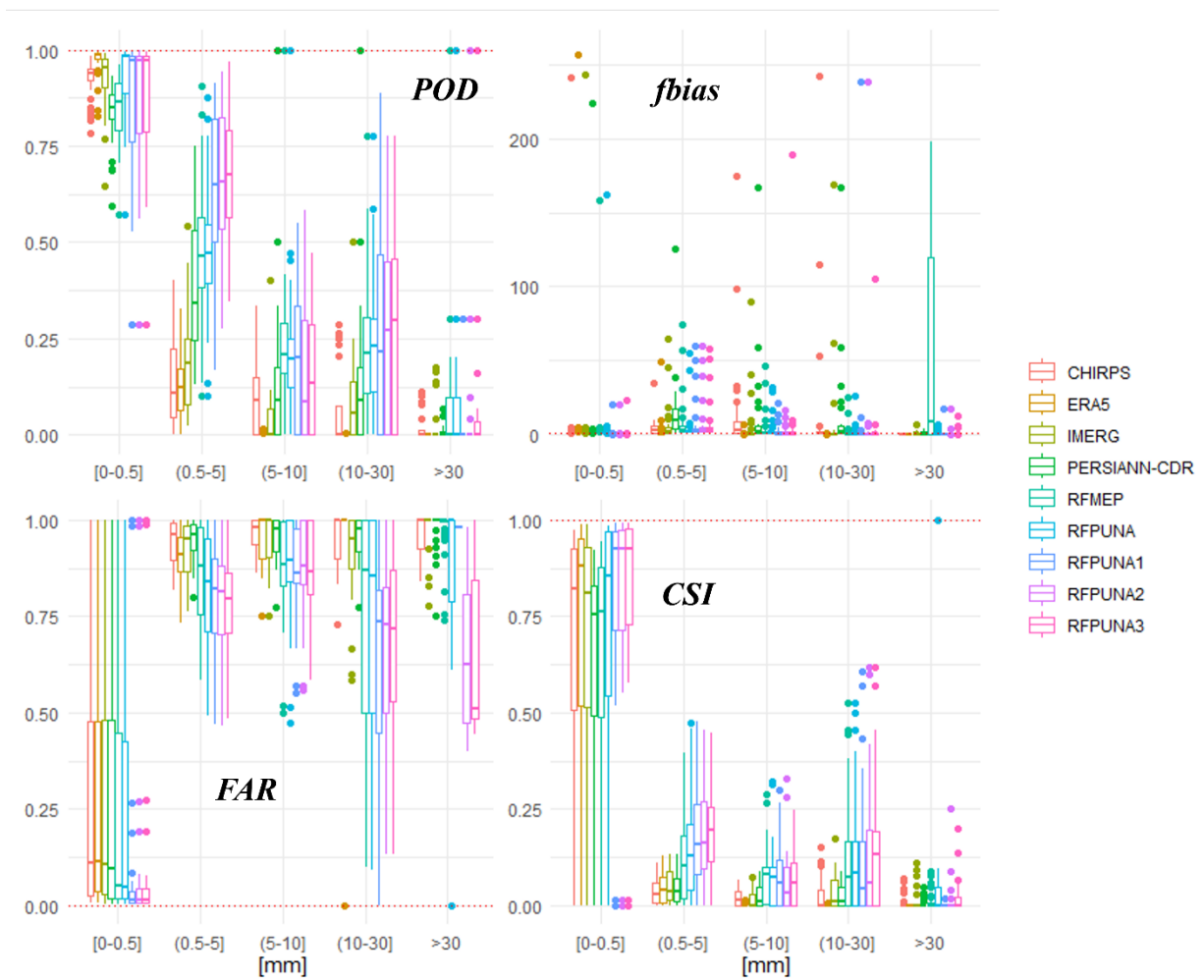


Figura 6.5: Métricas de los productos grillados de precipitación para el periodo 2007-2017.

En la Figura 6.5, se observa que los productos relativos a RFPUNA obtienen exactamente los mismos valores representados por la curva rosada de RFPUNA3. Otro aspecto que destaca es la frecuencia de sesgo (*fbias*) de los productos de CHIRPS e IMERG a escala diaria, y de ERA5 a escala mensual, las cuales se acercan al valor óptimo. Con respecto a ERA5, se observa que los valores de falsas alarmas (FAR) convergen al óptimo para valores más intensos de precipitación. Por último, no se observa una mejora sistemática de RFPUNA respecto al resto de los productos.

Para comprender la validez espacial de los resultados, en la Figura 6.6 se calcula el coeficiente de variación estacional con los resultados de cada producto en la zona. En azul, se observan estaciones que concentran más de la mitad de los valores anuales de precipitación en los meses de abril a septiembre; en verde si concentran las lluvias en los meses de octubre a marzo; y en rojo si falta información para efectuar el cálculo. Para complementar el análisis, se agrega de

fondo la clasificación climática *Köppen-Geiger* del área de estudio, representado por colores según el tipo de clima presente.

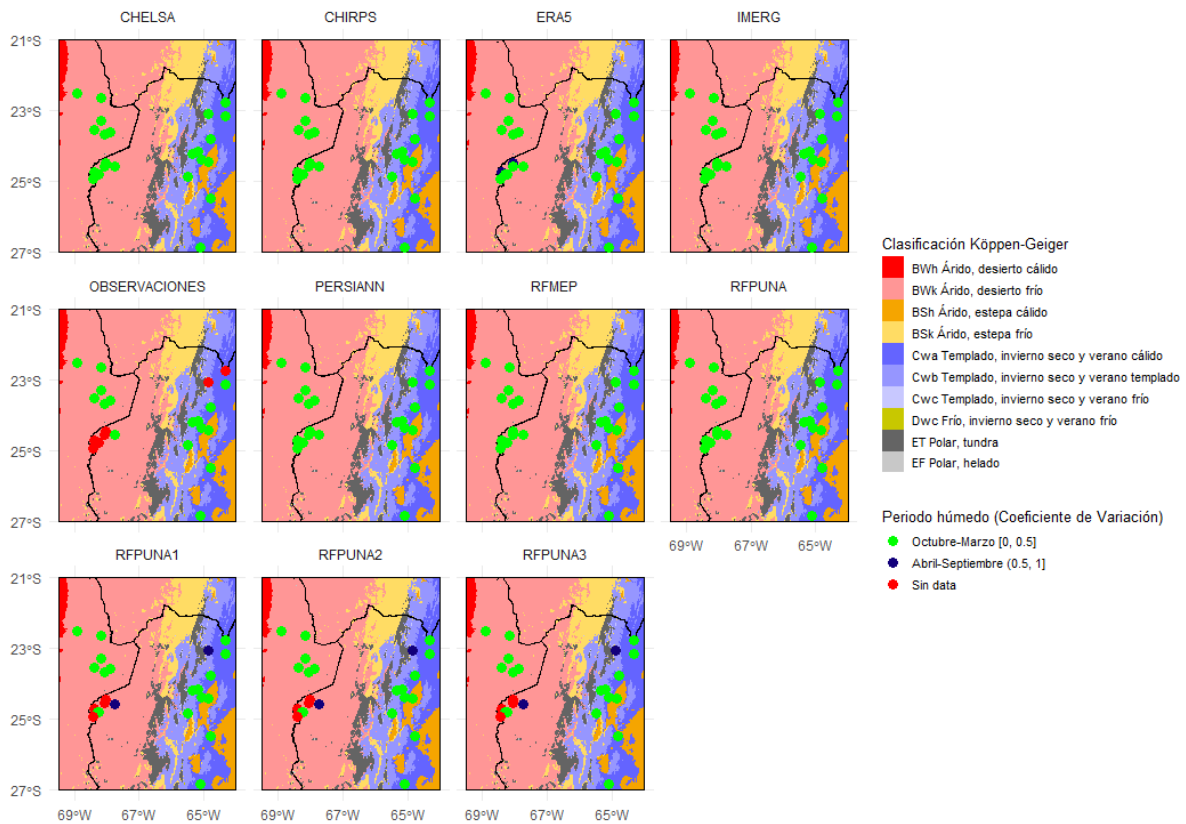


Figura 6.6: Coeficientes de variación estacional y distribución espacial de cada producto junto a la clasificación Köppen-Geiger del área de estudio.

Se observa un predominio de precipitaciones en los meses de octubre a marzo. De hecho, solo ERA5 y los productos estocásticos de RFPUNA encuentran puntos azules en el área, pero ninguno de ellos hace lógica con la clasificación climática.

A modo de presentar un análisis de la distribución espacial del producto en la extensión del área de estudio, la Figura 6.7 resume según diferentes agregaciones temporales la distribución espacial de la precipitación en el periodo 2007-2017. Allí se observan patrones espaciales similares entre productos, con valores de precipitación mayores al sureste del área. Adicionalmente, se observa que RFMEP entrega resultados similares a RFPUNA, lo cual se puede atribuir a las similitudes metodológicas y las utilización de las mismas estaciones para su calibración.

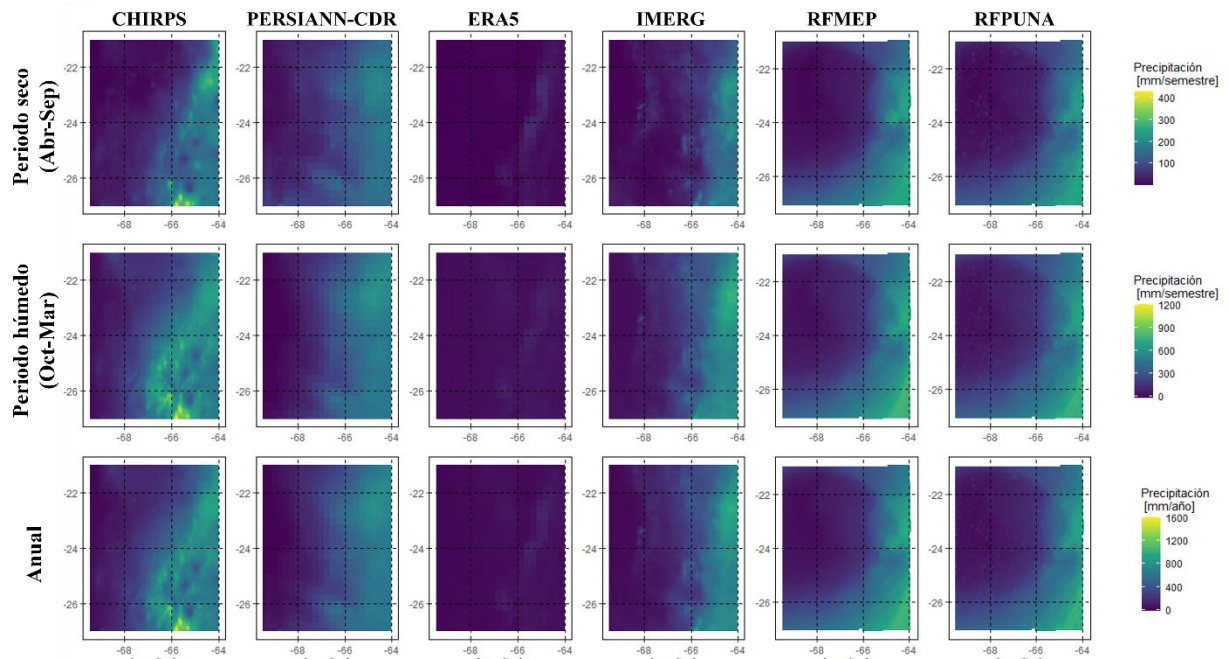


Figura 6.7: Distribución espacial de las precipitaciones simuladas por los distintos productos grillados.

Tomando en cuenta que existen algunas estaciones que no se usaron para entrenar el modelo, se utilizan para efectos de verificar las habilidades predictivas del producto. Para ello, se definen las bases de datos de entrenamiento y validación. El entrenamiento cuenta con las 39 estaciones según las cuales se construye el modelo, mientras que el set de validación se construye a partir de las estaciones restantes que presentan información meteorológica en al menos una fracción del periodo analizado. Los resultados de la verificación del producto RFPUNA, así como una evaluación del resto de los productos de referencia, se presentan en la Figura 6.8. En la figura se presentan, desde el rojo al azul, las estaciones de peor a mejor desempeño predictivo, y en negro, las estaciones de entrenamiento. De allí se desprende que los productos CHELSA, CHIRPSv2, PERSIANN-CDR, ERA5 e IMERG no presentan un buen desempeño, en contraste a lo exhibido por RF-MEP y RFPUNA.

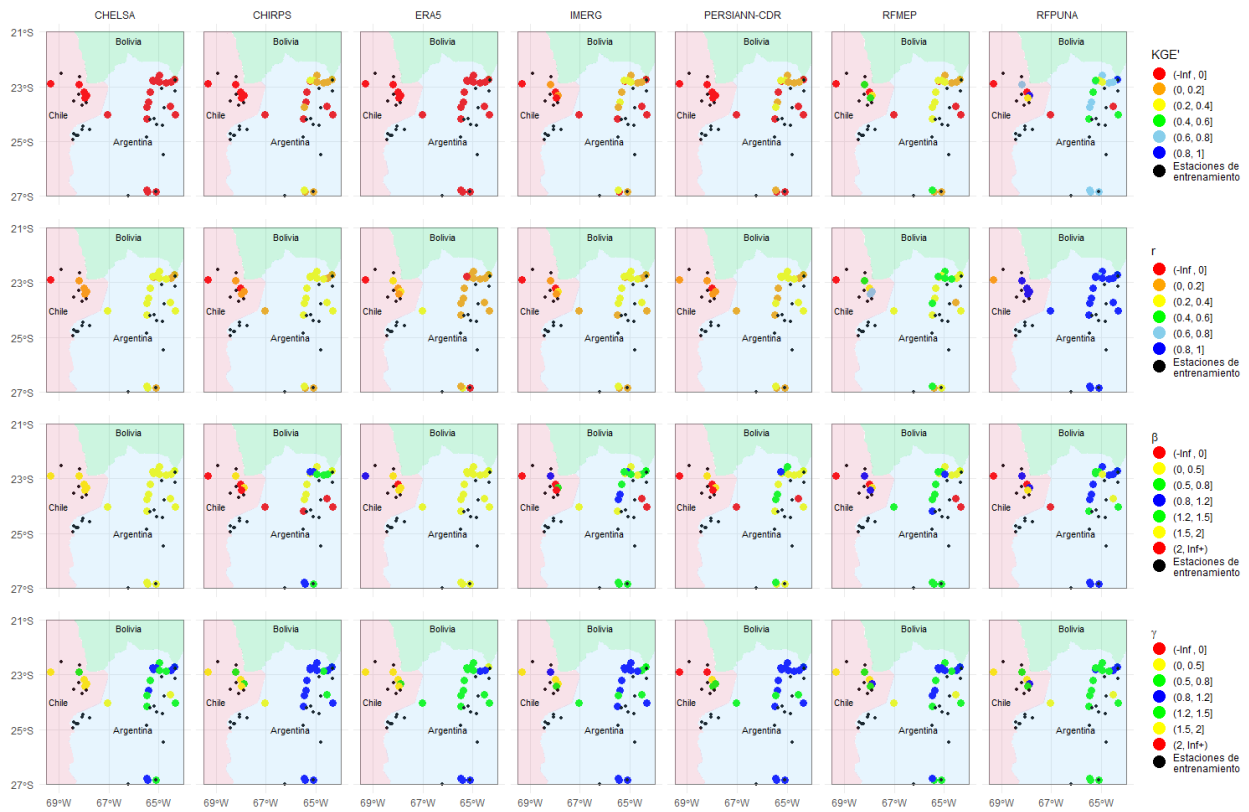


Figura 6.8: Distribución espacial de los valores KGE' para distintos productos grillados en el área de estudio.

Por último, se grafican Curvas de Variación Estacional (CVE) de cada estación para 4 distintas probabilidades de excedencia, conjunto al mejor KGE' obtenido, ya sea en escala diaria, mensual o anual. Las probabilidades fueron ajustadas mediante la distribución *Weibull* como una primera aproximación para observar el comportamiento del producto RFPUNA en comparación a otra distribución conocida. En la Figura 6.9 se observan dos de las estaciones, que sirven de ejemplo de una estación con buenos resultados para RFPUNA (Estación Ayquina, con un $KGE' = 0,96$ a escala anual) y una estación de mal desempeño para el producto (Estación Ollague, con un $KGE' = -1,32$ a escala anual), ilustrando el ajuste de las CVE y la estacionalidad de las simulaciones. En el ANEXO D: CURVAS DE VARIACIÓN ESTACIONAL se encuentran las CVE para las demás estaciones.

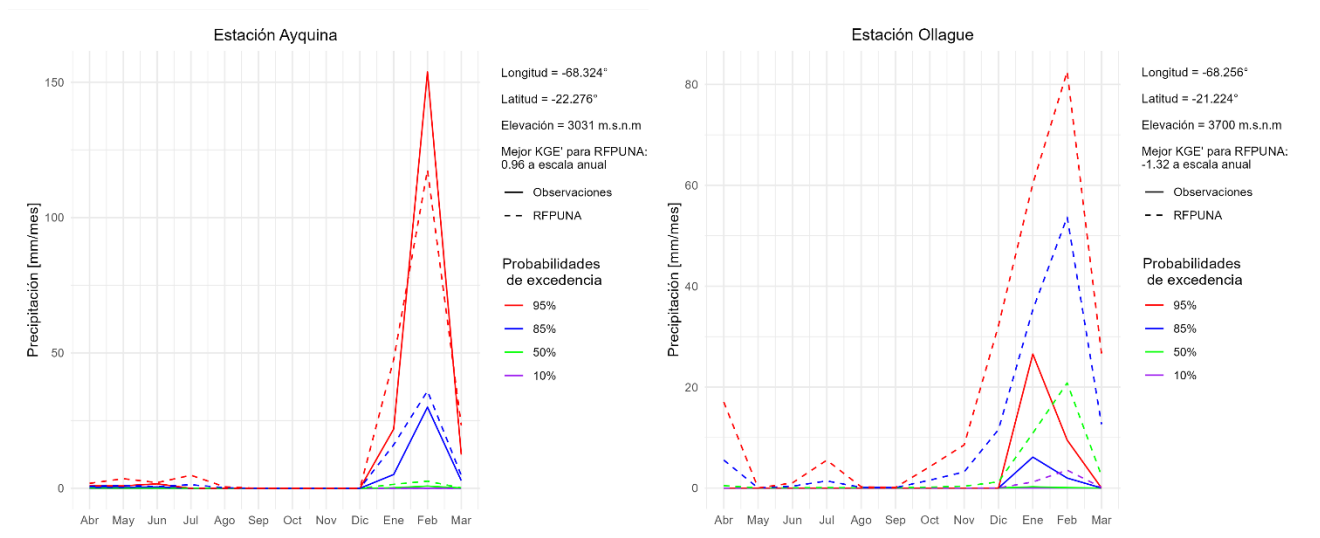


Figura 6.9: Curvas de variación estacional para las estaciones Ayquina y Ollague para 4 distintas probabilidades de excedencia para el periodo 2007-2017.

6.2.2 Temperatura

De manera análoga al análisis realizado para el caso de la precipitación, se calcula el KGE' de los productos de temperatura, el cual se presenta en la Figura 6.10 para los cuatro productos asociados a RFPUNA y los productos considerados como referencia (*i.e.*, CHELSA y ERA5). Cabe mencionar que el análisis se presenta para el KGE' y sus componentes según tres agregaciones temporales: diaria, mensual y anual.

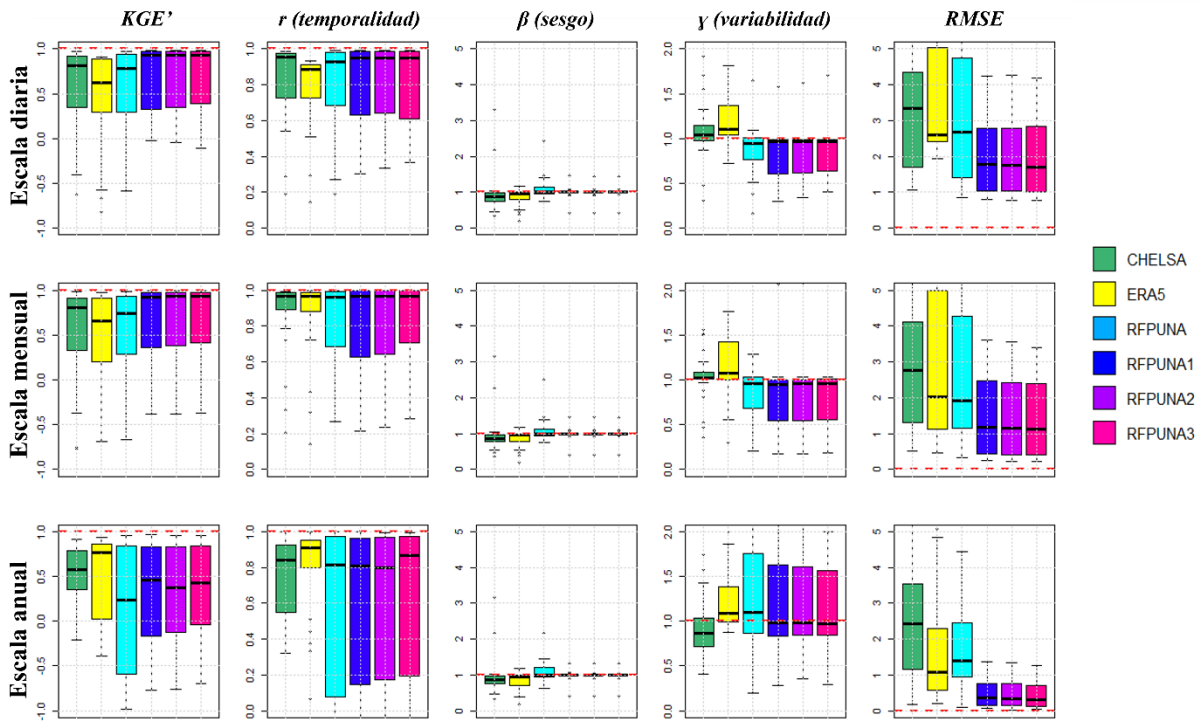


Figura 6.10: Gráficos de caja de productos grillados de temperatura y su KGE' con los parámetros r , β y γ para el periodo 2007-2016 en toda el área de estudio.

Al contrario de lo observado con la variable de precipitación de RFPUNA, para la temperatura no existe una mejora significativa con relación a los productos existentes. Lo anterior queda en evidencia al revisar los resultados presentados en la **¡Error! No se encuentra el origen d e la referencia.**, desde donde se desprende que la versión determinística de RFPUNA no destaca en ningún índice para ninguna agregación temporal. Similar al producto de temperatura, los gráficos de caja por territorio se encuentran en ANEXO A: RESULTADOS DE LA CALIBRACIÓN Y VERIFICACIÓN DEL MODELO

Tabla 6.3: Mediana de los resultados de los índices de KGE' para los distintos productos de temperatura⁶.

		CHELSA	ERA5	RFPUNA	RFPUNA1	RFPUNA2	RFPUNA3	
Todo	Día	KGE'	0.81	0.62	0.78	0.93	0.93	0.93
		r	0.95	0.89	0.93	0.95	0.95	0.95
		β	0.86	0.95	1.00	1.00	1.00	1.00
		γ	1.04	1.11	0.94	0.96	0.97	0.97
		RMSE [°C]	3.33	2.59	2.68	1.76	1.74	1.70
	Mes	KGE'	0.81	0.66	0.74	0.93	0.93	0.93
		r	0.97	0.97	0.96	0.96	0.96	0.97
		β	0.86	0.95	0.99	1.00	1.00	1.00
		γ	1.02	1.08	0.95	0.95	0.95	0.95

⁶ Se destacan en negrita los mejores valores por parámetro.

Chile	Año	RMSE [°C]	2.76	2.03	1.92	1.17	1.15	1.12	
		KGE'	0.57	0.77	0.23	0.45	0.38	0.43	
		<i>r</i>	0.84	0.91	0.82	0.81	0.80	0.87	
		β	0.86	0.95	0.99	1.00	1.00	1.00	
		γ	0.87	1.09	1.10	0.98	0.98	0.97	
		RMSE [°C]	2.42	1.09	1.40	0.35	0.34	0.32	
	Día	Día	KGE'	-0.30	0.09	-0.07	0.25	0.26	0.31
			<i>r</i>	0.67	0.63	0.44	0.53	0.55	0.57
			β	0.73	0.62	0.94	1.00	1.00	1.00
			γ	1.14	1.71	0.88	0.51	0.53	0.56
			RMSE [°C]	5.81	4.46	5.25	2.78	2.79	2.83
		Mes	KGE'	-0.37	-0.28	-0.26	0.15	0.16	0.22
			<i>r</i>	0.72	0.72	0.41	0.61	0.63	0.65
			β	0.79	0.68	0.94	1.00	1.00	1.00
			γ	0.81	1.60	0.59	0.40	0.41	0.42
RMSE [°C]			5.79	7.02	5.84	3.33	3.26	3.09	
Año		KGE'	0.34	-0.24	-0.91	-0.19	-0.15	0.04	
		<i>r</i>	0.54	0.44	0.61	0.60	0.61	0.62	
	β	0.78	0.61	0.96	1.01	1.01	1.01		
	γ	1.00	1.87	1.76	2.12	2.09	1.84		
	RMSE [°C]	4.03	2.14	3.38	0.89	0.87	0.83		
Argentina	Día	KGE'	0.88	0.85	0.89	0.96	0.96	0.96	
		<i>r</i>	0.96	0.91	0.97	0.97	0.97	0.97	
		β	0.89	0.96	1.00	1.00	1.00	1.00	
		γ	1.03	1.10	0.94	0.97	0.97	0.97	
		RMSE [°C]	5.70	6.28	4.47	5.90	5.90	5.91	
	Mes	KGE'	0.88	0.90	0.91	0.96	0.96	0.96	
		<i>r</i>	0.99	0.99	0.99	0.99	0.99	0.99	
		β	0.89	0.96	1.00	1.00	1.00	1.00	
		γ	1.04	1.07	0.96	0.97	0.97	0.97	
		RMSE [°C]	2.16	1.26	1.59	0.64	0.62	0.60	
	Año	KGE'	0.67	0.84	0.62	0.73	0.74	0.78	
		<i>r</i>	0.90	0.92	0.91	0.88	0.89	0.96	
		β	0.88	0.96	1.00	1.00	1.00	1.00	
		γ	0.87	1.08	1.02	0.96	0.96	0.95	
		RMSE [°C]	1.70	1.03	1.22	0.19	0.20	0.18	

Sin perjuicio de lo anterior, se observan mejoras al aplicar el algoritmo *Bluecat*, los cuales reportan los mejores resultados asociados a la mediana del conjunto de datos. Sin embargo, en comparación a los otros productos, en la Figura 6.10 se observa una dispersión importante en los gráficos de caja asociados a RFPUNA.

La distribución espacial de las temperaturas en la zona de estudio se grafica en la Figura 6.11. Allí, se observa una ausencia de valores bajos de temperatura para RFPUNA en comparación

a CHELSA o ERA5, otorgando valores bajo cero de temperaturas medias anuales para la zona más alta de la puna, ejercicio que nuestro producto no está realizando.

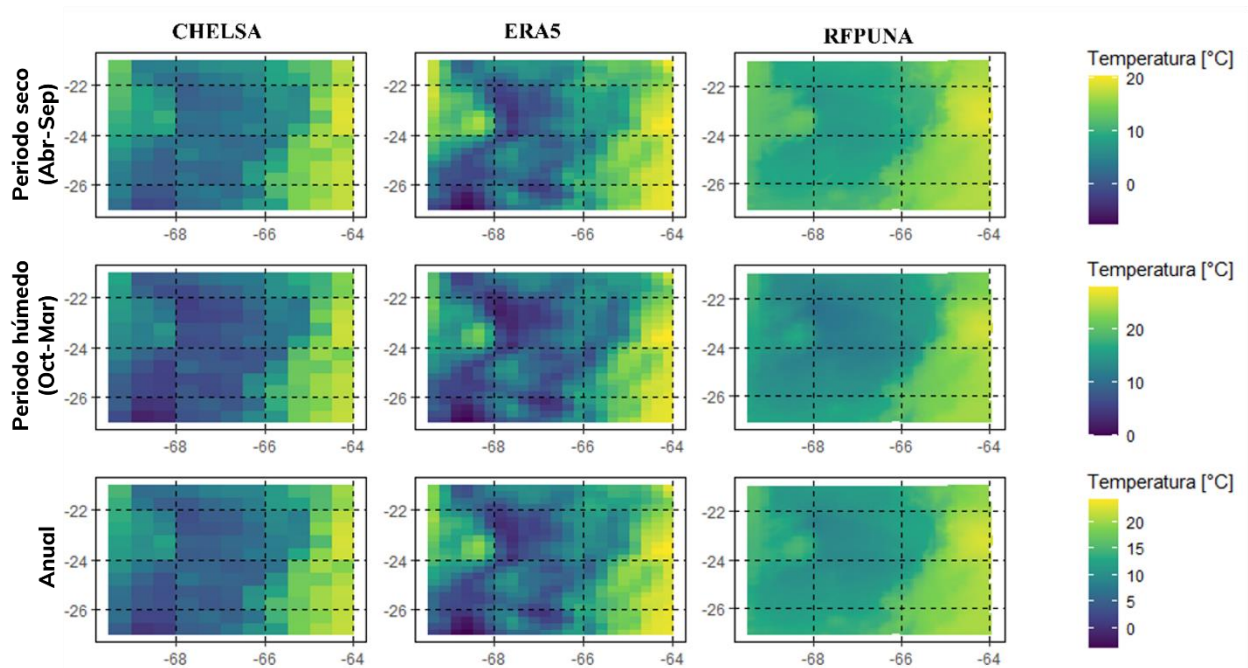


Figura 6.11: Distribución espacial de las temperaturas simuladas por los distintos productos grillados.

6.3 Resumen

Los resultados del producto RFPUNA de precipitación y temperatura se evalúan mediante diversas métricas, entre ellas, el KGE'. Por su parte, *Bluecat*, genera distintas series para una misma observación, por lo que se toma el promedio de KGE' de todas las simulaciones estocásticas para compararlos con los resultados determinísticos. A modo de resumir la información y el impacto que tiene la implementación de *Bluecat* en los resultados, la Figura 6.12 muestra sobre la línea 1:1 la relación entre el KGE' determinístico y estocástico. Allí se identifica, en general, una mejoría de las estimaciones con el uso de *Bluecat*. El gradiente de colores referencia la longitud de las observaciones, aproximadamente el 67.5°W divide las estaciones chilenas de las argentinas.

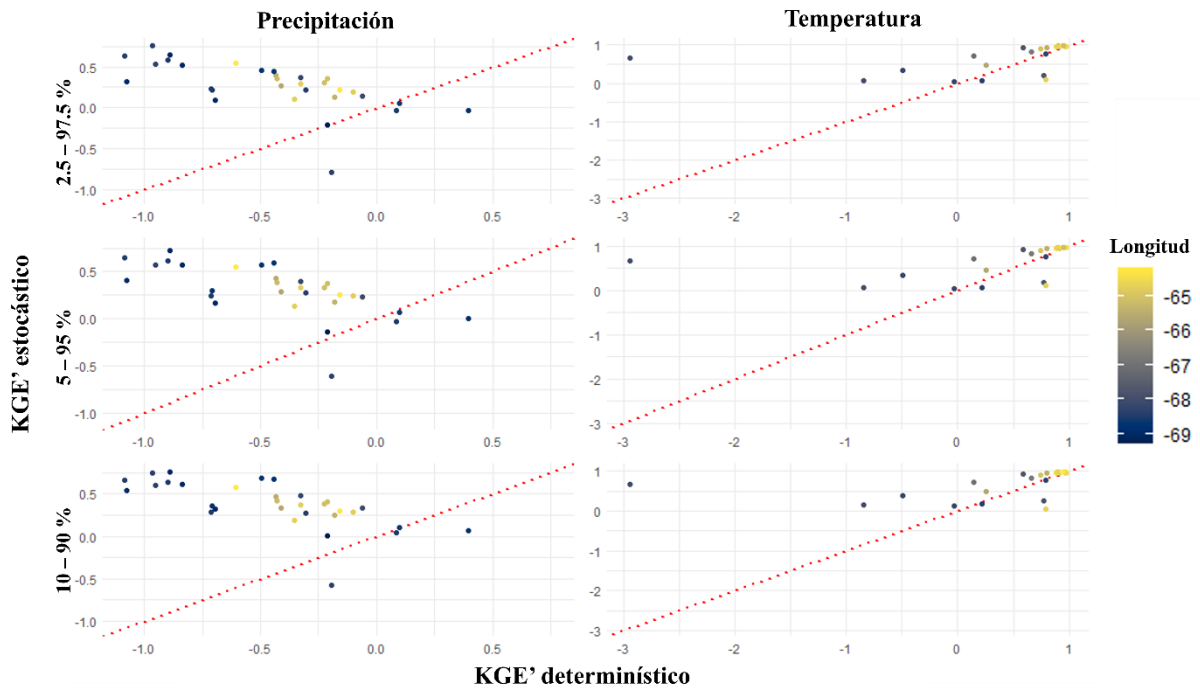


Figura 6.12: Comparación de KGE' determinístico contra el KGE' de la media estocástica según la longitud. Como referencia, la longitud -67.5^{a} W divide Chile de Argentina. La línea entrecortada roja representa la recta 1:1.

Con el objetivo de verificar los patrones espaciales del producto de precipitaciones y temperaturas RFPUNA, la *Tabla 6.4* reporta los índices de Morán obtenidos, usando como distancia el algoritmo *K-nearest neighbors* (KNN). Para precipitación no se observa autocorrelación, puesto que el p_{valor} es mayor al umbral de 0,05 y no es posible rechazar la hipótesis nula de que el conjunto tiene una distribución espacialmente aleatoria. Caso contrario para las temperaturas, donde sí es posible rechazar esta hipótesis. Sin embargo, las observaciones de temperatura son apenas menores al umbral, haciendo incierta la interpretación. De todos modos, se considera que existe una autocorrelación débil, al cumplir el criterio de significancia y contar con valores del índice de Z-score cercanos al 0, aun cuando el índice de Morán para las temperaturas es próximo a 1.

Tabla 6.4: Autocorrelación espacial según índice de Morán. Prec y Temp hacen referencia a las variables de precipitación y temperatura respectivamente.

Variable	Valor Índice de Morán	p_{valor} asociado	Z-score	¿Existe Autocorrelación?
RFPUNA: Prec	0,154	0,182	0,169	No
RFPUNA: Temp	0,948	3,62e-07	0,191	Sí, positiva y débil
Obs: Prec	-0,019	0,472	-0,271	No
Obs: Temp	0,372	0,048	0,223	Sí, positiva y débil

Finalmente, se presenta la distribución de las observaciones y simulaciones de RFPUNA y ERA5 para precipitación y temperatura en la Figura 6.13. Se observa una distribución de valores de precipitación dispar entre productos y observaciones, sobre todo ERA5, concentrando sus estimaciones en eventos menores a los 100 [mm], muy lejano a los promedios reales; en ese sentido, RFPUNA presenta estimaciones de precipitación más cercanas que ERA5

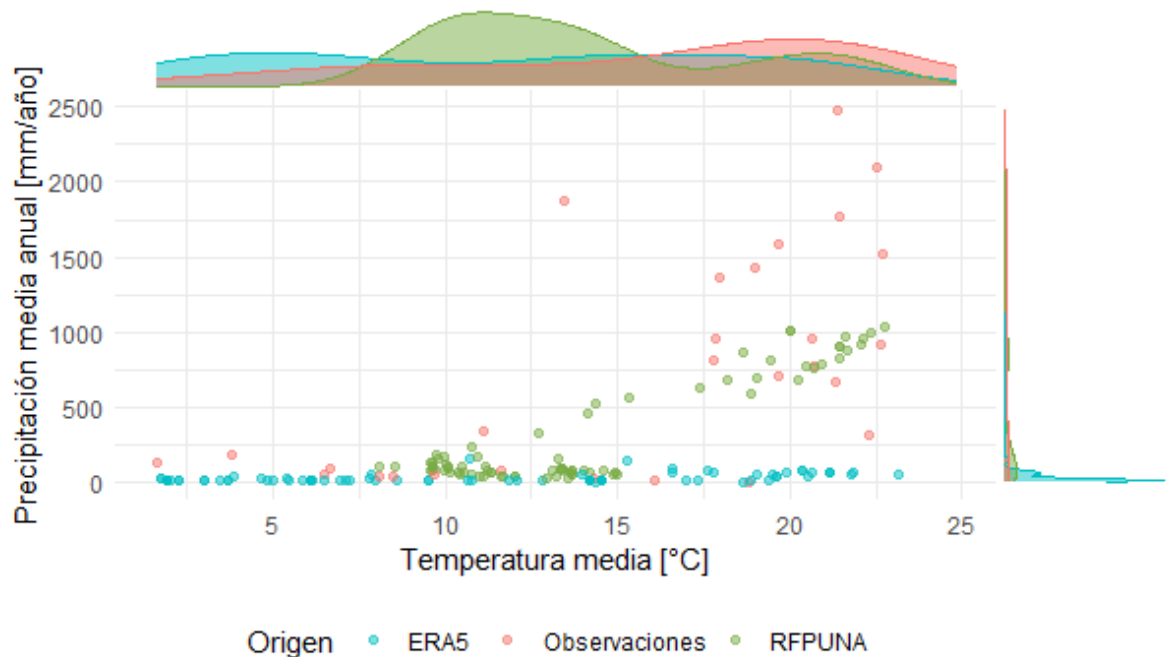


Figura 6.13: Distribución de precipitaciones y temperaturas medias anuales observadas y simuladas por el producto RFPUNA y ERA5 en el periodo 2007-2016.

Con respecto a las temperaturas, RFPUNA presenta distribuciones de valores intermedios entre lo simulado por ERA5 y lo observado. Además, tal como se observó en la Figura 6.11, RFPUNA no logra estimar las temperaturas bajo los 7°C aproximadamente.

7. DISCUSIONES Y COMENTARIOS

7.1 Alcances y Limitaciones del Producto

Al evaluar el desempeño y la aplicabilidad de un producto, es crucial considerar tanto sus alcances como sus limitaciones. Tomando en cuenta las dificultades de la zona de estudio para la estimación de precipitaciones, tanto RFPUNA como RFMEP, han demostrado ser productos más competentes en la captura de estos eventos que los otros productos analizados, especialmente en lo referente a temporalidad (0,44 y 0,43), sesgo (0,94 y 1,04) y variabilidad (0,75 y 0,65) respectivamente a escala diaria. De todas formas, las estimaciones están lejos de representar las precipitaciones con exactitud a escala diaria, solo a escala mensual se observa una mayor precisión en el producto propuesto (ver Figura 6.4 y **¡Error! No se encuentra el origen de la referencia.**).

Además, observando la Figura 6.8, se destaca el poder predictivo de ambos productos mencionados, ya que simularon precipitaciones con mayor precisión que los otros productos grillados, incluso en una proporción de 30-70 de validación y entrenamiento respectivamente, lo que es muy bajo para un *Validation Set Approach*.

Una de las principales limitaciones del producto se encuentra en las estimaciones de temperatura, acotándose a ofrecer resultados que, si bien son positivos, no logran superar a los productos de ERA5 ni CHELSA, con un KGE' anual de 0,23 contra los valores de 0,77 y 0,57 respectivamente. Sin embargo, se puede rescatar un buen desempeño a escala diaria (KGE'=0,78), mejor que ERA5 (KGE' = 0,62), pero no tan bueno como CHELSA (KGE' = 0,81).

Sin dudas, *Bluecat* es un gran avance en el desarrollo de estimaciones hidrológicas, cuantificando las incertidumbres de las simulaciones y entregando valores promedios que se ajustan incluso mejor a lo capturado por RFPUNA determinístico. No obstante, observando la Figura 6.12, se destaca una mejora sustancial en los valores de KGE' para las precipitaciones, aumentando cerca de una unidad los resultados determinísticos. Por su parte, las temperaturas no logran esa mejoría por tener mejores valores determinísticos. Sin embargo, en ambas variables se pueden obtener valores incluso peores a los determinísticos, lo que abre una discusión sobre el uso del algoritmo.

Otro hecho destacable son las diferencias de evaluación de los productos en territorio chileno y argentino, que es donde se concentran la mayoría de las estaciones. En Chile, el producto RFPUNA registra un KGE' anual de 0,6 en precipitación y -0,91 en temperatura, mientras que en Argentina se tienen valores de 0,26 y 0,62 respectivamente. El principal motivo de la gran diferencia en temperatura radica en la dispar disponibilidad de estaciones de esta variable entre ambos países, la evaluación y entrenamiento en Chile se realiza solo con 5 estaciones, las cuales resultan inadecuadas para cubrir un territorio tan extenso; por otra parte, Argentina cuenta con 33 estaciones de temperatura.

En resumen, este producto se destaca en la estimación de precipitaciones en comparación con otros similares, particularmente a nivel mensual, logrando un KGE' de 0,44, el más alto entre todos los productos determinísticos evaluados a diversas escalas temporales. Además, el producto RFPUNA demuestra ser eficiente en realizar predicciones de calidad con una cantidad limitada de

datos de entrenamiento. Sin embargo, su principal limitación reside en la estimación de temperaturas, donde no supera a otros productos grillados en el área de estudio; aun así, el producto tiene un mejor desempeño en Argentina a escala diaria y mensual. Cabe destacar también los avances logrados mediante el uso de *Bluecat*, que mejoran significativamente las estimaciones de precipitación y temperatura. No obstante, sigue siendo un alcance potencial en cuanto se logre distribuir la incertidumbre en toda la grilla.

7.2 Sobre el Potencial de Aplicaciones Hidrológicas

El producto propuesto presenta un mayor rendimiento respecto a la estimación de precipitaciones a escala diaria y mensual (KGE' de 0,25 y 0,44 respectivamente), mientras que a escala anual solo es superado por IMERG, con un KGE' de 0,40 contra un 0,32 de RFPUNA. Estos resultados permiten una mayor comprensión de la distribución espacial y temporal de las precipitaciones, pudiendo ofrecer múltiples oportunidades de mejoras en aplicaciones hidrológicas relacionadas a la variable climática. Con esto, se podrán estimar con mayor precisión los caudales superficiales y subterráneos en el área de estudio, y así, lograr configurar modelos hidrológicos robustos para realizar balances hídricos en la región.

Uno de estos modelos corresponde al modelo hidrológico de balance de energía y agua VIC (Liang et al., 1996; Wood et al., 1992). Este modelo fue utilizado en la actualización del Balance Hídrico Nacional chileno, analizando periodos históricos y proyectados. De la misma forma, se podrían realizar estos estudios en el área de estudio y poder predecir con mayor exactitud la disponibilidad y la distribución espacial y temporal del recurso hídrico.

Otras herramientas hidrológicas con un importante potencial de mejora son los modelos de cuenca *Soil and Water Assessment Tool* (SWAT) y MODFLOW para modelar la interacción entre aguas superficiales y subterráneas.

7.3 Trabajo Futuro

El producto presenta un gran potencial de mejora, aun con los resultados aceptables que se tienen con su configuración actual. En este contexto, el primer paso para explotar las oportunidades con RFPUNA corresponde a aumentar el número de estaciones de entrenamiento y verificación del modelo. Actualmente, la precipitación se entrena solo con 39 estaciones, mientras que la temperatura lo hace con 42. Contar con una mayor cantidad de estaciones, por ejemplo, privadas, apoyaría una actualización y potencial mejora del producto.

En este momento, el producto ha sido configurado con la mitad del total de estaciones disponibles en el periodo de estudio, esto debido a los esfuerzos computacionales necesarios para validar los resultados mediante la metodología de LOOCV. Por esto, se propone explorar un esquema de validación tipo *Validation Set Approach*, tal como se realizó en la Figura 6.8 y la calibración del umbral en ANEXO A: RESULTADOS DE LA CALIBRACIÓN Y VERIFICACIÓN DEL MODELO. Si bien, no se observan tan buenos resultados con este enfoque,

en comparación al LOOCV, lo más probable es que se deba a la desproporción entre el número de estaciones de entrenamiento y validación. Se sugiere una validación con el 30% de las estaciones, mientras el resto entrena.

Por último, queda pendiente el desafío de distribuir la incertidumbre cuantificada por *Bluecat* en toda la grilla. Lo anterior de manera tal de finalizar la construcción del producto RFPUNA estocástico. Para ello, se propone explorar en una primera instancia algunos métodos de interpolación geoestadística del tipo IDW, *Kriging*, o incluso desarrollar un *Random Forest* que distribuya los valores extremos de la incertidumbre.

8. CONCLUSIONES

El presente estudio tuvo por objetivo la elaboración de RFPUNA, un producto grillado local de precipitación y temperatura para la zona de la Puna Chile Argentina elaborado mediante *Random Forest*. Para ello, se recopilaron y filtraron estaciones meteorológicas del área; se descargaron distintos productos grillados y se evaluó el rendimiento en la Puna; se aplicaron herramientas de *Machine Learning* para estimar las variables climáticas en territorios de escasa información y topografía compleja; por último, se evaluó el ajuste y la verificación del producto RFPUNA.

Una de las ventajas del *Random Forest* es su autonomía para seleccionar como predictores las variables que disminuyen en mayor cantidad las impurezas del modelo. Por ello, no es necesario iterar el modelo variando variables de entrada, porque el algoritmo incorporará todo y decidirá, aunque esto implique un mayor esfuerzo computacional. De igual forma, siempre se pueden incorporar más variables al producto y ver cómo reacciona, siempre y cuando estas cuenten con una base científica-matemática que respalde su uso.

Con respecto a los resultados, se puede concluir que:

1. El producto de precipitación de RFPUNA desarrollado en la región de estudio, así como RFMEP, exhiben un desempeño que supera a los productos globales que se encuentran disponibles en la actualidad a escalas diarias y mensuales, con los mejores rendimientos de KGE' y RMSE. De acuerdo con la literatura revisada y las diferencias en la configuración de ambos modelos, se infiere que el empleo de métodos basados en *Random Forest* constituye una de las opciones más prometedoras para la estimación precisa de precipitaciones en áreas con topografía compleja y limitada disponibilidad de datos.
2. Las estimaciones de temperatura no presentan mejoras con respecto a los productos disponibles actualmente. En base a los resultados y la bibliografía, se puede sentar la hipótesis que las metodologías de *Random Forest* no son modelos óptimos para realizar estimaciones de temperatura.
3. La componente de precipitación de RFPUNA muestra un mejor desempeño en territorio chileno, destacando por sobre el resto de los productos globales; no así en Argentina, donde IMERG presenta mejor desempeño de KGE' a escala anual (0,57 contra 0,20 de RFPUNA), incluso mejor que las componentes estocásticas de RFPUNA. Caso contrario para la temperatura, en Chile el rendimiento es insuficiente por su variabilidad, siendo superado a escala anual por CHELSA, sin ser resultados satisfactorios (KGE' = 0,34); mientras que en Argentina todos los productos registran rendimientos satisfactorios a todas las escalas de tiempo, destacándose RFPUNA en escala diaria y mensual, con KGE' de 0,89 y 0,91.
4. La metodología *Bluecat* significa un gran avance y oportunidad para modelos determinísticos de precipitación. Sin embargo, queda pendiente determinar alguna fórmula efectiva que permita distribuir la incertidumbre en el área de estudio. Una de las

posibilidades son modelos geoestadísticos de *Kriging*, interpolaciones ponderadas tipo *Inverse Distance Weighted* (IDW) o la misma metodología de *Random Forest*

5. Los resultados fueron obtenidos en base al entrenamiento de pocas estaciones meteorológicas, lo que supone un potencial de mejora para la robustez del producto RFPUNA. Una vez conseguido el entrenamiento con todas las estaciones recopiladas se podrían obtener mejores conclusiones. De igual forma, se destaca el buen rendimiento de las estimaciones de precipitación, incluso con pocas estaciones de entrenamiento, con KGE' y RMSE superiores en escalas diarias (0,25 y 1,60 [mm]) y mensuales (0,44 y 14,88 [mm]).

Adicionalmente, existen algunos aspectos y hallazgos de interés derivados del estudio:

- La cantidad y calidad de las estaciones meteorológicas son insuficientes para un área tan extensa e importante económicamente.
- El desempeño de los productos grillados globales de precipitación en la zona de estudio no otorga garantías para su uso en proyectos hidrológicos. Incluso ERA5, producto global ampliamente utilizado, presenta un bajo desempeño ($KGE'_{\text{anual}} = -0,08$), realizando subestimaciones sistemáticas, como se observa en la Figura 6.13. Por otro lado, IMERG sí presenta un mejor rendimiento mensual y anual en el área de estudio, con KGE' de 0,20 y 0,40 respectivamente.
- Por otra parte, los productos de temperatura ERA5 y CHLSA presentan rendimientos notables en una zona de clima variable, logrando captar de forma correcta la magnitud y la variabilidad de las temperaturas en el área de estudio, incluso mejor que el producto propuesto (ver **¡Error! No se encuentra el origen de la referencia.**), abriendo la discusión sobre la aplicación de *Random Forest* para esta variable climática.
- El algoritmo *Bluecat* logra mejoras sustanciales en los resultados de precipitación y otras más leves para temperatura. No obstante, es incierto, pudiendo mejorar, mantener e incluso bajar el rendimiento con respecto a los valores determinísticos, como se observa en la Figura 6.12, donde hay un punto en Argentina (longitud aproximada de -64.5°W) con un KGE' determinístico de $-0,2$ y un KGE' estocástico de $-0,6$ en las precipitaciones con una banda de confianza de $2,5 - 97,5$.

Finalmente, el presente estudio desarrolla un producto de precipitación que presenta un mayor rendimiento respecto a los demás productos grillados globales estudiados a escala diaria y mensual. Esto permitiría una mayor comprensión de la distribución espacial y temporal de las precipitaciones en el área de estudio, pudiendo estimar con mayor precisión los caudales superficiales y subterráneos en el área de estudio. Por consecuencia, existen múltiples potenciales oportunidades de mejoras en aplicaciones hidrológicas relacionadas a precipitación y caudales, como la configuración de diversos modelos hidrológicos que capturen con mayor precisión la disponibilidad y la distribución espacial y temporal del recurso hídrico.

Toda la información y modelos usados para producir estos resultados incluidos en el estudio se encuentran públicos y disponibles en Zenodo (Gallegos 2024, <https://doi.org/10.5281/zenodo.10879989>)

BIBLIOGRAFÍA

- Abaurrea, J., Asín, J., Cebrián, A. C., & Centelles, A. (2004). Metodología para el control de calidad y homogeneidad de una base de datos de precipitación diaria.
- Allmendinger, R. W., Jordan, T. E., Kay, S. M., & Isacks, B. L. (1997). THE EVOLUTION OF THE ALTIPLANO-PUNA PLATEAU OF THE CENTRAL ANDES. *Annu. Rev. Earth Planet. Sci.*, 25, 139–174. www.annualreviews.org
- Arroyo, M. T. K., & Cavieres, L. A. (2013). High-Elevation Andean Ecosystems. *Encyclopedia of Biodiversity: Second Edition*, 96–110. <https://doi.org/10.1016/B978-0-12-384719-5.00428-7>
- Ashouri, H., Hsu, K. L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., Nelson, B. R., & Prat, O. P. (2015). PERSIANN-CDR: Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate Studies. *Bulletin of the American Meteorological Society*, 96(1), 69–83. <https://doi.org/10.1175/BAMS-D-13-00068.1>
- Aybar, C., Fernández, C., Huerta, A., Lavado, W., Vega, F., & Felipe-Obando, O. (2020). Construction of a high-resolution gridded rainfall dataset for Peru from 1981 to the present day. *Hydrological Sciences Journal*, 65(5), 770–785. <https://doi.org/10.1080/02626667.2019.1649411>
- Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Beck, H. E., McNamara, I., Ribbe, L., Nauditt, A., Birkel, C., Verbist, K., Giraldo-Osorio, J. D., & Xuan Thinh, N. (2020). RF-MEP: A novel Random Forest method for merging gridded precipitation products and ground-based measurements. *Remote Sensing of Environment*, 239, 111606. <https://doi.org/10.1016/J.RSE.2019.111606>
- Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Mendoza, P. A., Mcnamara, I., Beck, H. E., Thurner, J., Nauditt, A., Ribbe, L., & Thinh, N. X. (2021). On the selection of precipitation products for the regionalisation of hydrological model parameters. *Hydrol. Earth Syst. Sci.*, 25, 5805–5837. <https://doi.org/10.5194/hess-25-5805-2021>
- Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Ribbe, L., Nauditt, A., Giraldo-Osorio, J. D., & Thinh, N. X. (2018). Temporal and spatial evaluation of satellite rainfall estimates over different regions in Latin-America. *Atmospheric Research*, 213, 34–50. <https://doi.org/10.1016/J.ATMOSRES.2018.05.011>
- Beck, H. E., McVicar, T. R., Vergopolan, N., Berg, A., Lutsko, N. J., Dufour, A., Zeng, Z., Jiang, X., van Dijk, A. I. J. M., & Miralles, D. G. (2023). High-resolution (1 km) Köppen-Geiger maps for 1901–2099 based on constrained CMIP6 projections. *Scientific Data*, 10(1). <https://doi.org/10.1038/S41597-023-02549-6>
- Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., Van Dijk, A. I. J. M., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J., & Wood, E. F. (2017). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*, 21(12), 6201–6217. <https://doi.org/10.5194/HESS-21-6201-2017>

- Beck, S. L., & Zandt, G. (2002). The nature of orogenic crust in the central Andes. *Journal of Geophysical Research: Solid Earth*, 107(B10), ESE 7-1. <https://doi.org/10.1029/2000JB000124>
- Bennett, M., New, M., Marino, J., & Sillero-Zubiri, C. (2016a). *Climate complexity in the Central Andes: A study case on empirically-based local variations in the Dry Puna*. <https://doi.org/10.1016/j.jaridenv.2016.01.004>
- Bennett, M., New, M., Marino, J., & Sillero-Zubiri, C. (2016b). Climate complexity in the Central Andes: A study case on empirically-based local variations in the Dry Puna. *Journal of Arid Environments*, 128, 40–49. <https://doi.org/10.1016/J.JARIDENV.2016.01.004>
- Birkel, S. D., Mayewski, P. A., Perry, L. B., Seimon, A., & Andrade-Flores, M. (2022a). Evaluation of Reanalysis Temperature and Precipitation for the Andean Altiplano and Adjacent Cordilleras. *Earth and Space Science*, 9(3), e2021EA001934. <https://doi.org/10.1029/2021EA001934>
- Birkel, S. D., Mayewski, P. A., Perry, L. B., Seimon, A., & Andrade-Flores, M. (2022b). Evaluation of Reanalysis Temperature and Precipitation for the Andean Altiplano and Adjacent Cordilleras. *Earth and Space Science*, 9(3). <https://doi.org/10.1029/2021EA001934>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324/METRICKS>
- Burrough, P. A., & Mcdonnell, R. A. (1998). *Principles of Geographical Information Systems*.
- Chavez, S. P., & Takahashi, K. (2017). Orographic rainfall hot spots in the Andes-Amazon transition according to the TRMM precipitation radar and in situ data. *Journal of Geophysical Research*, 122(11), 5870–5882. <https://doi.org/10.1002/2016JD026282>
- Chen, J., Carlson, B. E., & Del Genio, A. D. (2002). Evidence for strengthening of the tropical general circulation in the 1990s. *Science*, 295(5556), 838–841. https://doi.org/10.1126/SCIENCE.1065835/SUPPL_FILE/1065835S3_THUMB.GIF
- Climate Prediction Center - ONI. (2023). Retrieved December 5, 2023, from https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php
- Climate reanalysis | ECMWF. (2023). Retrieved December 4, 2023, from <https://www.ecmwf.int/en/research/climate-reanalysis>
- Derin, Y., & Yilmaz, K. K. (2014). Evaluation of Multiple Satellite-Based Precipitation Products over Complex Topography. *Journal of Hydrometeorology*, 15(4), 1498–1516. <https://doi.org/10.1175/JHM-D-13-0191.1>
- Erazo, B., Bourrel, L., Frappart, F., Chimborazo, O., Labat, D., Dominguez-Granda, L., Matamoros, D., & Mejia, R. (2018). *Validation of Satellite Estimates (Tropical Rainfall Measuring Mission, TRMM) for Rainfall Variability over the Pacific Slope and Coast of Ecuador*. 10, 213. <https://doi.org/10.3390/w10020213>

- Estévez, J., Llabrés-Brustenga, A., Casas-Castillo, M. C., García-Marín, A. P., Kirchner, R., & Rodríguez-Solà, R. (2022). A quality control procedure for long-term series of daily precipitation data in a semiarid environment. *Theoretical and Applied Climatology*, *149*(3–4), 1029–1041. <https://doi.org/10.1007/S00704-022-04089-2/FIGURES/9>
- Falvey, M., & Garreaud, R. D. (2009). Regional cooling in a warming world: Recent temperature trends in the southeast Pacific and along the west coast of subtropical South America (1979–2006). *Journal of Geophysical Research: Atmospheres*, *114*(D4). <https://doi.org/10.1029/2008JD010519>
- Fernandez-Palomino, C. A., Hattermann, F. F., Krysanova, V., Lobanova, A., Vega-Jácome, F., Lavado, W., Santini, W., Aybar, C., & Bronstert, A. (2022). A Novel High-Resolution Gridded Precipitation Dataset for Peruvian and Ecuadorian Watersheds: Development and Hydrological Evaluation. *Journal of Hydrometeorology*, *23*(3), 309–336. <https://doi.org/10.1175/JHM-D-20-0285.1>
- Fleming, K., Awange, J. L., Kuhn, M., & Featherstone, W. E. (2011). Evaluating the TRMM 3B43 monthly precipitation product using gridded raingauge data over Australia. *Australian Meteorological and Oceanographic Journal*, *61*, 171–184. <http://people.eng.unimelb.edu.au/mpeel/koppen.html>
- Franzke, C. L. E., O’Kane, T. J., Berner, J., Williams, P. D., & Lucarini, V. (2015). Stochastic climate theory and modeling. *Wiley Interdisciplinary Reviews: Climate Change*, *6*(1), 63–78. <https://doi.org/10.1002/WCC.318>
- Getis, A. (2010). Spatial Autocorrelation. *Handbook of Applied Spatial Analysis*, 255–278. https://doi.org/10.1007/978-3-642-03647-7_14
- Ghalhari, G. F., Roudbari, A. D., & Asadi, M. (2016a). Identifying the spatial and temporal distribution characteristics of precipitation in Iran. *Arabian Journal of Geosciences*, *9*(12), 1–12. <https://doi.org/10.1007/S12517-016-2606-4/TABLES/4>
- Ghalhari, G. F., Roudbari, A. D., & Asadi, M. (2016b). Identifying the spatial and temporal distribution characteristics of precipitation in Iran. *Arabian Journal of Geosciences*, *9*(12), 1–12. <https://doi.org/10.1007/S12517-016-2606-4/TABLES/4>
- Gobierno de Argentina (s.f.). Tablero global del litio. Recuperado de <https://www.argentina.gob.ar/economia/mineria/siacam/tablero-global-del-litio>
- Handbook, F. (2005). Federal meteorological handbook no. 11. Doppler Radar Meteorological Observations, Part B, 265.
- Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., & New, M. (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research: Atmospheres*, *113*(D20), 20119. <https://doi.org/10.1029/2008JD010201>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo,

- G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., ... Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/QJ.3803>
- Hong, Y., Tang, G., Ma, Y., Huang, Q., Han, Z., Zeng, Z., Yang, Y., Wang, C., & Guo, X. (2018). *Remote Sensing Precipitation: Sensors, Retrievals, Validations, and Applications*. 1–23. https://doi.org/10.1007/978-3-662-47871-4_4-1
- Huffman, G. J., Adler, R. F., Bolvin, D. T., & Nelkin, E. J. (2010). The TRMM Multi-satellite Precipitation Analysis (TMPA). *Satellite Rainfall Applications for Surface Hydrology*, 3–22. https://doi.org/10.1007/978-90-481-2915-7_1/TABLES/2
- Jordan, T. E., Nester, P. L., Blanco, N., Hoke, G. D., Dávila, F., & Tomlinson, A. J. (2010). Uplift of the Altiplano-Puna plateau: A view from the west. *Tectonics*, 29(5). <https://doi.org/10.1029/2010TC002661>
- Jorquera, J., & Pizarro, A. (2023). Unlocking the potential of stochastic simulation through Bluecat: Enhancing runoff predictions in arid and high-altitude regions. *Hydrological Processes*, 37(12), e15046. <https://doi.org/10.1002/HYP.15046>
- Karbalaei Saleh, S., Amoushahi, S., & Gholipour, M. (2021). Spatiotemporal ecological quality assessment of metropolitan cities: a case study of central Iran. *Environmental Monitoring and Assessment*, 193(5), 1–20. <https://doi.org/10.1007/S10661-021-09082-2/FIGURES/8>
- Karger, D. N., Lange, S., Hari, C., Reyer, C. P. O., Conrad, O., Zimmermann, N. E., & Frieler, K. (2023). CHELSA-W5E5: Daily 1gkm meteorological forcing data for climate impact studies. *Earth System Science Data*, 15(6), 2445–2464. <https://doi.org/10.5194/ESSD-15-2445-2023>
- Khan, M. Y., Qayoom, A., Nizami, M. S., Siddiqui, M. S., Wasi, S., & Raazi, S. M. K. U. R. (2021). Automated Prediction of Good Dictionary EXamples (GDEX): A Comprehensive Experiment with Distant Supervision, Machine Learning, and Word Embedding-Based Deep Learning Techniques. *Complexity*, 2021. <https://doi.org/10.1155/2021/2553199>
- Koutsoyiannis, D., & Montanari, A. (2022a). Bluecat: A Local Uncertainty Estimator for Deterministic Simulations and Predictions. *Water Resources Research*, 58(1). <https://doi.org/10.1029/2021WR031215>
- Koutsoyiannis, D., & Montanari, A. (2022b). Climate Extrapolations in Hydrology: The Expanded Bluecat Methodology. *Hydrology 2022*, Vol. 9, Page 86, 9(5), 86. <https://doi.org/10.3390/HYDROLOGY9050086>
- Laviola, S., Levizzani, V., Cattani, E., & Kidd, C. (2013). The 183-WSL fast rain rate retrieval algorithm. Part II: Validation using ground radar measurements. *Atmospheric Research*, 134, 77–86. <https://doi.org/10.1016/J.ATMOSRES.2013.07.013>
- Li, J., Zhang, L., He, C., & Zhao, C. (2018). A Comparison of Markov Chain Random Field and Ordinary Kriging Methods for Calculating Soil Texture in a Mountainous Watershed, Northwest China. *Sustainability* 2018, Vol. 10, Page 2819, 10(8), 2819. <https://doi.org/10.3390/SU10082819>

- Liang, X., Wood, E. F., & Lettenmaier, D. P. (1996). Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification. *Global and Planetary Change*, 13(1-4), 195-206.
- Liu, S., Li, Y., Pauwels, V. R. N., & Walker, J. P. (2018). Impact of rain gauge quality control and interpolation on streamflow simulation: An application to the warwick catchment, Australia. *Frontiers in Earth Science*, 5, 302696. <https://doi.org/10.3389/FEART.2017.00114/BIBTEX>
- López Jiménez, V. L., Alejandra, M., Castro, R., Gineth, K., & Prieto, P. (2019). Propuesta para el mejoramiento de la red agrometeorológica en el departamento de Cundinamarca. *Publicaciones e Investigación*, 13(1), 63–78. <https://doi.org/10.22490/25394088.3266>
- López Steinmetz, R. L., & Salvi, S. (2021). Brine grades in Andean salars: When basin size matters A review of the Lithium Triangle. *Earth-Science Reviews*, 217, 103615. <https://doi.org/10.1016/J.EARSCIREV.2021.103615>
- Lucas, M. P., Longman, R. J., Giambelluca, T. W., Frazier, A. G., McLean, J., Cleveland, S. B., Huang, Y. F., & Lee, J. (2022). Optimizing Automated Kriging to Improve Spatial Interpolation of Monthly Rainfall over Complex Terrain. *Journal of Hydrometeorology*, 23(4), 561–572. <https://doi.org/10.1175/JHM-D-21-0171.1>
- Mahesh, B. (2018). Machine Learning Algorithms-A Review. *International Journal of Science and Research*. <https://doi.org/10.21275/ART20203995>
- Marazuela, M. A., Vázquez-Suñé, E., Ayora, C., García-Gil, A., & Palma, T. (2018). *Hydrodynamics of salt flat basins: The Salar de Atacama example*. <https://doi.org/10.1016/j.scitotenv.2018.09.190>
- Marconi, P., Arengo, F., & Clark, A. (2022). The arid Andean plateau waterscapes and the lithium triangle: flamingos as flagships for conservation of high-altitude wetlands under pressure from mining development. *Wetlands Ecology and Management*, 30(4), 827-852.
- Mcglashan, N., Brown, L., & Kay, S. (2008). Crustal thickness in the central Andes from teleseismically recorded depth phase precursors. *Geophysical Journal International*, 175(3), 1013–1022. <https://doi.org/10.1111/J.1365-246X.2008.03897.X/3/175-3-1013-FIG010.JPEG>
- Ministerio de Desarrollo Sostenible. (1994). Sub cuencas principales - Plan Nacional de Manejo de Cuencas Hidrográficas en Bolivia (PLAMACH-BOL). Recuperado de <https://demo.georchestra.org/geonetwork/srv/api/records/ed909c39-ae99-41a1-919c-07d235764efd>
- Ministerio de Economía y Finanzas Públicas de Argentina. (2024). Energía - Cuencas hídricas - COHIFE. Recuperado de <https://datos.gob.ar/dataset/energia-cuencas-hidricas---cohife>
- Ministerio de Obras Públicas de Chile. (s.f.). Inventario de cuencas y lagos. Recuperado de https://dga.mop.gob.cl/administracionrecursosohidricos/inventario_cuencas_lagos/Paginas/default.aspx
- Mohammed, J., Mengiste, Y., & P. Singh, V. (2023). Improving spatio-temporal precipitation estimates in data scarce river basins: an application of machine learning-based multi-source

- data merging. *Stochastic Environmental Research and Risk Assessment*, 37(4), 1353–1369. <https://doi.org/10.1007/S00477-022-02346-4/FIGURES/4>
- Musakwa, W., & Van Niekerk, A. (1 C.E.). Monitoring Urban Sprawl and Sustainable Urban Development Using the Moran Index: A Case Study of Stellenbosch, South Africa. <https://Services.Igi-Global.Com/Resolvedoi/Resolve.aspx?Doi=10.4018/Ijagr.2014070101>, 5(3), 1–20. <https://doi.org/10.4018/IJAGR.2014070101>
- Newman, A. J., Clark, M. P., Longman, R. J., & Giambelluca, T. W. (2019). Methodological Intercomparisons of Station-Based Gridded Meteorological Products: Utility, Limitations, and Paths Forward. *Journal of Hydrometeorology*, 20(3), 531–547. <https://doi.org/10.1175/JHM-D-18-0114.1>
- Nguyen, G. V., Le, X. H., Van, L. N., Jung, S., Yeon, M., & Lee, G. (2021). Application of Random Forest Algorithm for Merging Multiple Satellite Precipitation Products across South Korea. *Remote Sensing 2021, Vol. 13, Page 4033, 13(20)*, 4033. <https://doi.org/10.3390/RS13204033>
- Nurmi, P., & Department, O. (2003). *430 Recommendations on the verification of local weather forecasts*. <http://www.ecmwf.int/publications/library/ecpublications/>
- Roy, M. H., & Larocque, D. (2012). Robustness of random forests for regression. *Journal of Nonparametric Statistics*, 24(4), 993–1006. <https://doi.org/10.1080/10485252.2012.715161>
- Sarricolea, P., Dubreuil, V., Meseguer-Ruiz, O., Funatsu, B. M., Romero-Aravena, H., & Serrano-Notivoli, R. (2023). Precipitation extremes in the Puna of Atacama Desert, Chile: How to manage current and future uncertainty? (No. ART-2023-133736).
- Satgé, F., Hussain, Y., Molina-Carpio, J., Pillco, R., Laugner, C., Akhter, G., & Bonnet, M. P. (2021). Reliability of SM2RAIN precipitation datasets in comparison to gauge observations and hydrological modelling over arid regions. *International Journal of Climatology*, 41(S1), E517–E536. <https://doi.org/10.1002/JOC.6704>
- Seiler, C., Hutjes, R. W. A., & Kabat, P. (2013). Climate Variability and Trends in Bolivia. *Journal of Applied Meteorology and Climatology*, 52(1), 130–146. <https://doi.org/10.1175/JAMC-D-12-0105.1>
- Sharifi, E., Saghafian, B., & Steinacker, R. (2019). Copula-based stochastic uncertainty analysis of satellite precipitation products. *Journal of Hydrology*, 570, 739–754. <https://doi.org/10.1016/J.JHYDROL.2019.01.035>
- Shi, J. ;, Bao, Z. ;, Zhang, J. ;, Liu, C. ;, Jin, J. ;, Liu, Y. ;, Wang, G., Shi, J., Bao, Z., Zhang, J., Liu, C., Jin, J., Liu, Y., & Wang, G. (2022). Comparison of the Performance of IMERG Products and Interpolation-Based Precipitation Estimates in the Middle Reaches of Yellow River Basin. *Water 2022, Vol. 14, Page 1503, 14(9)*, 1503. <https://doi.org/10.3390/W14091503>
- Shi, Y., Chen, C., Chen, J., Mohammadi, B., Cheraghalizadeh, M., Abdallah, M., Mert Katipoğlu, O., Li, H., & Duan, Z. (2023). Evaluation of the RF-MEP Method for Merging Multiple Gridded Precipitation Products in the Chongqing City, China. *Remote Sensing 2023, Vol. 15, Page 4230, 15(17)*, 4230. <https://doi.org/10.3390/RS15174230>

- Strecker, M. R., Mulch, A., Uba, C., Schmitt, A. K., Chamberlain, C., Strecker, M. R., Mulch, A., Uba, C., Schmitt, A. K., & Chamberlain, C. (2006). Late Miocene onset of the South American Monsoon. *AGUFM*, 2006, T31E-06. <https://ui.adsabs.harvard.edu/abs/2006AGUFM.T31E..06S/abstract>
- Sun, H., Su, F., Yao, T., He, Z., Tang, G., Huang, J., Zheng, B., Meng, F., Ou, T., & Chen, D. (2021). General overestimation of ERA5 precipitation in flow simulations for High Mountain Asia basins. *Environmental Research Communications*, 3(12), 121003. <https://doi.org/10.1088/2515-7620/AC40F0>
- Torres-Batló, J., & Martí-Cardona, B. (2020). Precipitation trends over the southern Andean Altiplano from 1981 to 2018. *Journal of Hydrology*, 590, 125485. <https://doi.org/10.1016/J.JHYDROL.2020.125485>
- Trumbull, R. B., Riller, U., Oncken, O., Scheuber, E., Munier, K., & Hongn, F. (2006). The Time-Space Distribution of Cenozoic Volcanism in the South-Central Andes: a New Data Compilation and Some Tectonic Implications. *The Andes*, 29–43. https://doi.org/10.1007/978-3-540-48684-8_2
- U.S. Geological Survey (2023). Lithium Statics and Information. Recuperado de <https://pubs.usgs.gov/periodicals/mcs2023/mcs2023-lithium.pdf>
- Wang, C., Tang, G., Han, Z., Guo, X., & Hong, Y. (2018). Global intercomparison and regional evaluation of GPM IMERG Version-03, Version-04 and its latest Version-05 precipitation products: Similarity, difference and improvements. *Journal of Hydrology*, 564, 342–356. <https://doi.org/10.1016/j.jhydrol.2018.06.064>
- Wang, Z., Zhong, R., Lai, C., & Chen, J. (2017). Evaluation of the GPM IMERG satellite-based precipitation products and the hydrological utility. *Atmospheric Research*, 196, 151–163. <https://doi.org/10.1016/J.ATMOSRES.2017.06.020>
- Whitman, D., Isacks, B. L., & Kay, S. M. (1996). Lithospheric structure and along-strike segmentation of the Central Andean Plateau: seismic Q, magmatism, flexure, topography and tectonics. *Tectonophysics*, 259(1–3), 29–40. [https://doi.org/10.1016/0040-1951\(95\)00130-1](https://doi.org/10.1016/0040-1951(95)00130-1)
- Woldemeskel, F. M., Sivakumar, B., & Sharma, A. (2013). Merging gauge and satellite rainfall with specification of associated uncertainty across Australia. *Journal of Hydrology*, 499, 167–176. <https://doi.org/10.1016/J.JHYDROL.2013.06.039>
- Wood, E. F., Lettenmaier, D. P., & Zartarian, V. G. (1992). A land-surface hydrology parameterization with subgrid variability for general circulation models. *Journal of Geophysical Research: Atmospheres*, 97(D3), 2717–2728.
- Xie, P., & Arkin, P. A. (1996). Analyses of Global Monthly Precipitation Using Gauge Observations, Satellite Estimates, and Numerical Model Predictions. *JCLI*, 9(4), 840–858. [https://doi.org/10.1175/1520-0442\(1996\)009](https://doi.org/10.1175/1520-0442(1996)009)

- Yürekli, K., Simsek, H., Cemek, B., & Karaman, S. (2007). Simulating climatic variables by using stochastic approach. *Building and Environment*, 42(10), 3493–3499. <https://doi.org/10.1016/J.BUILDENV.2006.10.046>
- Zambrano-Bigiarini, M.; Baez-Villanueva, O.M., Giraldo-Osorio, J. RFmerge: Merging of Satellite Datasets with Ground Observations using Random Forests. R package version 0.1-5. URL <https://hzambran.github.io/RFmerge/>. doi:10.5281/zenodo.3581515.
- Zhang, L., Li, X., Zheng, D., Zhang, K., Ma, Q., Zhao, Y., & Ge, Y. (2021). Merging multiple satellite-based precipitation products and gauge observations using a novel double machine learning approach. *Journal of Hydrology*, 594, 125969. <https://doi.org/10.1016/J.JHYDROL.2021.125969>

ANEXO A: RESULTADOS DE LA CALIBRACIÓN Y VERIFICACIÓN DEL MODELO

A.1 Resultados de Calibración

Para optimizar el resultado de RFPUNA, se debe calibrar el umbral de incertidumbre, el cual define sobre qué monto de agua sobre la grilla se puede considerar un evento de precipitación y cuándo no. Se definió un mínimo de 0.01 [mm] y un máximo de 1 [mm] para hacer la evaluación; entre estos, se definieron 18 valores equidistantes, tabulados en Tabla A.1.

Tabla A.1: Referencias para identificar los productos asociados a los umbrales de calibración de la Figura A.2

Nombre	ID	Valor umbral [mm]	Color
RFPUNA	RFPUNA	-	#4E79A7
RFPUNA1	R1	0,01	#A0CBE8
RFPUNA2	R2	0,06	#F28E2B
RFPUNA3	R3	0,11	#FFBE7D
RFPUNA4	R4	0,17	#59A14F
RFPUNA5	R5	0,22	#8CD17D
RFPUNA6	R6	0,27	#B6992D
RFPUNA7	R7	0,32	#F1CE63
RFPUNA8	R8	0,37	#499894
RFPUNA9	R9	0,43	#86BCB6
RFPUNA10	R10	0,48	#E15759
RFPUNA11	R11	0,53	#FF9D9A
RFPUNA12	R12	0,58	#79706E
RFPUNA13	R13	0,64	#BAB0AC
RFPUNA14	R14	0,69	#D37295
RFPUNA15	R15	0,74	#FABFD2
RFPUNA16	R16	0,79	#B07AA1
RFPUNA17	R17	0,84	#D4A6C8
RFPUNA18	R18	0,9	#9D7660
RFPUNA19	R19	0,95	#D7B5A6
RFPUNA20	R20	1	#4E79A7

A partir de estos, se calcula el KGE' de las 20 simulaciones más el producto sin matriz de intermitencia, denominado RFPUNA a secas, en la Figura A.1. Luego, se calculan los demás índices en la Tabla A.2

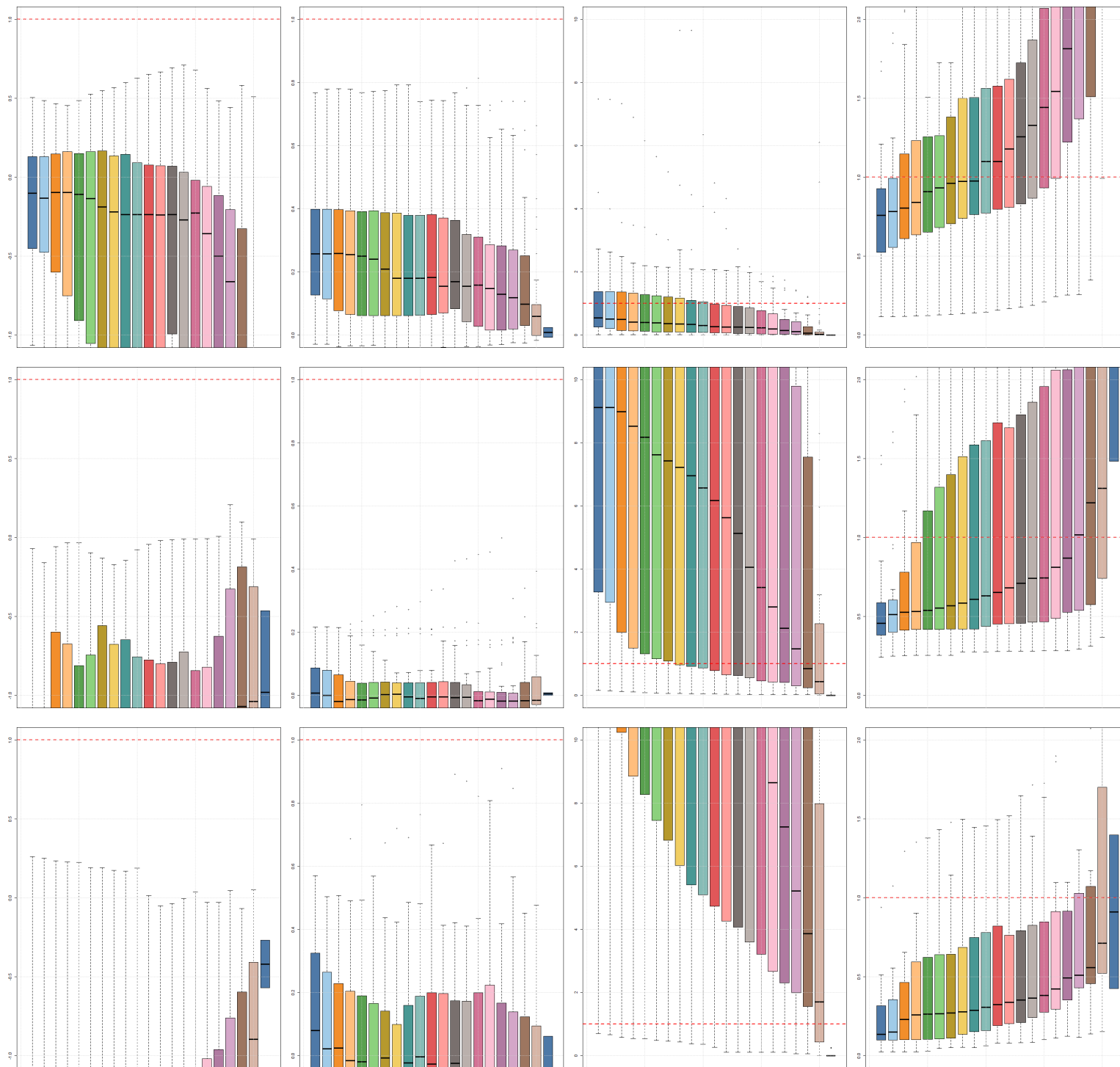


Figura A.1: Calibración del valor umbral de RFPUNA mediante KGE' con los parámetros r , β y γ en el periodo 2007-2017. La línea roja marca el valor óptimo de cada parámetro y la línea negra de los gráficos de caja indica el valor mediano. La leyenda de los productos está disponible en la Tabla A.1.

Tabla A.2: Mejor producto de intermitencia para distintos índices y magnitudes de lluvia.

Índice	[0,1)	[1,5)	[5,10)	[10,30)	>30
POD	R19	R1	R19	R19	Todos
<i>fbias</i>	RFPUNA	R10	R10	R10	RFPUNA
FAR	R2	R10	R10	R20	RFPUNA
CSI	R19	R10	R10	R20	RFPUNA

A.2 Resultados de Verificación

En esta sección se adjuntan resultados más detallados y específicos de lo presentado en la sección de RESULTADOS. En general, se presentan los valores tabulados del promedio de KGE' de los productos grillados y los gráficos de caja de estos para distintas agregaciones temporales, separando por semestres de más lluvia en la zona (de octubre a marzo) y los meses más secos (abril a septiembre). Además, se visualizan los resultados por territorio estudiado, analizando el rendimiento de las estaciones por país, sin considerar Bolivia debido a contar solo con 3 estaciones.

A.2.1. Precipitación

En la Tabla A.3 se observan los valores promedio de KGE' para todos los productos analizados. Calcular el promedio es tan importante como la mediana, ya que esta última no entrega referencias en la dispersión de los datos.

Tabla A.3: Media de los resultados de los índices de KGE' para los distintos productos de precipitación. Se destacan en negrita los mejores valores por parámetro. RFP hace referencia a RFPUNA.

		CHIRPS	PERS ⁷	ERA5	IMERG	RFMEP	RFP ⁸	RFP1	RFP2	RFP3	
Todo	Día	KGE'	-2,56	-2,86	-0,48	-1,18	-1,66	-0,64	0,18	0,23	0,29
		<i>r</i>	0,18	0,24	0,15	0,21	0,43	0,46	0,50	0,53	0,58
		β	4,04	4,34	0,55	2,39	3,09	1,96	1,18	1,18	1,18
		γ	0,60	0,45	0,74	0,87	0,62	0,75	0,49	0,52	0,56
		RMSE [mm]	4,78	4,29	3,98	4,95	3,75	3,70	3,23	3,16	3,06
	Mes	KGE'	-33,93	-33,60	-4,45	-29,94	-21,51	-20,07	-13,59	-13,34	-13,77
		<i>r</i>	0,52	0,57	0,51	0,54	0,65	0,65	0,59	0,60	0,63
		β	34,69	34,37	4,74	30,64	22,55	21,07	15,35	15,11	15,56
		γ	0,55	0,50	0,61	0,72	0,69	0,82	0,59	0,61	0,62
		RMSE [mm]	30,77	29,96	32,28	29,36	28,76	27,79	28,74	29,20	27,82
	Año	KGE'	-1,69	-2,07	-0,08	-0,30	-0,83	0,04	0,48	0,52	0,55
		<i>r</i>	0,52	0,43	0,51	0,61	0,59	0,64	0,68	0,70	0,74
		β	3,23	3,58	0,28	1,76	2,30	1,39	1,08	1,08	1,09
		γ	0,34	0,36	0,83	0,72	0,68	0,89	0,65	0,68	0,69
		RMSE [mm]	140,34	164,39	266,87	118,49	145,20	138,17	57,68	56,19	54,91
Ch	Día	KGE'	-2,83	-3,58	-0,22	-1,34	-1,91	-0,41	0,29	0,34	0,41

⁷ PERS hace referencia al producto grillado PERSIANN-CDR.

⁸ RFP hace referencia al producto grillado RFPUNA.

		r	0,14	0,20	0,17	0,16	0,45	0,48	0,54	0,57	0,62
		β	4,36	5,19	0,42	2,53	3,44	1,81	1,10	1,10	1,10
		γ	0,46	0,27	0,51	0,74	0,60	0,77	0,50	0,54	0,59
		RMSE [mm]	2,25	1,98	1,77	2,20	1,83	1,77	1,55	1,51	1,45
	Mes	KGE'	-3,04	-3,82	0,04	-1,38	-2,05	-0,45	0,16	0,17	0,24
		r	0,42	0,51	0,58	0,48	0,67	0,68	0,65	0,66	0,69
		β	4,77	5,63	0,45	2,83	3,78	2,07	1,57	1,58	1,53
		γ	0,44	0,36	0,60	0,64	0,69	0,88	0,65	0,67	0,68
	Año	RMSE [mm]	16,94	16,64	14,35	15,89	13,58	12,44	11,47	11,19	10,26
		KGE'	-2,54	-3,29	0,02	-0,70	-1,30	0,00	0,60	0,63	0,67
		r	0,51	0,46	0,65	0,62	0,71	0,76	0,77	0,78	0,83
		β	4,06	4,87	0,38	2,12	2,93	1,60	1,09	1,10	1,10
Argentina	Día	γ	0,23	0,27	0,58	0,63	0,73	0,97	0,73	0,77	0,78
		RMSE [mm]	72,73	83,29	58,20	53,42	42,11	30,68	28,01	27,07	25,46
		KGE'	-2,22	-1,94	-0,80	-0,98	-1,34	-0,94	0,05	0,10	0,15
		r	0,23	0,29	0,13	0,26	0,41	0,43	0,46	0,49	0,53
	Mes	β	3,62	3,27	0,72	2,21	2,63	2,15	1,28	1,27	1,28
		γ	0,77	0,68	1,02	1,04	0,64	0,73	0,47	0,49	0,53
		RMSE [mm]	7,98	7,21	6,78	8,44	6,19	6,15	6,94	6,96	7,02
		KGE'	-73,04	-71,33	-10,13	-66,12	-46,15	-44,92	-29,91	-29,37	-30,41
	Año	r	0,64	0,64	0,42	0,62	0,63	0,62	0,53	0,53	0,55
		β	70,22	68,51	9,83	63,68	44,83	43,63	31,71	31,18	32,23
		γ	0,68	0,69	0,63	0,82	0,69	0,75	0,53	0,54	0,55
		RMSE [mm]	43,91	42,61	49,32	42,16	43,18	42,37	45,14	46,30	44,50
	KGE'	-0,42	-0,24	-0,24	0,30	-0,12	0,11	0,32	0,36	0,39	
	r	0,52	0,40	0,30	0,59	0,39	0,45	0,55	0,58	0,62	
	β	1,91	1,54	0,12	1,17	1,31	1,04	1,06	1,06	1,06	
	γ	0,52	0,48	1,20	0,86	0,60	0,76	0,54	0,57	0,57	
	RMSE [mm]	247,39	292,81	597,26	221,51	308,43	308,37	101,05	98,76	97,96	

Luego, se grafican los KGE' de los productos para el primer semestre del año hidrológico (O.N.D.E.F.M., o en su defecto ONDEFM) y del segundo semestre (AMJJAS) en las figuras Figura

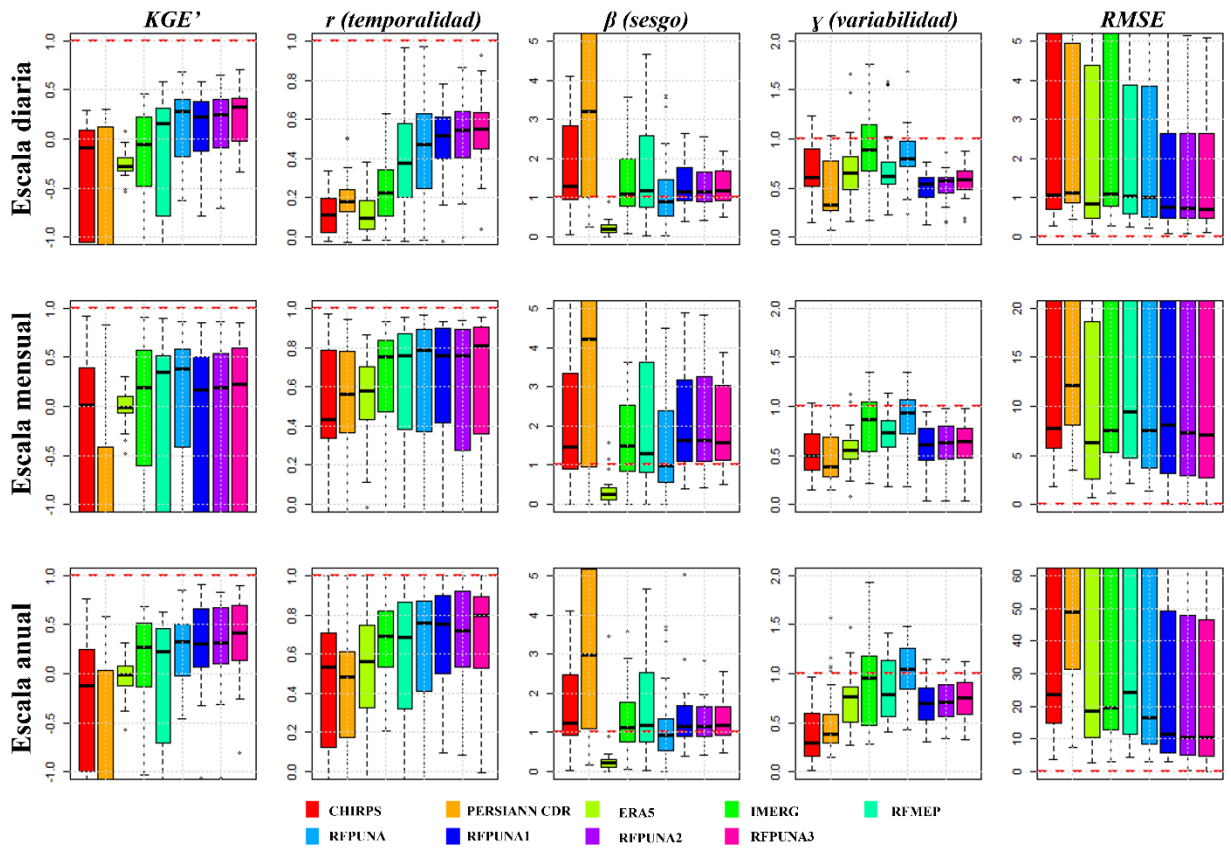


Figura A.3 respectivamente.

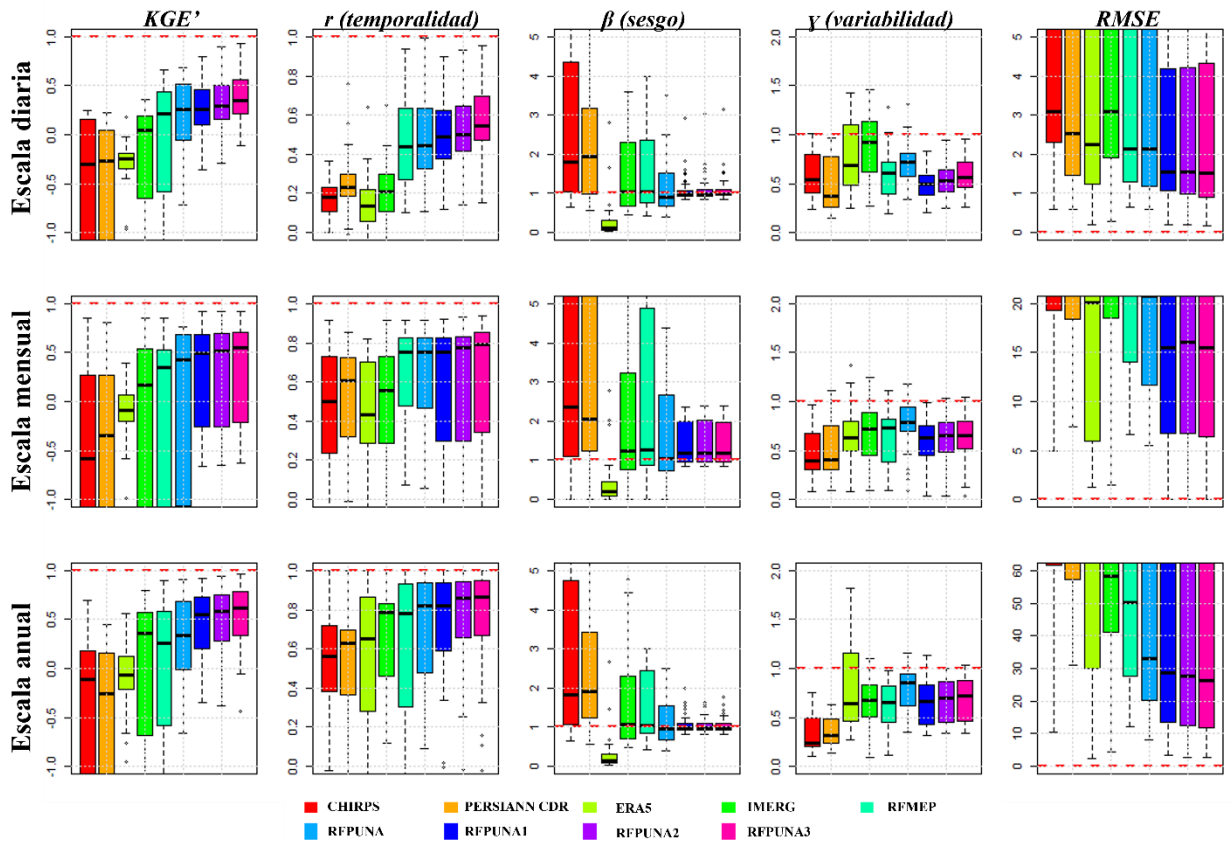


Figura A.2: Gráficos de caja de productos grillados de precipitación y su KGE' con los parámetros r , β y γ para los meses del primer semestre del año hidrológico (O.N.D.E.F.M.) en el periodo 2007-2017.

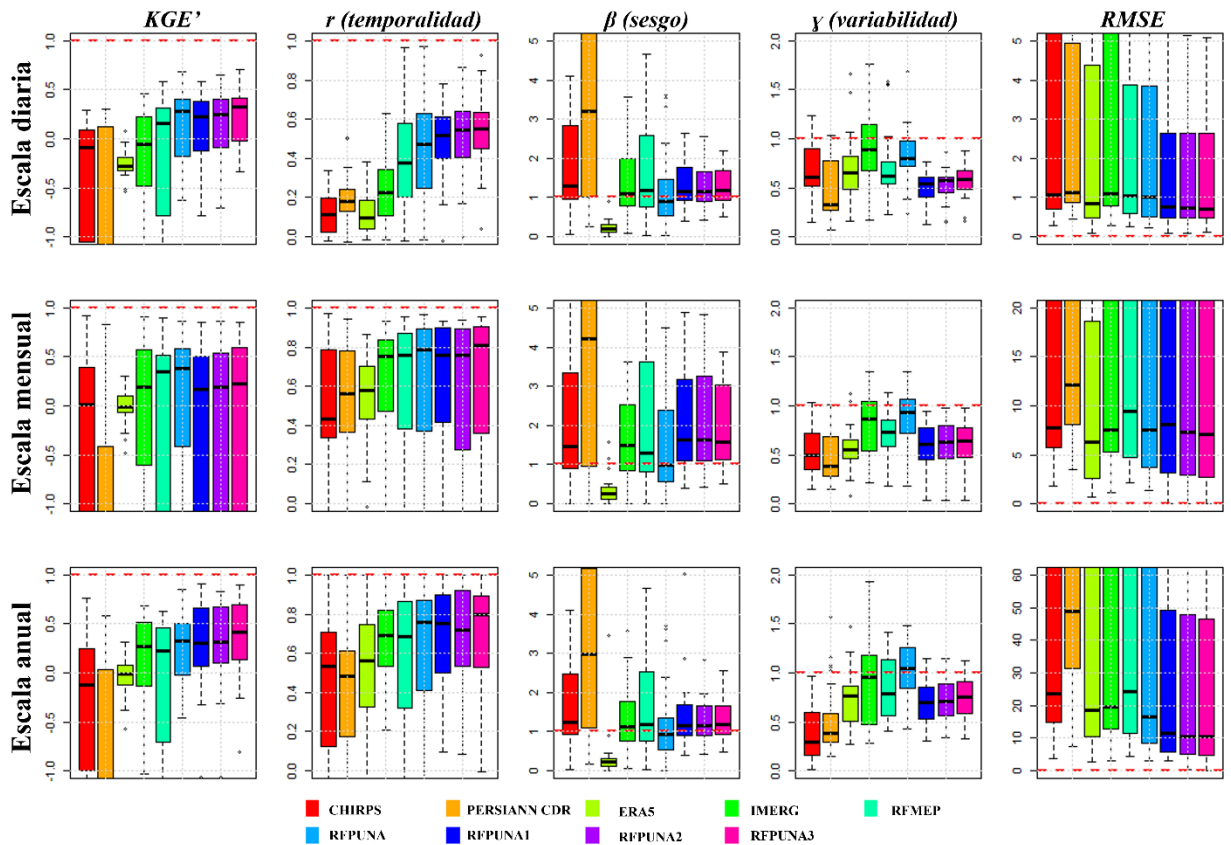


Figura A.3: Gráficos de caja de productos grillados de precipitación y su KGE' con los parámetros r , β y γ para los meses del segundo semestre del año hidrológico (A.M.J.J.A.S.) en el periodo 2007-2017.

Por último, en las FIGURAS se observan los gráficos de caja para Chile y Argentina respectivamente.

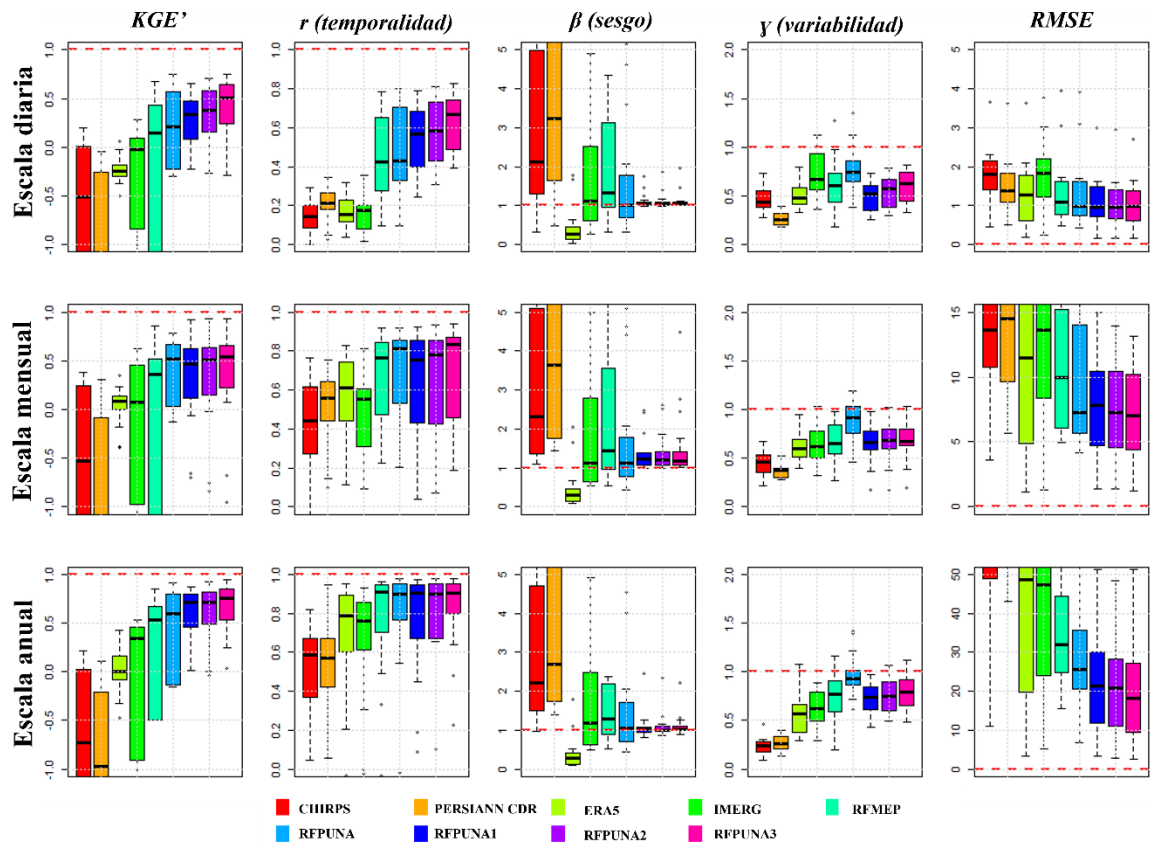


Figura A.4: Gráficos de caja de productos grillados de precipitación y su KGE' con los parámetros r , β y γ para territorio chileno, en el periodo 2007-2017.

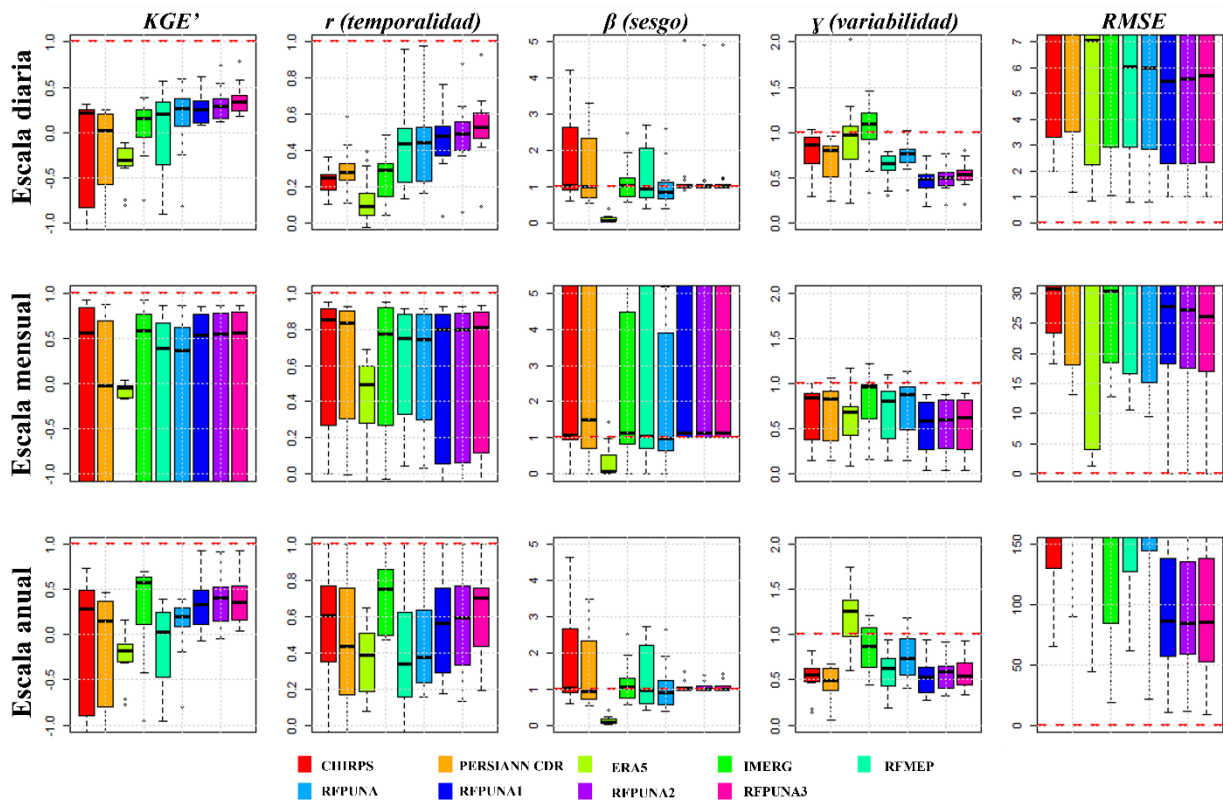


Figura A.5: Gráficos de caja de productos grillados de precipitación y su KGE' con los parámetros r , β y γ para territorio argentino, en el periodo 2007-2017.

A.2.2 Temperatura

Análogo a lo realizado en precipitación, se calcula la media del KGE' para temperatura en la Tabla A.4 y los rendimientos KGE' para los periodos estacionales en las figuras Figura A.4 y Figura A.5.

Tabla A.4: Media de los resultados de los índices de KGE' para los distintos productos de temperatura.

		CHELSA	ERA5	RFPUNA	RFPUNA1	RFPUNA2	RFPUNA3	
Todo	Día	KGE'	0,53	0,39	0,46	0,69	0,70	0,71
		r	0,80	0,75	0,76	0,80	0,80	0,81
		β	0,95	0,86	1,23	1,00	0,99	0,99
		γ	1,11	1,40	0,88	0,83	0,84	0,85
		RMSE [°C]	3,65	4,12	3,22	2,31	2,30	2,27
	Mes	KGE'	0,51	0,43	0,45	0,67	0,68	0,69
		r	0,80	0,80	0,76	0,81	0,82	0,84
		β	0,95	0,87	1,22	1,00	1,00	1,00
		γ	1,08	1,31	0,87	0,83	0,83	0,83
		RMSE [°C]	3,34	3,27	2,95	1,94	1,92	1,87

	Año	KGE'	0,37	-0,31	-0,21	0,01	0,03	0,08
		<i>r</i>	0,72	0,76	0,55	0,56	0,57	0,60
		β	0,95	0,86	1,24	0,99	0,99	0,99
		γ	1,03	2,09	1,61	1,54	1,52	1,48
		RMSE [°C]	2,66	2,24	2,07	0,86	0,85	0,84
Chile	Día	KGE'	-0,19	-0,40	-0,55	0,35	0,37	0,40
		<i>r</i>	0,52	0,51	0,40	0,58	0,58	0,59
		β	1,25	0,71	1,83	1,06	1,06	1,06
		γ	1,09	1,92	0,82	0,55	0,57	0,60
		RMSE [°C]	6,45	5,82	5,45	2,98	2,98	2,94
	Mes	KGE'	-0,35	-0,36	-0,62	0,26	0,27	0,30
		<i>r</i>	0,42	0,42	0,28	0,50	0,52	0,56
		β	1,26	0,74	1,82	1,09	1,09	1,08
		γ	1,00	1,65	0,73	0,49	0,49	0,50
		RMSE [°C]	6,59	6,03	5,67	3,16	3,12	3,05
	Año	KGE'	-0,09	-1,91	-2,13	-0,76	-0,69	-0,59
		<i>r</i>	0,52	0,50	0,44	0,50	0,51	0,51
		β	1,32	0,66	1,90	1,05	1,05	1,05
		γ	1,06	3,56	2,99	2,24	2,19	2,09
		RMSE [°C]	5,24	4,46	3,64	0,89	0,89	0,87
Argentina	Día	KGE'	0,73	0,63	0,75	0,83	0,84	0,85
		<i>r</i>	0,88	0,82	0,87	0,89	0,89	0,89
		β	0,86	0,92	1,06	1,00	1,00	1,00
		γ	1,11	1,24	0,88	0,88	0,89	0,90
		RMSE [°C]	7,67	8,13	6,08	7,43	7,44	7,45
	Mes	KGE'	0,77	0,67	0,77	0,84	0,85	0,86
		<i>r</i>	0,91	0,91	0,90	0,92	0,93	0,94
		β	0,86	0,92	1,05	1,00	1,00	1,00
		γ	1,09	1,20	0,89	0,88	0,88	0,88
		RMSE [°C]	2,34	2,35	2,17	1,11	1,09	1,06
	Año	KGE'	0,47	0,04	0,31	0,49	0,50	0,54
		<i>r</i>	0,77	0,81	0,57	0,65	0,65	0,69
		β	0,85	0,92	1,07	1,00	1,00	1,00
		γ	1,04	1,77	1,20	1,05	1,05	1,03
		RMSE [°C]	1,88	1,47	1,66	0,32	0,31	0,30

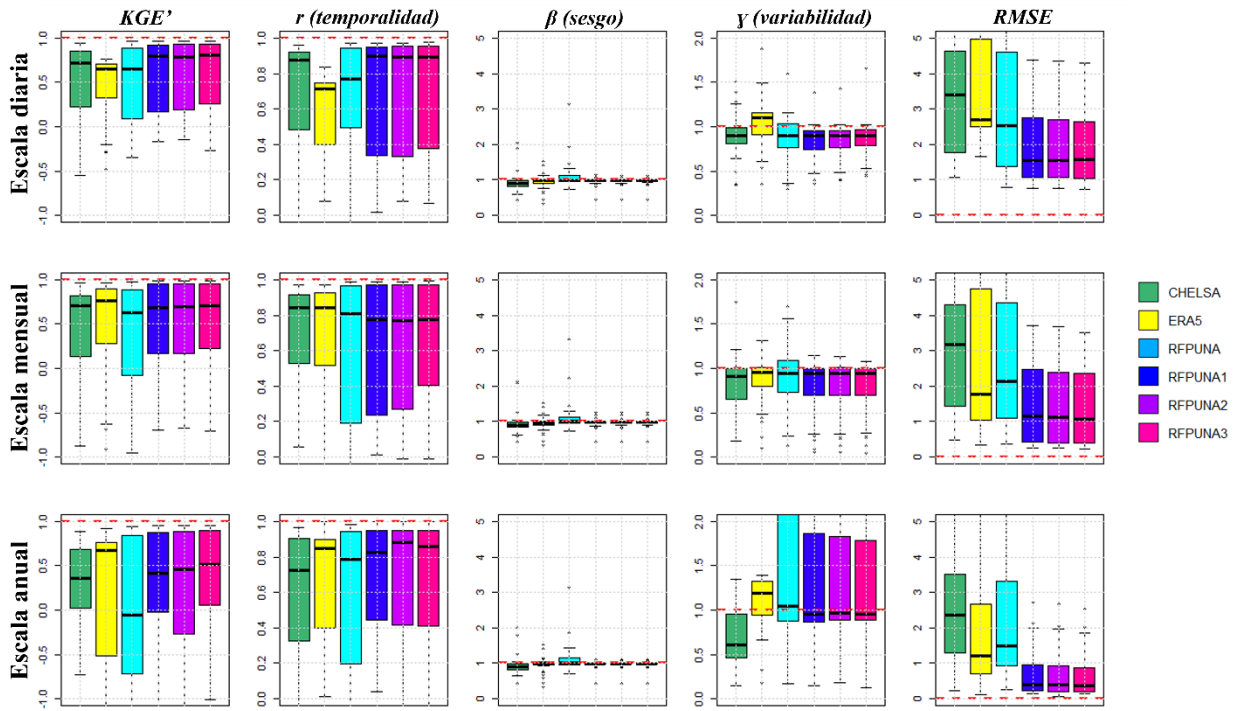


Figura A.4: Gráficos de caja de productos grillados de temperatura y su KGE' con los parámetros r , β y γ para los meses del primer semestre del año hidrológico (O.N.D.E.F.M.) en el periodo 2007-2016.

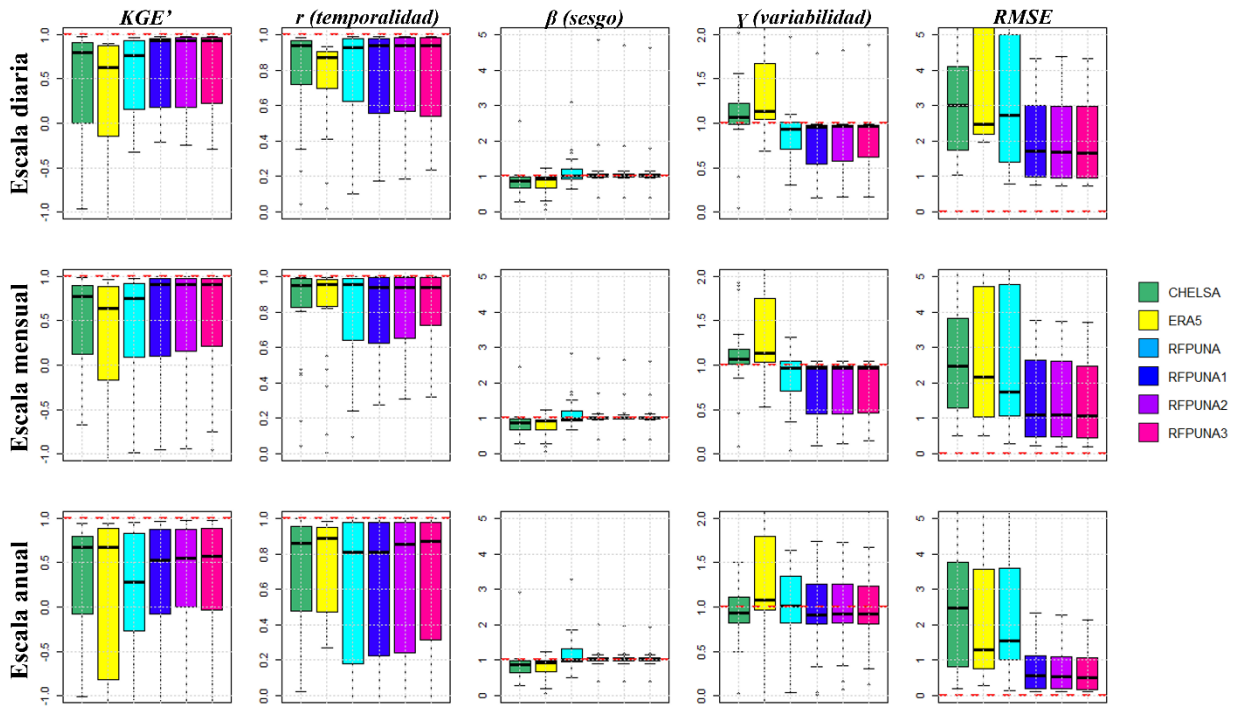


Figura A.5: Gráficos de caja de productos grillados de temperatura y su KGE' con los parámetros r , β y γ para los meses del segundo semestre del año hidrológico (A.M.J.J.A.S.) en el periodo 2007-2016.

ANEXO B: REGISTRO DE ESTACIONES

Se recopilan las principales características de las estaciones, con su nombre completo, su ID (definido como referencia), longitud, latitud, elevación y cuál es la institución administrativa encargada. Por confidencialidad, no se presentan las estaciones privadas.

- **Chile**

En territorio chileno se consideraron 47 estaciones públicas, todas provenientes de la DGA. Si bien existen estaciones que no se encuentran dentro del área de estudio, se consideraron por su buen registro de información, su cercanía a la zona y estar a cotas similares a las demás. En la Tabla B.1 se enseña el contexto geográfico de cada estación, considerando longitud, latitud y elevación para el sistema de referencia WGS 84.

Tabla B.1: Estaciones públicas en territorio chileno.

Nombre	ID	Longitud	Latitud	Elevación [msnm]	Institución administrativa
Amincha	Ch_1_pub	-68,339422	-21,200828	3860	DGA
Ollague	Ch_2_pub	-68,255593	-21,223952	3700	DGA
Cebollar	Ch_3_pub	-68,338585	-21,53139	3730	DGA
Lequena	Ch_4_pub	-68,662403	-21,658677	3320	DGA
Rio Loa Antes Represa Lequena	Ch_5_pub	-68,66209	-21,655884	3315	DGA
Quillagua	Ch_6_pub	-69,535684	-21,655201	802	DGA
Ascotan	Ch_7_pub	-68,275618	-21,725198	3970	DGA
Quinchamale	Ch_8_pub	-68,59844	-21,909029	3080	DGA
San Pedro De Conchi	Ch_9_pub	-68,536369	-21,93362	3217	DGA
Parshall N 2	Ch_10_pub	-68,519703	-21,945479	3318	DGA
Conchi Viejo	Ch_11_pub	-68,725852	-21,95384	3491	DGA
Ojos San Pedro	Ch_12_pub	-68,339134	-21,987873	3800	DGA
Rio Siloli Antes B,T, Fcab	Ch_13_pub	-68,028915	-22,013748	4000	DGA
Silala	Ch_14_pub	-68,032312	-22,017164	4305	DGA
Conchi Embalse	Ch_15_pub	-68,624047	-22,02698	3010	DGA
Conchi Muro Embalse	Ch_16_pub	-68,620821	-22,026955	3000	DGA
Inacaliri	Ch_17_pub	-68,066125	-22,027288	4040	DGA
Rio Loa En Salida Embalse Conchi	Ch_18_pub	-68,621157	-22,027895	2950	DGA
Cupo	Ch_19_pub	-68,316643	-22,1148	3370	DGA
Linzor	Ch_20_pub	-68,022845	-22,229801	4100	DGA
Turi	Ch_21_pub	-68,288026	-22,241125	3070	DGA
Toconce	Ch_22_pub	-68,166327	-22,264684	3310	DGA
Ayquina	Ch_23_pub	-68,324026	-22,275905	3031	DGA
Rio Salado A, J, Curti	Ch_24_pub	-68,242786	-22,279088	3080	DGA
Sifon Ayquina	Ch_25_pub	-68,349702	-22,283625	3000	DGA
Salado Embalse	Ch_26_pub	-68,203402	-22,285169	3200	DGA
Rio Salado En Sifon Ayquina	Ch_27_pub	-68,344832	-22,289288	2980	DGA
Caspana	Ch_28_pub	-68,212197	-22,336956	3260	DGA

Chiu-Chiu	Ch_29_pub	-68,648061	-22,345445	2524	DGA
El Tatio	Ch_30_pub	-68,016253	-22,351007	4370	DGA
Coya Sur	Ch_31_pub	-68,000	-23,000	3000	DGA
Calama	Ch_32_pub	-68,906317	-22,5209	2300	DGA
Rio Grande	Ch_33_pub	-68,167375	-22,651977	3250	DGA
Sierra Gorda	Ch_34_pub	-69,321292	-22,890806	1616	DGA
San Pedro De Atacama	Ch_35_pub	-68,200528	-22,910384	2450	DGA
Toconao Experimental	Ch_36_pub	-67,997741	-23,189194	2500	DGA
Toconao Reten	Ch_37_pub	-68,009401	-23,193919	2460	DGA
Talabre	Ch_38_pub	-67,889638	-23,315846	3300	DGA
Camar	Ch_39_pub	-67,962	-23,408595	2700	DGA
Socaire	Ch_40_pub	-67,891654	-23,58787	3251	DGA
Peine	Ch_41_pub	-68,066942	-23,681879	2460	DGA
Imilac	Ch_42_pub	-68,782777	-24,163377	3000	DGA
Monturaqui	Ch_43_pub	-68,435262	-24,341721	3430	DGA
Aguas Verdes	Ch_44_pub	-69,962551	-25,401037	1560	DGA
Pastos Grandes	Ch_45_pub	-69,5642	-27,1142	2260	DGA
Sierra Overa	Ch_46_pub	-69,8881	-25,78	1900	DGA
Las Vegas	Ch_47_pub	-69,6656	-26,6781	2250	DGA

Para mayor comodidad al referenciar, se llamará a las estaciones por su ID y no por su nombre. Luego, se tabulan las características principales de cada estación en la Tabla B.2, considerando los periodos de ambas variables climáticas, su monto medio anual y una columna de observaciones, donde se describe a grandes rasgos si lo descrito en los períodos es veraz o se trata de información incompleta con vacíos entre medio. Es necesario acotar que no todas las estaciones miden temperatura, el signo ‘-’ se utiliza para completar cuando no exista esta variable en la estación.

Tabla B.2: Características de las estaciones públicas en territorio chileno.

ID	Periodo precipitación	Periodo temperatura	Precipitación media anual [mm]	Temperatura media anual [°C]	Años aprobados seguidos	Observaciones
Ch_1_pub	1979-1981	-	36	-	0	Incompleta
Ch_2_pub	1975-2019	1975-2019	47	7,3	7	Incompleta
Ch_3_pub	1976-2007	1976-1987	53,9	7,5	12	Incompleta
Ch_4_pub	1975-2020	1983-1989	85,9	9,1	19	Completa
Ch_5_pub	1999-2020	-	72,2	-	17	Incompleta (2019)
Ch_6_pub	1978-2019	1975-1979	0,5	19,4	25	Precipitación completa, temperatura incompleta
Ch_7_pub	1975-2017	1975-1983	59,5	5,6	16	Incompleta
Ch_8_pub	1996-2019	-	25,0	-	11	Completa
Ch_9_pub	1975-1997	-	33,4	-	14	Completa
Ch_10_pub	1975-2013	1975-1993	28,4	11,1	23	Precipitación completa, temperatura incompleta
Ch_11_pub	1975-2018	-	45,4	-	18	Incompleta
Ch_12_pub	1978-2017	1981-1989	49,9	4,7	18	Completa

Ch_13_pub	2012-2019	-	221,7	-	1	Incompleta
Ch_14_pub	2000-2019	-	103,6	-	13	Completa
Ch_15_pub	1975-2019	1975-1994	19,8	11,6	15	Completa
Ch_16_pub	1978-1991	1980-1991	18,3	11,4	10	Completa
Ch_17_pub	1975-2019	1975-1992	116,1	6,1	15	Precipitación completa, temperatura incompleta
Ch_18_pub	2012-2019	-	95,3	-	2	Completa
Ch_19_pub	1978-2019	-	67,2	-	19	Completa
Ch_20_pub	1975-2020	1975-2019	155,1	4,8	15	Incompleta
Ch_21_pub	1980-1993	1980-1993	37,1	10,3	8	Completa
Ch_22_pub	1975-2019	1993-1994	93,1	12,6	14	Completa
Ch_23_pub	1975-2019	1993-1994	40,1	13,2	27	Completa
Ch_24_pub	2012-2013	-	54,0	-	0	Completa
Ch_25_pub	1975-1977	-	41,4	-	2	Completa
Ch_26_pub	1975-2020	2005-2020	63,9	7,7	30	Incompleta
Ch_27_pub	2011-2019	-	24,7	-	13	Incompleta
Ch_28_pub	1975-2008	1975-1994	70,7	10,5	14	Completa
Ch_29_pub	1975-2019	1979-1994	5,6	13,4	25	Completa
Ch_30_pub	1977-2020	1983-2019	124,4	2,1	13	Incompleta
Ch_31_pub	1975-1991	1975-1991	0,7	21,0	14	Incompleta
Ch_32_pub	1975-2020	1975-1994	3,9	14,6	14	Incompleta
Ch_33_pub	1976-2019	-	74,6	-	26	Completa
Ch_34_pub	1994-2020	2014-2020	1,2	16,4	14	Completa
Ch_35_pub	1975-2017	1975-1989	31,4	15,5	11	Precipitación incompleta, temperatura completa
Ch_36_pub	1975-2009	1981-1994	30,0	17,1	14	Precipitación completa, temperatura incompleta
Ch_37_pub	1975-1991	-	39,3	-	16	Completa
Ch_38_pub	1995-2019	-	70,6	-	11	Completa
Ch_39_pub	1978-2019	-	41,5	-	39	Completa
Ch_40_pub	1975-2017	1975-1994	39,3	10,7	13	Precipitación completa, temperatura incompleta
Ch_41_pub	1975-2019	1975-1994	20	16,8	25	Precipitación completa, temperatura incompleta
Ch_42_pub	1975-1977	-	4,3	-	2	Completa
Ch_43_pub	1975-1979	1975-1980	58,5	8,1	5	Precipitación completa, temperatura incompleta
Ch_44_pub	1987-2019	1988-1994	8,09	16,7	13	Precipitación completa, temperatura incompleta
Ch_45_pub	1975-2019	-	33,1	-	22	Completa
Ch_46_pub	1986-1988	-	19,7	-	1	Completa
Ch_47_pub	1983-2020	-	33,9	-	15	Completa

- **Argentina**

En territorio argentino se consideran 57 estaciones públicas provenientes de distintas fuentes. Similar a las estaciones públicas chilenas, se consideran estaciones fuera del área de estudio si se encuentran en un área de influencia cercano y cuentan con buena calidad de información. Estas se encuentran tabuladas en las Tabla B.3 y Tabla B.4.

Tabla B.3: Estaciones públicas en territorio argentino.

Nombre	ID	Longitud	Latitud	Elevación [msnm]	Institución administrativa
Salta Aero	Ar_1_pub	-65,483	-24,85	1221	SMN
Oran Aero	Ar_2_pub	-64,329	-23,153	356	SMN
Metan	Ar_3_pub	-64,8	-25,483	855	SMN
Jujuy Aero	Ar_4_pub	-65,083	-24,383	905	SMN
Jujuy U N	Ar_5_pub	-65,183	-24,167	1302	SMN
La Quiaca	Ar_6_pub	-65,6	-22,1	3459	SMN
Santiago del estero Aero	Ar_7_pub	-64,3	-27,767	199	SMN
Termas de Río Hondo	Ar_8_pub	-64,933	-27,483	280	SMN
Tucumán Aero	Ar_9_pub	-65,1	-26,85	450	SMN
General Enrique Mosconi Tarta	Ar_10_pub	-63,794	-22,62	449	SMN
Bermejo	Ar_11_pub	-64,317	-22,767	381	SMN
Teniente Benjamín Matienzo	Ar_12_pub	-65,105	-26,841	456	SMN
Caimancito	Ar_16_pub	-64,539583	-23,712056	355,64	RHN
Madrejones	Ar_50_pub	-64,115	-22,3738	467	RHC
Río Blanco	Ar_51_pub	-64,2808	-23,0702	335	RHC
Escuela San Andrés	Ar_52_pub	-64,8675	-23,0848	1733	RHC
Humahuaca	Ar_53_pub	-65,3465	-23,2081	2944	RHC
Tilcara	Ar_54_pub	-65,3932	-23,5767	2475	RHC
Río San Lorenzo	Ar_55_pub	-64,7859	-23,7927	487	RHC
Chañarcito	Ar_56_pub	-65,4741	-23,7587	2185	RHC
Río Reyes	Ar_57_pub	-65,4901	-24,1685	1753	RHC
Palma Sola	Ar_58_pub	-64,3361	-24,024	790	RHC
Río Xibi-Xibi	Ar_59_pub	-65,3522	-24,2125	1369	RHC
Bajada de Pinto	Ar_60_pub	-64,8454	-24,4358	524	RHC
Los Toldos	Ar_61_pub	-64,6985	-22,2802	1581	RHC
Pie de Médano	Ar_210_pub	-66,2177	-26,99547	2180	RHN
Puerta de San José	Ar_214_pub	-67,00479	-27,5541	1404	RHN
Potrero del Clavillo	Ar_409_pub	-65,974	-27,4	1383	RHN
KM-19	Ar_410_pub	-65,636	-27,1009	610	RHN
Potrero de las tablas	Ar_411_pub	-65,426	-26,85777	950	RHN
Casa de Piedra	Ar_452_pub	-65,9123	-27,282	1200	RHN
Piedra Grande	Ar_453_pub	-65,806	-27,313	500	RHN
Yampa II	Ar_455_pub	-65,841	-27,178	1500	RHN
El Nogalito	Ar_456_pub	-65,471667	-26,785472	1100	RHN
Las Mesadas	Ar_457_pub	-65,931	-27,214	1300	RHN
Aguas Blancas	Ar_604_pub	-64,364606	-22,7249	407,44	RHN

Balapuca	Ar_605_pub	-64,463	-22,492	581,38	RHN
Alarache	Ar_607_pub	-64,586	-22,249	937,05	RHN
Cuatro Cedros	Ar_629_pub	-64,515825	-22,829403	484,01	RHN
San Telmo	Ar_637_pub	-64,971778	-22,573361	429,91	RHN
Astillero	Ar_638_pub	-64,119	-22,374	500	RHN
San José	Ar_655_pub	-64,703639	-22,8655	877	RHN
Iruya	Ar_672_pub	-65,215861	-22,795611	2741	RHN
Las Higueras	Ar_673_pub	-65,097333	-22,754278	1972	RHN
San Antonio	Ar_674_pub	-64,971167	-22,825556	1400	RHN
San Isidro	Ar_675_pub	-65,24175	-22,756528	2956	RHN
Nazareno	Ar_676_pub	-65,101472	-22,513778	3101	RHN
Paltorco	Ar_677_pub	-65,088	-22,405	3720	RHN
Santa Victoria Oeste	Ar_678_pub	-64,971	-22,252	450	RHN
Tuc Tuca	Ar_679_pub	-65,267472	-22,404444	3950	RHN
Güemes	Ar_687_pub	-65,043639	-24,687528	734	RHN
Poscaya	Ar_688_pub	-65,0805	-22,454	3238	RHN
Trigo Huaico	Ar_689_pub	-65,037	-22,374	3300	RHN
Colanzulí	Ar_690_pub	-65,224222	-22,887944	3575	RHN
El Molino	Ar_691_pub	-65,151472	-22,597722	2600	RHN
El Pabellón	Ar_692_pub	-65,339083	-22,553472	4265	RHN
Pozo Sarmiento	Ar_693_pub	-64,196833	-23,140556	326,06	RHN
San Luis	Ar_697_pub	-66	-26,083333	1800	RHN

Tabla B.4: Características de las estaciones públicas en territorio argentino.

ID	Periodo precipitación	Periodo temperatura	Precipitación media anual [mm]	Temperatura media anual [°C]	Años aprobados seguidos	Observaciones
Ar_1_pub	1975-2020	1975-2020	759,8	17,5	46	Completa
Ar_2_pub	1975-2020	1975-2020	997	22,3	46	Completa
Ar_3_pub	1993-2020	1993-2020	784,6	19,3	3	Incompleta
Ar_4_pub	1975-2020	1975-2020	731,9	19,5	46	Completa
Ar_5_pub	1987-2020	1987-2020	930,6	17,6	33	Completa
Ar_6_pub	1975-2020	1975-2020	341,4	10,4	46	Completa
Ar_7_pub	1975-2020	1975-2020	643,9	20,9	46	Completa
Ar_8_pub	1975-2020	2014-2020	559,6	20,8	4	Incompleta
Ar_9_pub	1981-2020	1981-2020	989,3	20,3	39	Completa
Ar_10_pub	1975-2019	1975-2019	649,1	22,5	0	Incompleta
Ar_11_pub	1988-2008	1986-2019	171,1	22,7	0	Incompleta
Ar_12_pub	1981-2019	1981-2019	883,3	19,8	0	Completa
Ar_16_pub	1993-2022	1992-2022	347,1	22,3	26	Precipitaciones incompletas, temperatura completa
Ar_50_pub	1975-2000	2021-2022	1309,1	24,7	14	Completa
Ar_51_pub	1975-2022	-	172,4	-	15	Incompleta

Ar_52_pub	2017-2022	-	147,1	-	3	Completa
Ar_53_pub	1984-2022	2019-2022	399,7	11,6	30	Completa
Ar_54_pub	1982-2022	-	434,0	-	38	Completa
Ar_55_pub	1975-2022	-	1393,1	-	11	Incompleta
Ar_56_pub	1983-2022	-	538,1	-	33	Completa
Ar_57_pub	1982-2022	-	338,0	-	35	Completa
Ar_58_pub	1983-2022	2019-2022	472,5	22,1	28	Completa
Ar_59_pub	1982-2022	-	289,0	-	38	Completa
Ar_60_pub	1982-2022	-	283,7	-	38	Completa
Ar_210_pub	1991-2013	-	229,6	-	21	Completa
Ar_214_pub	1982-2022	-	574,6	-	28	Completa
Ar_409_pub	1980-2022	1978-2022	1932,6	13,5	40	Completa
Ar_410_pub	1975-2022	1993-2022	1496,3	18,9	28	Completa
Ar_411_pub	1975-2022	1992-2022	1349,5	17,8	35	Completa
Ar_452_pub	1988-2022	-	942,2	-	32	Completa
Ar_453_pub	1975-2022	-	1051,4	-	28	Completa
Ar_455_pub	1987-2022	-	1024,5	-	33	Completa
Ar_456_pub	1975-2022	-	1370,4	-	30	Incompleta
Ar_457_pub	1975-2022	-	1096,4	-	46	Completa
Ar_604_pub	1988-2022	1975-2022	1932,7	22,4	32	Completa
Ar_605_pub	1988-2022	1975-2022	2375,8	21,3	32	Completa
Ar_607_pub	1988-2020	1975-2000	1614,8	19,0	31	Completa
Ar_629_pub	1988-2022	1975-2022	1898,8	21,4	29	Completa
Ar_637_pub	1975-2022	1975-2022	1330,0	22,2	7	Precipitaciones completas, temperaturas incompletas
Ar_638_pub	1975-2022	1975-1987	1645,7	20,8	46	Completa
Ar_655_pub	1975-2022	1992-2022	1944,8	19,8	25	Precipitaciones incompletas, temperaturas completas
Ar_672_pub	1975-2022	-	2095,8	-	44	Incompleta
Ar_673_pub	1984-2022	-	560,3	-	36	Completa
Ar_674_pub	1984-2022	-	513,5	-	31	Incompleta
Ar_675_pub	1984-2022	-	738,0	-	33	Completa
Ar_676_pub	2021-2022	-	475,4	-	0	Completa
Ar_677_pub	1984-2022	-	513,3	-	0	Incompleta
Ar_678_pub	2019-2022	-	534,8	-	1	Incompleta
Ar_679_pub	2018-2022	2019-2022	416,1	6,2	0	Precipitaciones incompletas, temperaturas completas
Ar_687_pub	2019-2022	-	221,6	-	0	Incompleta
Ar_688_pub	2019-2022	-	110,7	-	1	Incompleta
Ar_689_pub	2019-2022	-	939,0	-	1	Incompleta
Ar_690_pub	2019-2022	1992-2013	71,1	10,5	1	Incompleta
Ar_691_pub	2019-2022	-	793,0	-	1	Incompleta

Ar_692_pub	2019-2022	-	343,5	-	0	Incompleta
Ar_693_pub	2019-2022	1988-2022	289,4	23,4	0	Precipitaciones incompletas, temperaturas completas
Ar_697_pub	2019-2022	-	301,1	-	1	Incompleta

- **Bolivia**

Finalmente, se tabulan las estaciones bolivianas en las tablas Tabla B.5 y Tabla B.6.

Tabla B.5: Estaciones públicas en territorio boliviano.

Nombre	ID	Longitud	Latitud	Elevación [msnm]	Institución administrativa
Tarija	Bo_1_pub	-64,7	-21,55	1854	GHCN
Yacuiba	Bo_2_pub	-63,65	-21,95	656	GHCN
Tcnl, Rafael Pabón	Bo_3_pub	-63,406	-21,255	398	GHCN
Narváez	Bo_4_pub	-64,278939	-21,409802	1712	RHC
Salinas	Bo_5_pub	-64,2525	-21,8268	1031	RHC
Cañas	Bo_6_pub	-64,9075	-21,8956	2178	RHC
El Cajón	Bo_7_pub	-64,1286	-22,328	498	RHC

Tabla B.6: Características de las estaciones públicas en territorio boliviano.

ID	Periodo precipitación	Periodo temperatura	Precipitación media anual [mm]	Temperatura media anual [°C]	Años aprobados seguidos	Observaciones
Bo_1_pub	1975-2019	1975-2019	330,5	19,2	0	Precipitaciones incompletas, temperaturas completas
Bo_2_pub	1975-2019	1975-2019	596,8	22,7	0	Precipitaciones incompletas, temperaturas completas
Bo_3_pub	1980-2019	1980-2019	383,6	24,9	0	Incompleta
Bo_4_pub	2019-2022	-	371,4	-	0	Incompleta
Bo_5_pub	2019-2022	-	486,6	-	0	Incompleta
Bo_6_pub	2019-2022	-	179,3	-	0	Incompleta

Control de Calidad

- Precipitación

En primer lugar, mediante métodos de detección, se deben eliminar los valores “imposibles” de precipitación y temperatura. Luego, a través de un ponderador α , obtener un umbral para clasificar como año “aprobado” o “reprobado” según la superación de este, tal como se muestra en la ecuación (4.1).

Con respecto a este análisis cuantitativo, se destacan 4 estaciones que cumplen con 45 años seguidos de información completa se encuentran en Argentina, estas son: Oran Aero (Ar_2_pub), Santiago del Estero Aero (Ar_7_pub), Las Mesadas (Ar_457_pub) y Astillero (Ar_638_pub).

Bajo el criterio de detección, se pueden distinguir al menos dos estaciones con esta clase de errores al observar precipitaciones medias mensuales valores fuera de todo evento histórico a nivel regional. Estas estaciones son privadas y fueron descartadas de inmediato.

Con respecto al control cualitativo, se realizan los test de las tablas Tabla B.7 y Tabla B.8 para analizar valores fuera de rango, el cual es definido por el usuario (en este caso, se definió ese umbral sobre los 40 [mm] de lluvia); las repeticiones diarias, que se basan en la cantidad de días que se repite un mismo valor (se definió arbitrariamente que exista una alarma si se repite tres días seguidos del mismo valor); y los *outliers* climáticos, definidos estadísticamente según los percentiles de la observación.

En la Tabla B.7, se muestra una estación con estas tres alertas, pero sin eventos aparentemente extraños. Todo lo contrario, sucede con la estación de la Tabla B.8, donde se repite un mismo valor desde principios de noviembre hasta mediados de diciembre del 2001. Esto da cuenta de una mala práctica de parte de los operadores, quienes, por múltiples razones, alteran en toda su lógica los registros de una estación.

Tabla B.7: Test cualitativo de malas prácticas y outliers para una buena estación (Ar_10_pub).

Variables	Año	Mes	Día	Valor	Test
Precipitación	1977	6	7	7,9	<i>climatic outliers</i>
Precipitación	1978	4	14	35,1	<i>climatic outliers</i>
Precipitación	1978	12	19	45	<i>daily out of range</i>
Precipitación	1982	1	30	50	<i>daily out of range</i>
Precipitación	1982	11	30	40,9	<i>daily out of range</i>
Precipitación	1984	2	28	40,9	<i>daily out of range</i>
Precipitación	1985	4	23	30	<i>climatic outliers</i>
Precipitación	1988	1	18	99,1	<i>climatic outliers;daily out of range</i>
Precipitación	1988	12	5	46	<i>daily out of range</i>
Precipitación	1990	2	20	43,9	<i>daily out of range</i>
Precipitación	1990	11	16	43,9	<i>daily out of range</i>
Precipitación	1991	1	10	41,9	<i>daily out of range</i>
Precipitación	1991	1	24	47	<i>daily out of range</i>
Precipitación	1993	3	28	3	<i>daily repetition</i>
Precipitación	1993	3	29	3	<i>daily repetition</i>
Precipitación	1993	3	30	3	<i>daily repetition</i>
Precipitación	1993	12	9	51,1	<i>daily out of range</i>
Precipitación	1993	12	31	193	<i>climatic outliers;daily out of range</i>
Precipitación	1995	9	24	39,9	<i>climatic outliers</i>
Precipitación	1995	12	5	45	<i>daily out of range</i>
Precipitación	1996	1	7	43,9	<i>daily out of range</i>
Precipitación	1996	3	16	52,1	<i>daily out of range</i>

Precipitación	2000	1	25	78	<i>climatic outliers;daily out of range</i>
Precipitación	2001	10	15	48	<i>climatic outliers;daily out of range</i>
Precipitación	2001	12	11	108	<i>climatic outliers;daily out of range</i>
Precipitación	2002	10	27	5,1	<i>daily repetition</i>
Precipitación	2002	10	28	5,1	<i>daily repetition</i>
Precipitación	2002	10	29	5,1	<i>daily repetition</i>
Precipitación	2002	11	19	59,9	<i>climatic outliers;daily out of range</i>
Precipitación	2003	12	12	49	<i>daily out of range</i>
Precipitación	2003	12	27	102,9	<i>climatic outliers;daily out of range</i>
Precipitación	2004	10	30	52,1	<i>climatic outliers;daily out of range</i>
Precipitación	2005	2	13	41,9	<i>daily out of range</i>
Precipitación	2005	3	28	54,1	<i>daily out of range</i>
Precipitación	2007	1	15	48	<i>daily out of range</i>
Precipitación	2018	11	5	56,1	<i>daily out of range</i>
Precipitación	2018	12	16	48	<i>daily out of range</i>
Precipitación	2019	1	14	40,1	<i>daily out of range</i>
Precipitación	2019	2	1	47,2	<i>daily out of range</i>
Precipitación	2019	3	31	55,1	<i>daily out of range</i>

Tabla B.8: Test cualitativo de malas prácticas y outliers para una mala estación (privada).

	Año	Mes	Día	Valor	Test
Precipitación	2000	1	22	15,6	<i>climatic_outliers</i>
Precipitación	2000	12	14	8,7	<i>climatic_outliers</i>
Precipitación	2001	11	2	0,1	<i>daily_repetition</i>
Precipitación	2001	11	3	0,1	<i>daily_repetition</i>
Precipitación	2001	11	4	0,1	<i>daily_repetition</i>
Precipitación	2001	11	6	0,1	<i>daily_repetition</i>
Precipitación	2001	11	7	0,1	<i>daily_repetition</i>
Precipitación	2001	11	8	0,1	<i>daily_repetition</i>
Precipitación	2001	11	9	0,1	<i>daily_repetition</i>
Precipitación	2001	11	10	0,1	<i>daily_repetition</i>
Precipitación	2001	11	11	0,1	<i>daily_repetition</i>
Precipitación	2001	11	12	0,1	<i>daily_repetition</i>
Precipitación	2001	11	13	0,1	<i>daily_repetition</i>
Precipitación	2001	11	14	0,1	<i>daily_repetition</i>
Precipitación	2001	11	15	0,1	<i>daily_repetition</i>
Precipitación	2001	11	16	0,1	<i>daily_repetition</i>
Precipitación	2001	11	17	0,1	<i>daily_repetition</i>
Precipitación	2001	11	18	0,1	<i>daily_repetition</i>
Precipitación	2001	11	19	0,3	<i>climatic_outliers</i>
Precipitación	2001	11	20	0,1	<i>daily_repetition</i>
Precipitación	2001	11	21	0,1	<i>daily_repetition</i>
Precipitación	2001	11	22	0,1	<i>daily_repetition</i>
Precipitación	2001	11	23	0,1	<i>daily_repetition</i>
Precipitación	2001	11	24	0,1	<i>daily_repetition</i>

Precipitación	2001	11	25	0,1	<i>daily_repetition</i>
Precipitación	2001	11	27	0,1	<i>daily_repetition</i>
Precipitación	2001	11	28	0,1	<i>daily_repetition</i>
Precipitación	2001	11	29	0,1	<i>daily_repetition</i>
Precipitación	2001	11	30	0,1	<i>daily_repetition</i>
Precipitación	2001	12	1	0,1	<i>daily_repetition</i>
Precipitación	2001	12	2	0,1	<i>daily_repetition</i>
Precipitación	2001	12	3	0,1	<i>daily_repetition</i>
Precipitación	2001	12	5	0,1	<i>daily_repetition</i>
Precipitación	2001	12	6	0,1	<i>daily_repetition</i>
Precipitación	2001	12	7	0,1	<i>daily_repetition</i>
Precipitación	2001	12	8	0,1	<i>daily_repetition</i>
Precipitación	2001	12	10	0,1	<i>daily_repetition</i>
Precipitación	2001	12	11	0,1	<i>daily_repetition</i>
Precipitación	2001	12	12	0,1	<i>daily_repetition</i>
Precipitación	2001	12	13	0,1	<i>daily_repetition</i>
Precipitación	2002	4	9	23,4	<i>climatic_outliers</i>
Precipitación	2003	11	11	0,8	<i>climatic_outliers</i>
Precipitación	2013	5	17	10,9	<i>climatic_outliers</i>
Precipitación	2017	1	23	18	<i>climatic_outliers</i>

El análisis de malas prácticas permite identificar errores sistemáticos en la medición y/o inscripción de datos por parte de los trabajadores de las estaciones. Entre los análisis más reveladores está la distribución semanal de datos y el WD frac asociado a cada uno de los días de la semana, permitiendo reconocer la existencia de tendencias en el registro de datos, como la ausencia de información en días no hábiles. Otra imagen interesante de observar es la distribución del último dígito de los datos, identificando abusos en el redondeo de los datos en contra de la precisión de estos.

Similar a la sección anterior, se analizarán dos casos, uno con resultados satisfactorios y otra estación con notorias malas prácticas.

- **Buenas prácticas: Estación Privada (Ar_1_ma)**

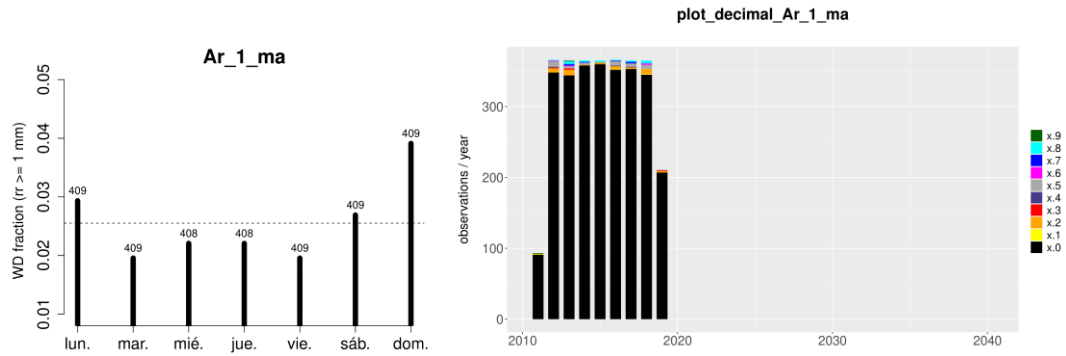


Figura B. 1 a,b: Distribución de datos de la estación Ar_1_ma. En a) se observa la distribución de los días de semana junto al WD frac de cada día. En b) se muestra la distribución del último dígito en la estación.

En a) se puede apreciar que todos los días de la semana tienen la misma cantidad de datos y que el WD frac cuenta con una amplia dispersión de sus datos con respecto al promedio. Por otra parte, en b) se observa una buena variación de últimos dígitos, fuera de cualquier redondeo sistemático que disminuya la precisión de los datos.

No se puede concluir que se realizan malas prácticas de forma sistemática, pero, al desconocer la real configuración de los datos y presentar una dispersa distribución de la fracción de días húmedos, no se puede afirmar que es una estación exenta de malas prácticas.

- **Malas prácticas: Estación Privada 2 (Ch_1_sqm)**

Del análisis de la Figura B. 1.a se puede observar una dispereja distribución de información a lo largo de los días de la semana, oscilando entre los 1187 a los 1165 datos en los jueves y domingos respectivamente, seguido de cerca por los sábados, con la segunda minoría entre el total con 1181 datos.

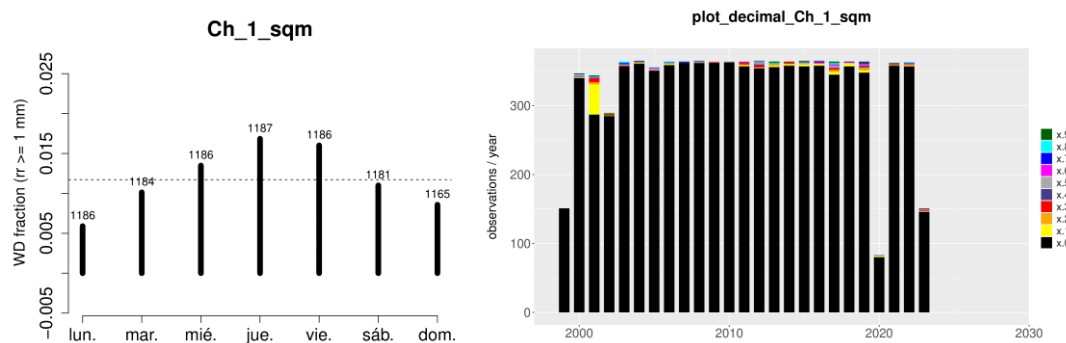


Figura B. 2 a,b: Distribución de datos de la estación Ch_1_sqm. En a) se observa la distribución de los días de semana junto al WD frac de cada día. En b) se muestra la distribución del último dígito en la estación.

Luego, observando la Figura B. 2.b, se reconocen años con buena distribución de últimos dígitos, como el 2001, pero luego aparecen años sin ninguna variedad de decimales fuera del 0, como en los años 2009 y 2010, para volver a observar años con correcta distribución de dígitos a partir de

2011. Además, se observa que en los años 2002 y 2020 se registraron menos de 300 días de información

De estas dos figuras se puede presumir un periodo de descuido en el ingreso de datos, tanto en los fines de semana, como en algunos años específicos ya mencionados, y en la precisión de estos, redondeando datos al número entero más cercano.

Resumiendo, siguiendo las recomendaciones de (Estévez et al., 2022), se debe considerar un porcentaje de al menos el 90% de los datos anuales de cada estación. Observando la Tabla 4.2, considerando ese porcentaje y dependiendo de la cantidad de años seguidos de información, se tiene un rango de 7 a 85 estaciones. Considerando 30 años seguidos, se pueden recopilar 43 estaciones meteorológicas.

Las estaciones anteriormente mencionadas destacan por su buena calidad en los análisis cuantitativos. Sin embargo, no están exentas de malas prácticas, destacando redondeos sistemáticos en las estaciones públicas chilenas; las argentinas, por el contrario, presentan buena calidad en cuanto a malas prácticas humanas se refiere. Por la misma razón, se debe tener en consideración el uso de cada estación por periodos, descartando aquellos que puedan estar alterados.

ANEXO C: DATOS DE ENTRADA AL MODELO

En el presente anexo, se adjuntan las matrices de correlación de variables para áreas de estudio más acotadas. En Figura C.1, se muestra la interacción de las variables en las estaciones en territorio chileno, mientras que la

Figura C.2 hace lo mismo en tierras argentinas.

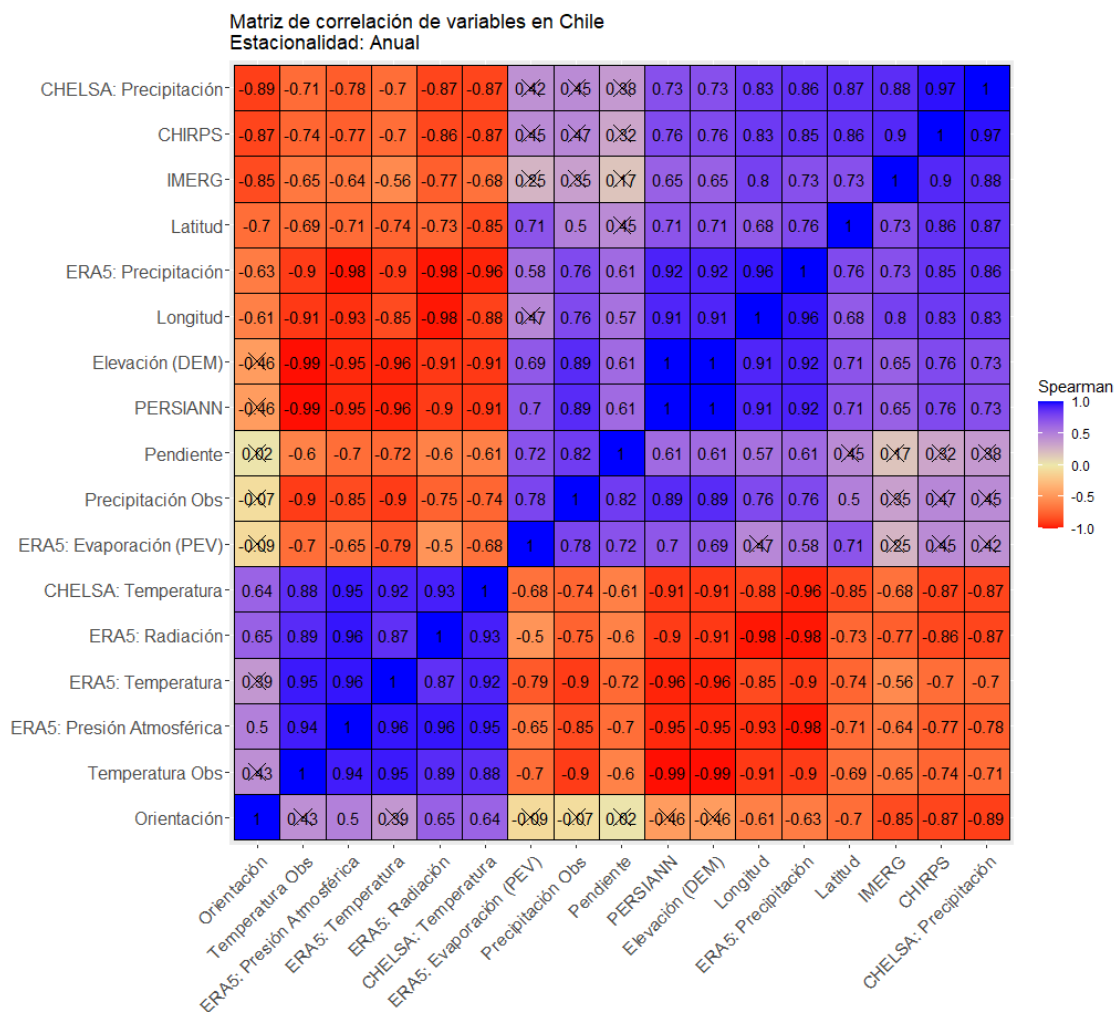


Figura C.1: Correlación de las variables de estudio a escala anual en el área chilena del estudio. Se considera el periodo entre 2007 y 2016.

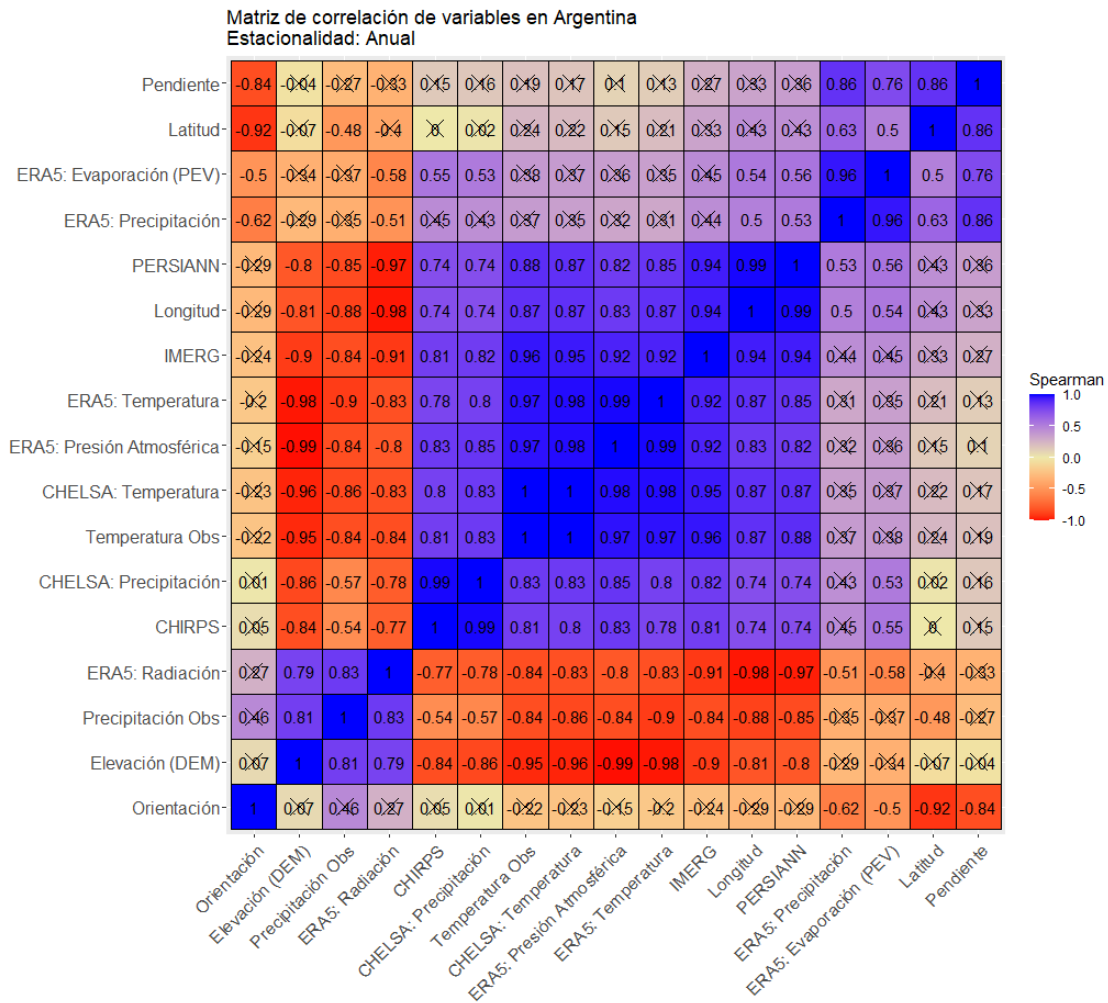


Figura C.2: Correlación de las variables de estudio a escala anual en el área argentina del estudio. Se considera el periodo entre 2007 y 2016.

ANEXO D: CURVAS DE VARIACIÓN ESTACIONAL

En el presente anexo, se adjuntan las CVE de todas las estaciones utilizadas en el estudio, similares a las realizadas en la sección de RESULTADOS, en la Figura 6.9. En Figura C.1, se muestra la interacción de las variables en las estaciones en territorio chileno, mientras que la

Figura C.2 hace lo mismo en tierras argentinas.

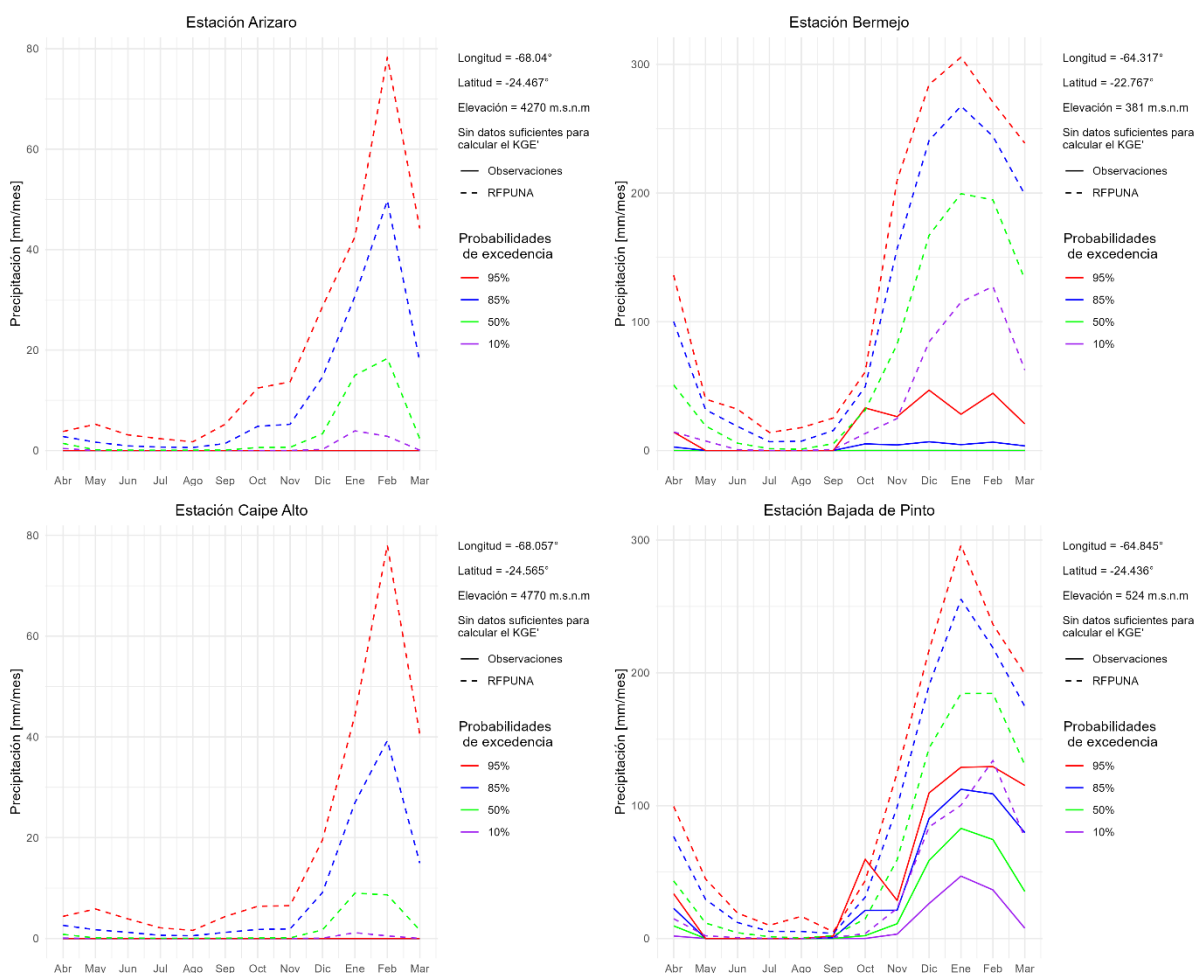


Figura D.0.1: Curvas de variación estacional para las estaciones Arizaro, Bermejo, Bajada de Pinto y Caipe Alto para 4 distintas probabilidades de excedencia para el periodo 2007-2017.

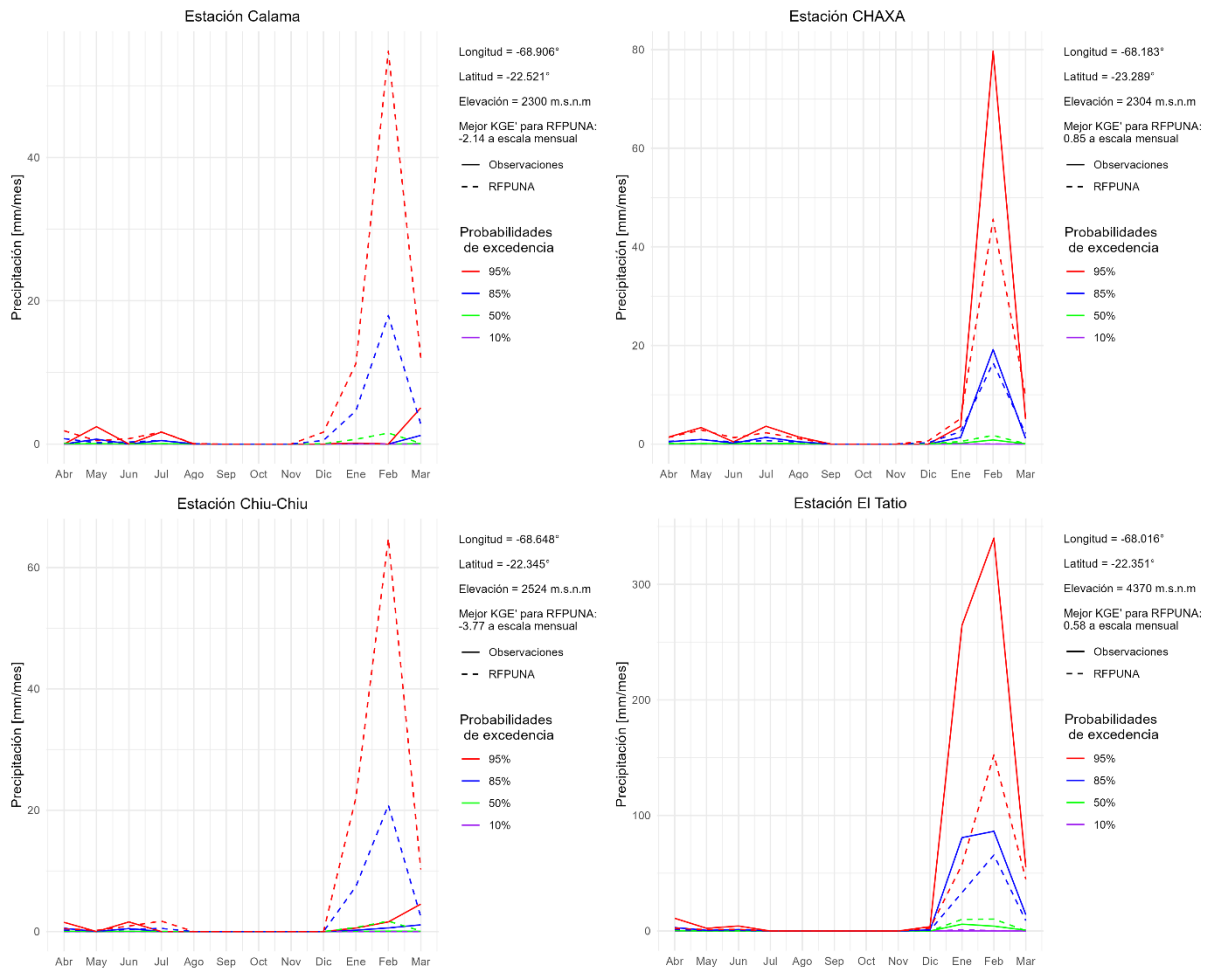


Figura D.0.2: Curvas de variación estacional para las estaciones Calama, CHAXA, Chiu-Chiu y El Tatio para 4 distintas probabilidades de excedencia para el periodo 2007-2017.

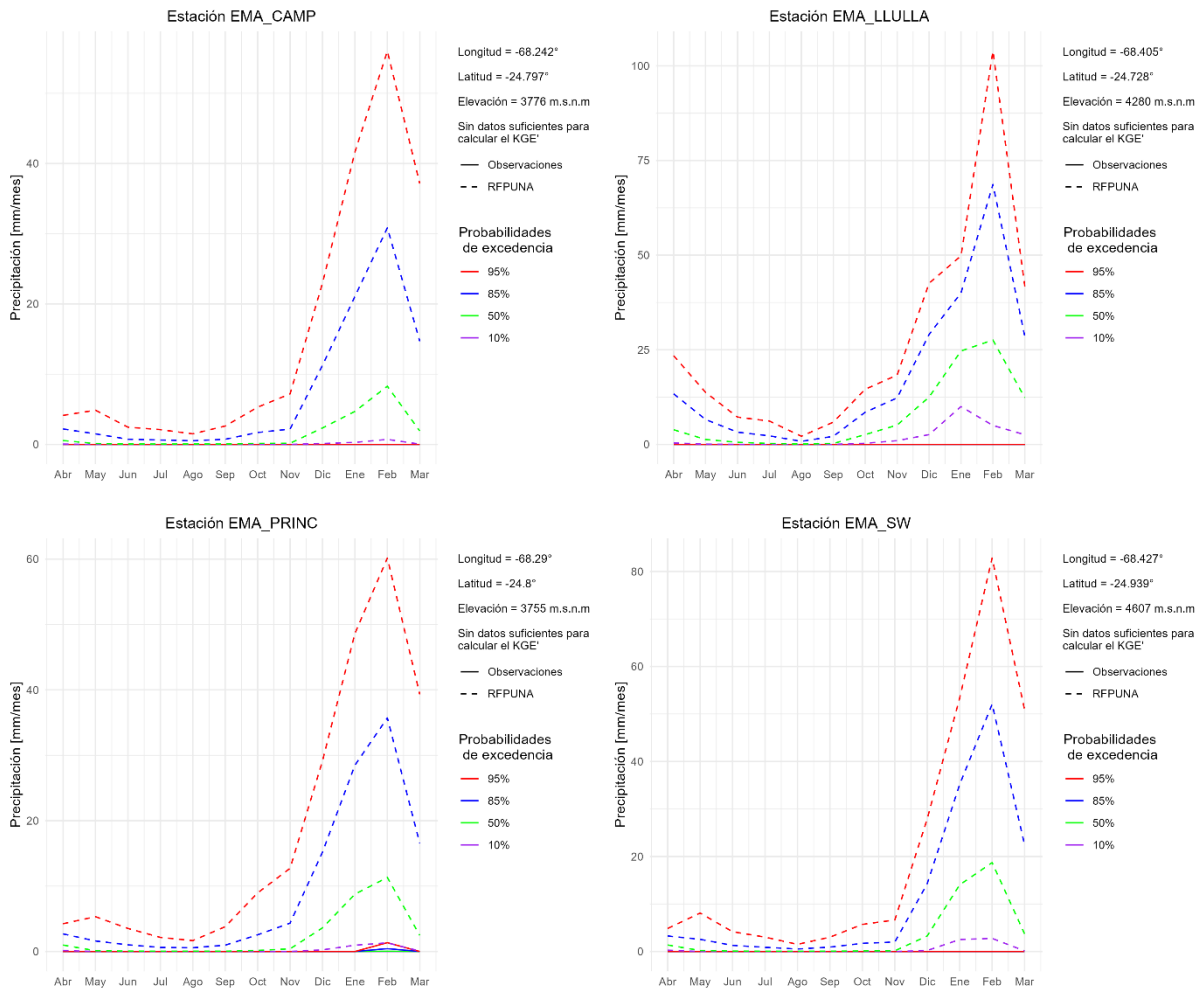


Figura D.0.3: Curvas de variación estacional para las estaciones EMA CAMP, EMA LLULLA, EMA PRINC y EMA SW para 4 distintas probabilidades de excedencia para el periodo 2007-2017.

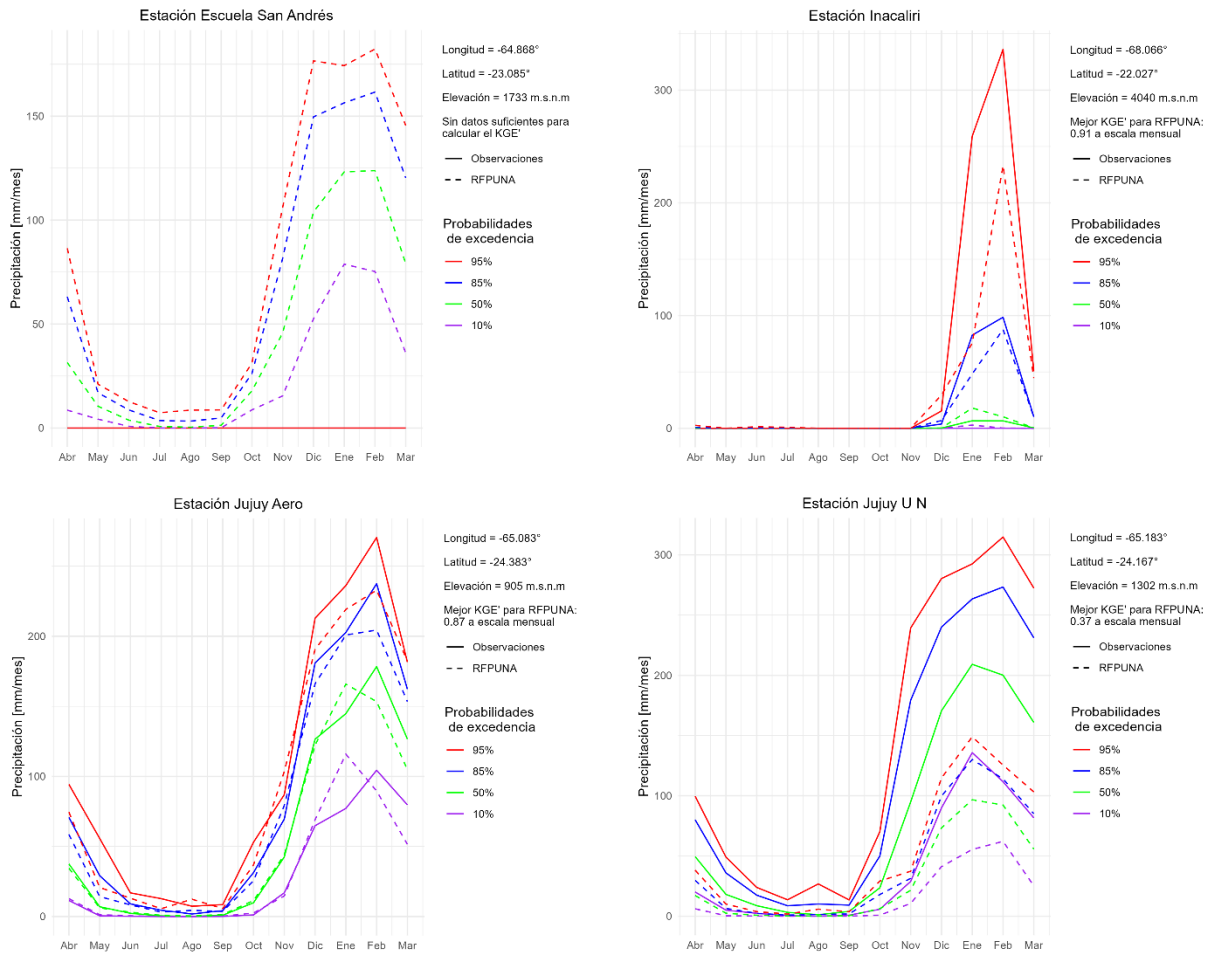


Figura D.0.4: Curvas de variación estacional para las estaciones Escuela San Andrés, Inacaliri, Jujuy Aero y JUJUY U N para 4 distintas probabilidades de excedencia para el periodo 2007-2017.

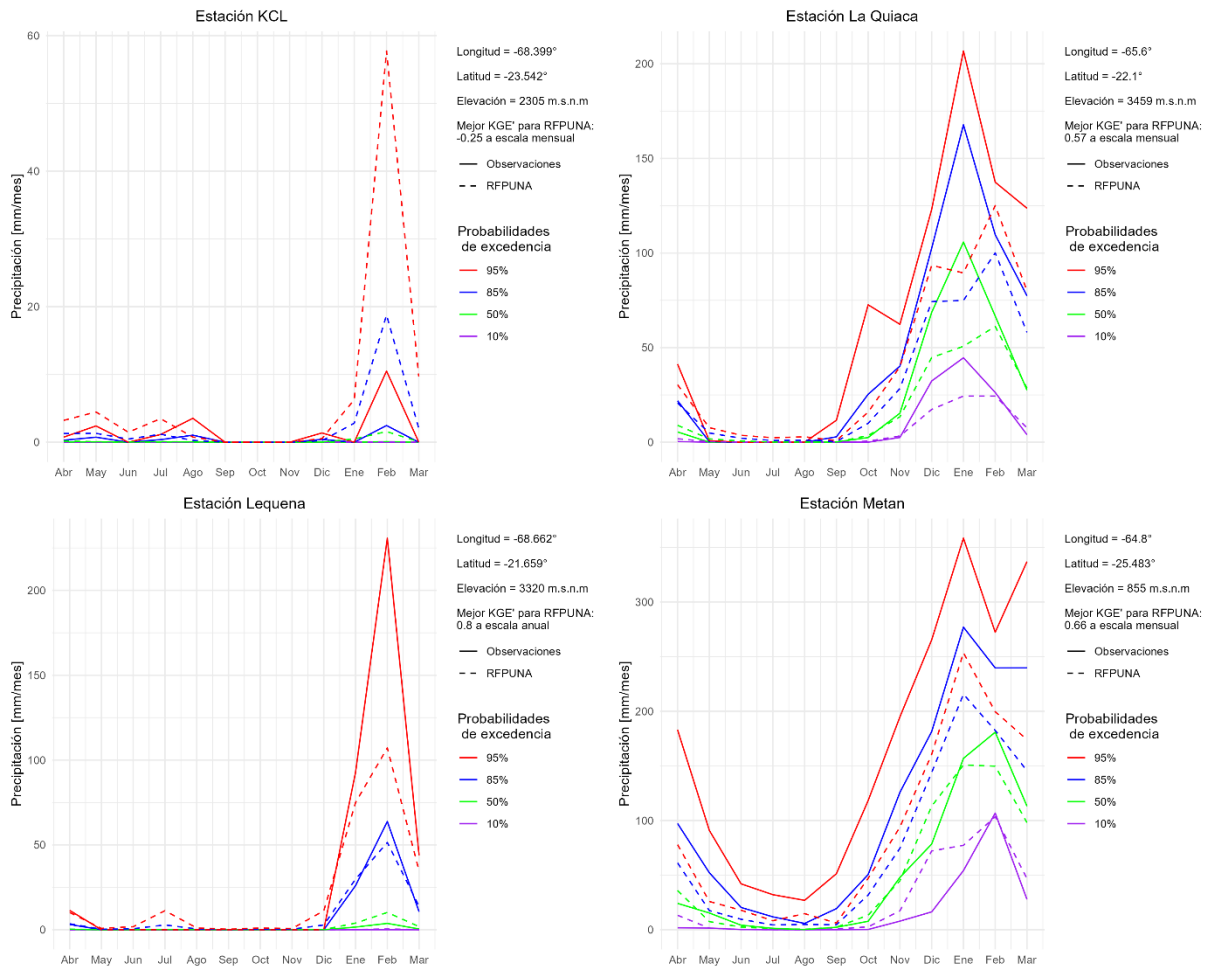


Figura D.0.5: Curvas de variación estacional para las estaciones KCL, La Quiaca, Lequena y Metan para 4 distintas probabilidades de excedencia para el periodo 2007-2017.

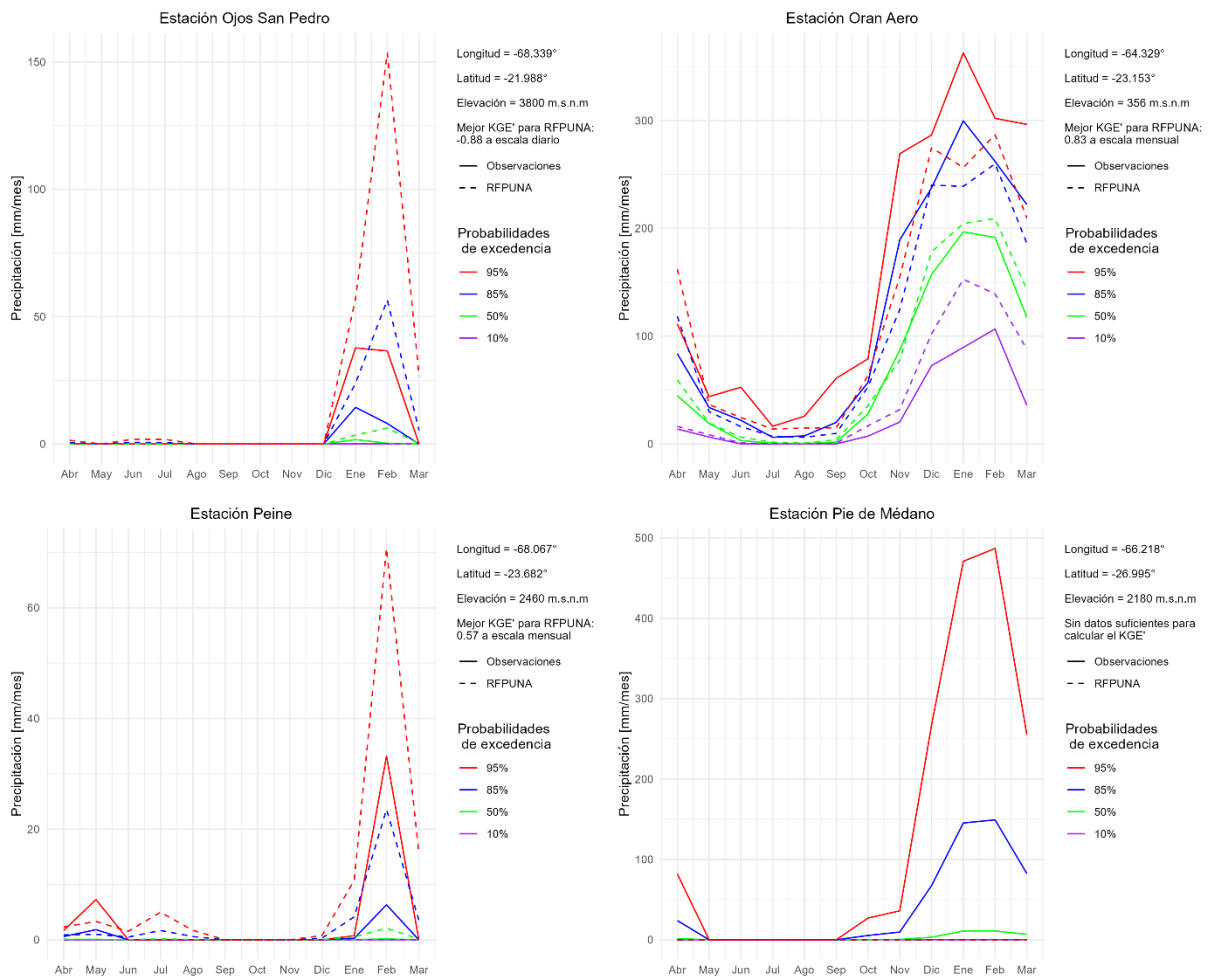


Figura D.0.6: Curvas de variación estacional para las estaciones Ojos San Pedro, Oran Aero, Peine y Pie de Médano para 4 distintas probabilidades de excedencia para el periodo 2007-2017.

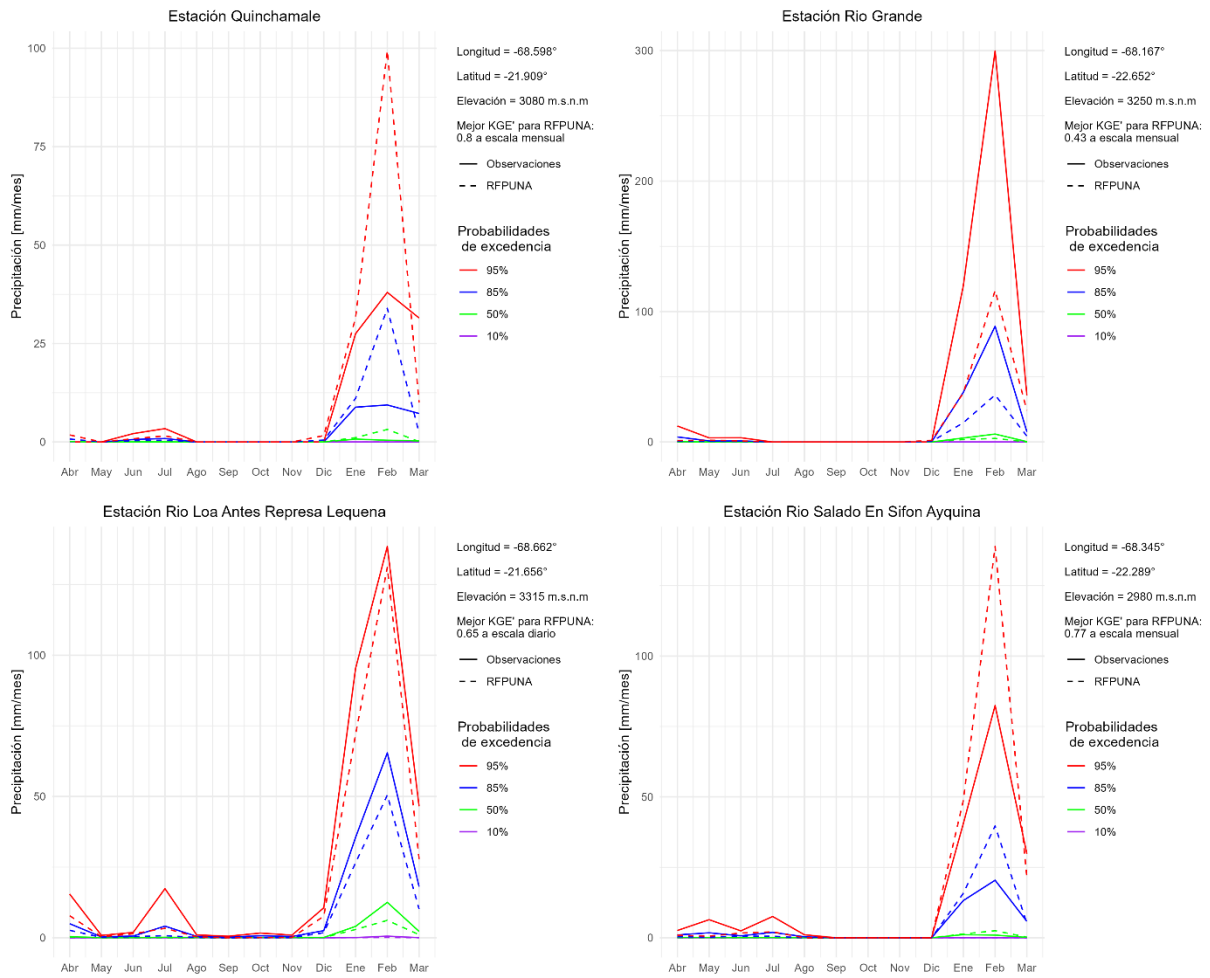


Figura D.0.7: Curvas de variación estacional para las estaciones Quinchamale, Río Grande, Río Los Andes antes Represa Lequena y Río Salado en Sifón Ayquina para 4 distintas probabilidades de excedencia para el periodo 2007-2017.

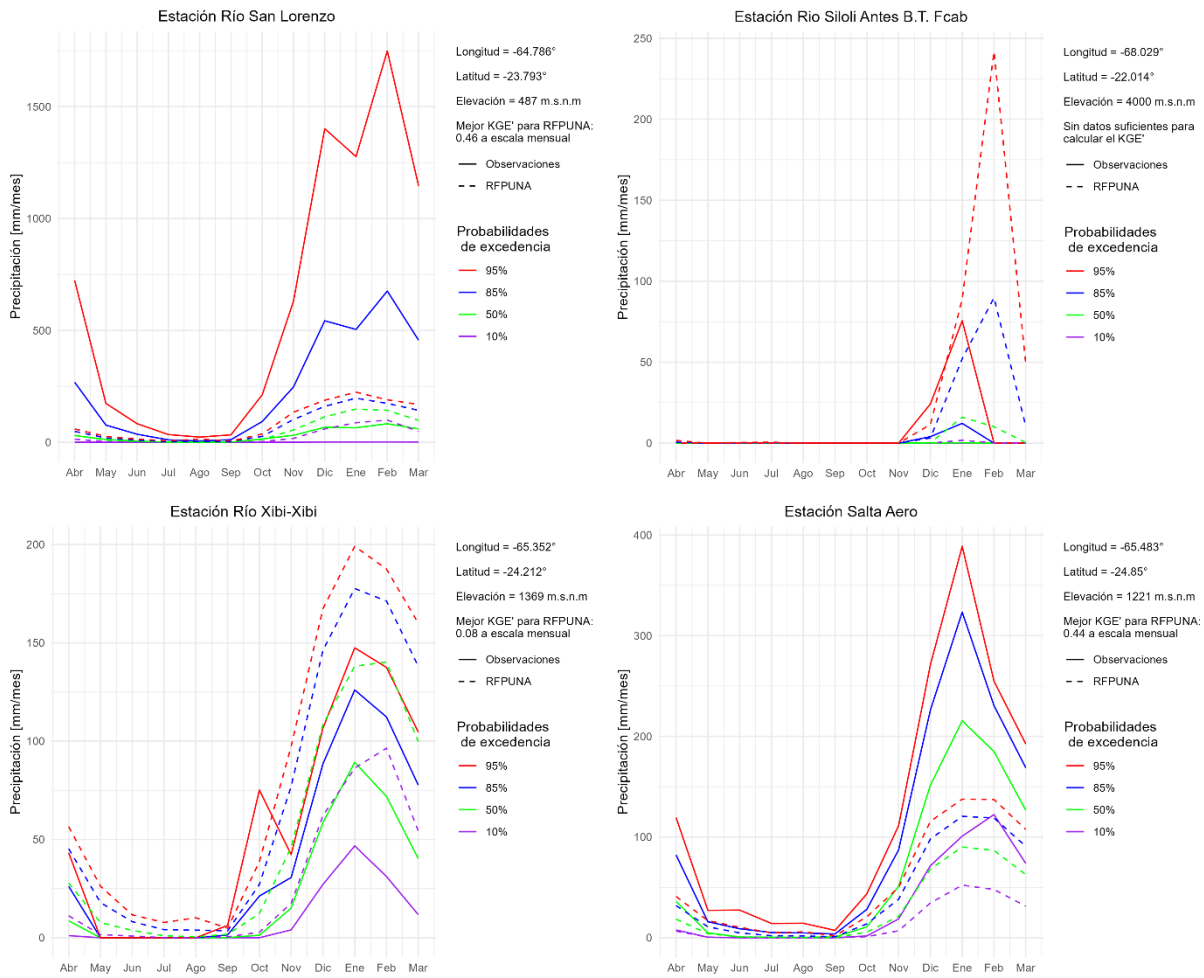


Figura D.0.8: Curvas de variación estacional para las estaciones Río San Lorenzo, Río Siloli antes B.T. Fcab, Río Xibi-Xibi y Salta Aero para 4 distintas probabilidades de excedencia para el periodo 2007-2017.

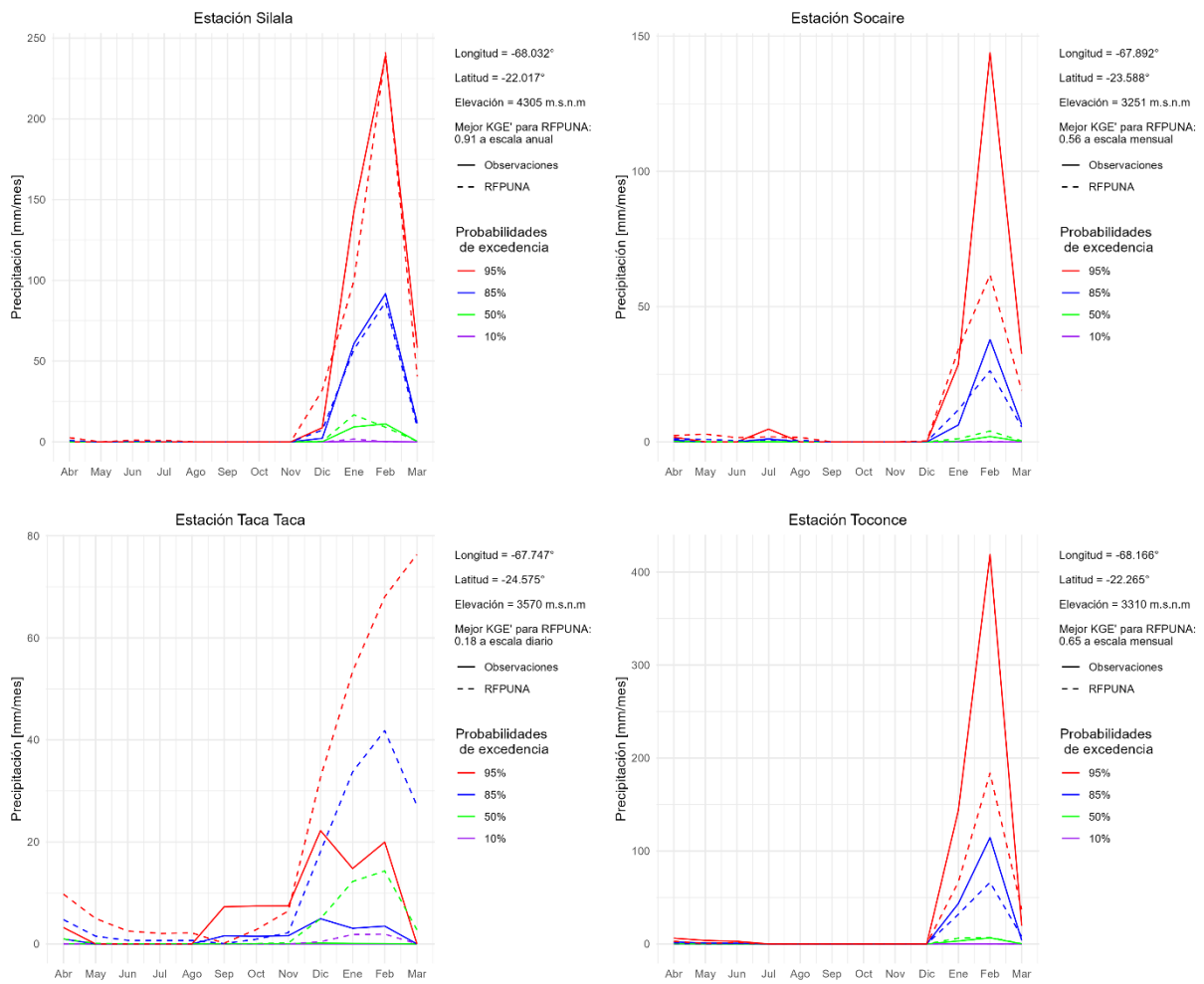


Figura D.0.9: Curvas de variación estacional para las estaciones Silala, Socaire, Taca Taca y Toconce para 4 distintas probabilidades de excedencia para el periodo 2007-2017.

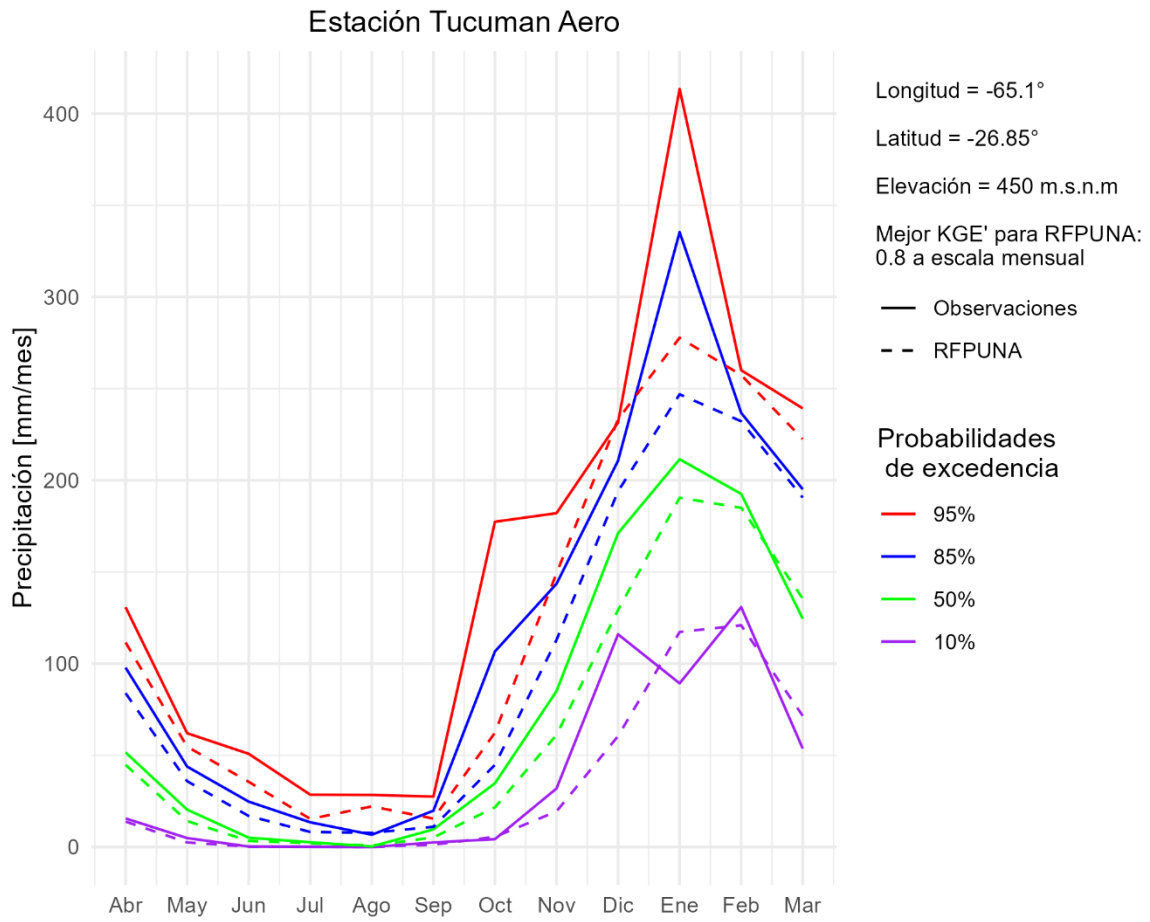


Figura D.0.10: Curvas de variación estacional para la estación Tucumán Aero para 4 distintas probabilidades de excedencia para el periodo 2007-2017.