

UNIVERSIDAD NACIONAL DE INGENIERIA

FACULTAD DE INGENIERÍA CIVIL



TESIS

**APLICACIÓN DE DISTANCIAS NO EUCLÍDEAS EN LA
INTERPOLACIÓN ESPACIAL DE PRECIPITACIÓN**

PARA OBTENER EL TÍTULO PROFESIONAL DE INGENIERO CIVIL

ELABORADO POR

ERICK KEVIN CLAROS CHAGUA

ASESOR

Dr. JULIO KUROIWA ZEVALLOS

LIMA- PERÚ

2022

© 2022, Universidad Nacional de Ingeniería. Todos los derechos reservados

**“El autor autoriza a la UNI a reproducir de la Tesis en su totalidad o en parte,
con fines estrictamente académicos.”**

Claros Chagua, Erick Kevin

erick.claros@uni.pe

972961876

Dedicatoria

A mi familia, por siempre estar a mi lado y brindarme su apoyo en todo este tiempo. A Dios, por guiar mi camino y cuidarme. A todas las personas que siempre están conmigo.

Agradecimientos

Al Dr. Julio Kuroiwa Zevallos, asesor de la presente tesis, por su apoyo y consejos brindados durante el desarrollo de este trabajo de investigación y en los años de mi formación universitaria.

A los docentes que contribuyeron a mi aprendizaje, a comprender mejor los principios de la Hidráulica e Hidrología y en particular ayudándome a resolver ciertas dudas durante el desarrollo de esta tesis.

A mis amigos de la Universidad Nacional de Ingeniería y del Laboratorio Nacional de Hidráulica con los cuales compartí grandes experiencias.

Al Laboratorio Nacional de Hidráulica por todo el apoyo brindado y en particular por la disponibilidad del clúster de alto rendimiento computacional Tipón, obtenido mediante el CONVENIO N° 257-INNOVATEPERU-EC- 2016.

Al Instituto de Investigación de la Facultad de Ingeniería Civil de la Universidad Nacional de Ingeniería, por el financiamiento durante el desarrollo de la presente investigación.

ÍNDICE

RESUMEN.....	4
ABSTRACT	6
PRÓLOGO	8
LISTA DE CUADROS.....	10
LISTA DE FIGURAS.....	11
LISTA DE SÍMBOLOS Y SIGLAS	13
CAPITULO I: INTRODUCCIÓN	15
1.1 GENERALIDADES	15
1.2 PROBLEMÁTICA.....	16
1.3 OBJETIVOS	16
1.3.1 Objetivo General.....	16
1.3.2 Objetivos Específicos.	16
1.4 ANTECEDENTES INVESTIGATIVOS	17
CAPITULO II: MARCO TEÓRICO Y CONCEPTUAL	19
2.1. MARCO TEÓRICO.....	19
2.1.1 Métodos de interpolación determinísticos.....	19
2.1.1.1 Polígonos de Thiessen	19
2.1.1.2 IDW	20
2.1.1.3 IDW con corrección de gradiente altitudinal.....	21
2.1.2 Métodos de interpolación geoestadísticos	22
2.1.2.1 Estimación de la correlación espacial.....	22
2.1.2.2 Kriging ordinario	27
2.1.2.3 Kriging con deriva externa	29
2.1.3 Validación cruzada	31
2.1.4 Índices de desempeño.....	32
2.1.4.1 Eficiencia de Nash-Sutcliffe	32
2.1.4.2 Raíz del error medio cuadrático.....	32

2.1.4.3	Coeficiente de correlación de Pearson	33
2.1.4.4	Error de sesgo.....	33
2.1.4.5	Eficiencia de Kling-Gupta	33
2.1.5	Algunas aproximaciones de distancias no euclídeas	34
2.1.5.1	Distancia de perfil topográfico.....	34
2.1.5.2	Distancia generalizada más corta.....	35
2.1.6	Modelos hidrológicos	36
2.1.6.1	Calibración y validación	37
2.1.6.2	Generalidades del modelo GR2M.....	38
2.1.6.3	Descripción del modelo GR2M	39
2.2	MARCO CONCEPTUAL.....	41
2.2.1	Precipitación.....	41
2.2.2	Temperatura.....	42
2.2.3	Evapotranspiración.....	42
2.2.4	Caudal de escorrentía superficial.....	43
2.2.5	Variabilidad espacial.....	43
2.2.6	Interpolación espacial.....	43
2.2.7	Distancia no euclídea	44
CAPITULO III: ÁREA DE ESTUDIO E INFORMACIÓN		45
3.1	FUENTES DE INFORMACIÓN.....	45
3.1.1	Topografía	45
3.1.2	Hidrometeorología	45
3.2	ÁREA DE ESTUDIO	45
3.3	INFORMACIÓN DE PRECIPITACIÓN.....	47
3.4	ANÁLISIS DE LA INFORMACIÓN HIDROMETEOROLÓGICA.....	51
3.4.1	Valores extremos.....	51
3.4.2	Quiebres y segmentos inconsistentes obvios	53
3.4.3	Homogeneidad de los datos	54

CAPITULO IV: MÉTODOS Y DESARROLLO	57
4.1 ZONAS DE ANÁLISIS	57
4.2 DISTANCIAS NO EUCLÍDEAS PROPUESTAS.....	58
4.2.1 Distancia no euclídea 1 (NE 1)	59
4.2.2 Distancia no euclídea 2 (NE 2)	61
4.2.2.1 Primera etapa.....	62
4.2.2.2 Segunda etapa	62
4.2.3 Distancia no euclídea 3 (NE 3)	64
4.3 PROCESO DE INTERPOLACIÓN.....	66
4.3.1.1 Selección de λ	68
4.3.1.2 Ejemplos.....	68
4.3.1.3 Similitud con otras aproximaciones.....	70
4.4 EVALUACIÓN DE PERFORMANCE	71
4.4.1 Validación cruzada	71
4.4.2 Comparación con otros métodos	71
4.5 UNA APLICACIÓN EN MODELAMIENTO HIDROLÓGICO	74
CAPITULO V: ANÁLISIS Y DISCUSIÓN DE RESULTADOS.....	77
5.1 EVALUACIÓN DE DESEMPEÑO	77
5.2 COMPARACIÓN CON OTROS MÉTODOS	80
5.3 EJEMPLOS	86
5.4 DESEMPEÑO EN MODELAMIENTO HIDROLÓGICO	90
CONCLUSIONES.....	96
RECOMENDACIONES.....	98
REFERENCIAS BIBLIOGRÁFICAS.....	99
ANEXOS.....	107
Anexo 1.....	107
Anexo 2.....	116

RESUMEN

Cuando se formulan modelos hidrológicos para predecir la escorrentía es fundamental caracterizar la distribución espacial y temporal de la precipitación. Sin embargo, la baja densidad de estaciones es un limitante para lograr este propósito. Por este motivo se debe emplear información adicional que esté disponible para mejorar la estimación de la precipitación.

En la presente investigación se estudió la precisión de un método de interpolación espacial, que hace uso de distancias no euclídeas asociadas a la elevación, para la estimación de precipitación a paso mensual (periodo 1970-2017) en la región andina peruana. Este método considera un componente lineal que representa el gradiente altitudinal de la precipitación y un componente calculado por una ponderación en base a distancias no euclídeas. En esta aplicación, las distancias no euclídeas pretenden representar el efecto de accidentes y separaciones topográficas. Considerando que el cálculo de las distancias no euclídeas planteadas puede ser computacionalmente pesado, se evaluaron tres variantes. La primera distancia no euclídea representa la menor distancia por sobre el perfil topográfico entre dos ubicaciones, mientras que la segunda variante se estima como la longitud del más corto recorrido espacial entre estas ubicaciones considerando posibles trayectorias en la superficie topográfica y atmósfera. La tercera variante sigue un proceso de cálculo similar a la segunda, con la diferencia de que los recorridos en la atmósfera no están permitidos. Además, dada la gran variabilidad espacial, se consideraron doce zonas en el área de estudio, seis en la vertiente del Pacífico, cinco en la vertiente del Atlántico y una en la vertiente del Titicaca. Asimismo, los resultados de esta metodología fueron comparados con los de otros cinco métodos, tres de los cuales hacen uso de la elevación; y se evaluó su aplicación en el modelamiento hidrológico de dos cuencas con el modelo GR2M.

Los resultados obtenidos muestran que el uso del método en base a distancias no euclídeas permite obtener una estimación más precisa de la precipitación en comparación a otras metodologías en casi todas las zonas; y entre las tres variantes planteadas, la segunda muestra mejor performance. Asimismo, los métodos de interpolación que usan la elevación mostraron un mejor desempeño en comparación a los otros métodos en la mayoría de las zonas, especialmente

en las que hay mayores cambios topográficos. Además, en la evaluación realizada del modelamiento hidrológico de dos cuencas, los resultados fueron también satisfactorios en comparación a otros métodos de interpolación.

ABSTRACT

When hydrological models are formulated to predict runoff, it is essential to characterize the spatial and temporal distribution of precipitation. However, the low density of stations is a limitation to achieve this purpose. For this reason, additional information that is available should be used to improve the estimation of precipitation.

In the present investigation, the precision of a spatial interpolation method was studied, which makes use of non-Euclidean distances associated with elevation, for the estimation of precipitation at a monthly rate (period 1970-2017) in the Peruvian Andean region. This method considers a linear component that represents the elevational gradient of precipitation and a component calculated by weighting based on non-Euclidean distances. In this application, non-Euclidean distances are intended to represent the effect of topographic features and separations. Considering that the calculation of the proposed non-Euclidean distances can be computationally heavy, three variants were evaluated. The first non-Euclidean distance represents the shortest distance above the topographic profile between two locations, while the second variant is estimated as the length of the shortest spatial path between these locations considering possible trajectories on the topographic surface and atmosphere. The third variant follows a calculation process similar to the second, with the difference that runs in the atmosphere are not allowed. In addition, given the great spatial variability, twelve zones were considered in the study area, six on the Pacific slope, five on the Atlantic slope, and one on the Titicaca slope. Likewise, the results of this methodology were compared with those of five other methods, three of which make use of elevation; and its application in the hydrological modeling of two basins with the GR2M model was evaluated.

The results obtained show that the use of the method based on non-Euclidean distances allows obtaining a more precise estimate of precipitation compared to other methodologies in almost all areas; and among the three variants proposed, the second shows better performance. Likewise, the interpolation methods that use elevation showed better performance compared to the other methods in most areas, especially in those with greater topographic changes. In addition, in the

evaluation of the hydrological modeling of two basins, the results were also satisfactory in comparison to other interpolation methods.

PRÓLOGO

El Perú tiene una superficie de aproximadamente 1.3 millones de km², que lo ubica entre los veinte países de mayor superficie en el mundo. Posee, además, una gran diversidad climática. El geógrafo Javier Pulgar Vidal identificó ocho regiones naturales, que van desde zonas hiperáridas hasta bosques lluviosos e incluso en las últimas décadas se ha verificado la existencia de una gran variedad de microclimas. Aunque el Perú es un país extenso y con una gran variedad de climas y microclimas la densidad de estaciones climáticas es muy baja. En los años 60 del siglo pasado se instalaron una serie de estaciones climatológicas ordinarias, pluviográficas y pluviométricas en diferentes zonas del Perú y en los años 80 una gran parte de ellas fueron desactivadas. La Organización Meteorológica Mundial sugiere que cada estación pluviométrica cubra entre 600 y 900 km² y en zonas montañosas cubra entre 100 y 250 km². En el Perú el SENAMHI ha publicado datos de 370 estaciones meteorológicas que miden precipitación lo que, en promedio, arroja un valor de 3,473 mm por estación, valor que se encuentra muy por encima del valor recomendado.

Ante la falta de datos se emplea una serie de métodos como el polígono de Thiessen, Kriging, etc. Sin embargo, estos métodos no toman en cuenta fenómenos como el efecto de la convección orográfica en la precipitación. En mi experiencia personal he podido comprobar que, aunque dos estaciones se encontraban a solo 8 km de distancia, la precipitación medida durante una estación lluviosa en una de ellas era 2.4 veces la precipitación medida en la otra debido principalmente a la diferencia en altitud, en el orden de 1200 m.

Con la finalidad de mejorar la estimación de las precipitaciones para modelación hidrológica, la tesis del Sr. Bachiller Erick Claros Chagua propone tres métodos de estimación de precipitación basándose en la metodología de distancias no euclídeas.

Finalmente, se emplea el modelo hidrológico GR2M en dos cuencas para demostrar la bondad del uso del método de distancia no euclídea que mejor performance arrojó en la evaluación de los tres métodos de distancia no euclídea propuestos.

Los resultados de la modelación hidrológica fueron prometedores. Se pudo reproducir adecuadamente los hidrogramas de dos estaciones ubicadas en la vertiente del Pacífico, tanto en la Costa Central como en la Costa Sur, obteniéndose valores altos en los índices de performance en la calibración y validación de los hidrogramas obtenidos.

Se espera que esta contribución constituya una herramienta útil que pueda ser empleada por los profesionales del área de recursos hídricos en regiones de topografía y geografía compleja como las que existen a lo largo de la región andina.

Dr. Julio M. Kuroiwa Zevallos
Docente Principal TC
Departamento Académico de Hidráulica e
Hidrología

LISTA DE CUADROS

Tabla N°	Título	Página N°
1	Datos adicionales de zonas de análisis	58
2	Periodos de calentamiento, calibración y validación	76
3	Índices de desempeño medios para interpolación propuesta considerando opciones de distancias no euclídeas – Vertiente del Pacífico (zonas de la 1 a la 6)	77
4	Índices de desempeño medios para interpolación propuesta considerando opciones de distancias no euclídeas – Vertiente del Atlántico y Titicaca (zonas de la 7 a la 12)	78
5	Índices de desempeño medios para métodos de interpolación de comparación – Vertiente del Pacífico (zonas de la 1 a la 6)	80
6	Índices de desempeño medios para métodos de interpolación de comparación – Vertiente del Atlántico y Titicaca (zonas de la 7 a la 12)	81
7	Performance (Nash) de modelamiento hidrológico con diferentes series de precipitación – Cuenca Caplina	93
8	Performance (KGE / Bias [%]) de modelamiento hidrológico con diferentes series de precipitación – Cuenca Obrajillo	93
9	Performance (Nash) de modelamiento hidrológico con diferentes series de precipitación – Cuenca Obrajillo	95
10	Performance (KGE / Bias [%]) de modelamiento hidrológico con diferentes series de precipitación – Cuenca Obrajillo	95

LISTA DE FIGURAS

Figura N°	Título	Página N°
1	Interpolación mediante distintos métodos. Polígonos de Thiessen (a), IDW $p = 1$ (b), IDW $p = 2$ (c), IDW $p = 5$ (d) e IDW con corrección por altitud (e)	20
2	Esquema de interpolación mediante metodología propuesta por Valery (2010)	22
3	Nube de semivarianzas (arriba) y variograma experimental o empírico (abajo).	24
4	Componentes típicos de un variograma: pepita, meseta y rango.	25
5	Ejemplos de variogramas teóricos	26
6	Interpolación usando el método de Kriging ordinario (a) y Kriging con deriva externa (b)	29
7	Distancia de perfil topográfico en comparación a distancias euclídeas.	34
8	Distancia generalizada más corta D_λ para diferentes coeficientes λ desde una estación en valle (izquierda) y una en montaña (derecha)	36
9	Esquema de la cuenca y variables del modelo GR2M Esquema de modelo GR2M	38
10	Esquema de modelo GR2M	39
11	Área de estudio, mapa de elevaciones SRTM y red de estaciones	46
12	Disponibilidad en el tiempo de estaciones de la base de datos	47
13	Densidad de elevaciones de estaciones y DEM en el área de estudio	48
14	Precipitación media anual en estaciones	49
15	Número de años de registro en estaciones	50
16	Ejemplos de filtrado de estaciones con valores extremos	52
17	Ejemplos de filtrado de estaciones con segmentos inconsistentes	53
18	Coefficiente de correlación para estaciones seleccionadas	56
19	Zonas de análisis y estaciones	57
20	Distancia mínima por sobre el perfil longitudinal	59
21	Distancia no euclídea 1 (NE 1) desde estación en valle (Yauyos)	60
22	Distancia no euclídea 1 (NE 1) desde estación montañosa (Milloc)	61
23	Distancia no euclídea 2 (NE 2) desde estación en valle (Yauyos)	63
24	Distancia no euclídea 2 (NE 2) desde estación montañosa (Milloc)	64
25	Distancia no euclídea 3 (NE 3) desde estación en valle (Yauyos)	65
26	Distancia no euclídea 3 (NE 3) desde estación montañosa (Milloc)	66
27	Proceso de interpolación en zona 3 usando distancias NE 2 (1982-12)	68
28	Proceso de interpolación en zona 3 usando distancias NE 2 (2009-01)	69

29	Proceso de interpolación en zona 3 usando distancias NE 2 (1980-08)	70
30	Cálculo de variogramas en noviembre de 2009 para KED en zona 8: en base a validación cruzada para algunas estaciones (a) y considerando todos los datos (b)	73
31	Localización de cuencas de estudio	74
32	Interpolación de temperatura y cálculo de evapotranspiración potencial - cuenca Obrajillo	76
33	Diagrama de cajas de distribución de índices de desempeño por zonas.	79
34	Diagrama de cajas de índices de desempeño para los métodos de interpolación evaluados (zonas de la 1 a la 4)	83
35	Diagrama de cajas de índices de desempeño para los métodos de interpolación evaluados (zonas de la 5 a la 8)	84
36	Diagrama de cajas de índices de desempeño para los métodos de interpolación evaluados (zonas de la 9 a la 12)	85
37	Mapas de precipitación obtenidos de métodos de interpolación en comparación en agosto de 1998 y febrero de 1999 – Zona 3	86
38	Mapas de precipitación obtenidos de métodos de interpolación en comparación en agosto de 2008 y febrero de 2009 – Zona 3	87
39	Mapas de precipitación obtenidos de métodos de interpolación en comparación en agosto de 1998 y febrero de 1999 – Zona 6	88
40	Mapas de precipitación obtenidos de métodos de interpolación en comparación en agosto de 2008 y febrero de 2009 – Zona 3	89
41	Interpolación de precipitación con diferentes métodos en la cuenca Caplina.	90
42	Interpolación de precipitación con diferentes métodos en la cuenca Obrajillo.	91
43	Simulación hidrológica mediante el modelo GR2M para diferentes series de precipitación - cuenca Caplina.	92
44	Simulación hidrológica mediante el modelo GR2M para diferentes series de precipitación - cuenca Obrajillo.	94

LISTA DE SÍMBOLOS Y SIGLAS

a	: rango de variograma
$Bias$: error de sesgo
C_0	: parámetro de efecto pepita en variograma
C_1	: parámetro asociado a la meseta ($C_0 + C_1$) en variograma
d	: distancia
$D_\lambda()$: función de distancia no euclídea asociada al coeficiente λ
$E():$: valor esperado
h	: distancia entre dos ubicaciones, usada para definir variogramas
Obs	: valor observado
$P_L()$: función de precipitación como modelo lineal dependiente de la elevación
r	: coeficiente de correlación de Pearson
$R()$: residual de estimación
s	: ubicación
Sim	: valor simulado
w	: peso para ponderación
z	: elevación
γ	: semivarianza
$\Gamma()$: función de trayectoria
ε	: error
θ	: parámetro relacionado al gradiente altitudinal
λ	: coeficiente de penalización de distancias verticales
μ	: media poblacional
σ^2	: varianza poblacional
ANA	: Autoridad Nacional del Agua
DEM	: digital elevation model Modelo de elevación digital
ETP	: evapotranspiración potencial
GEV	: generalized extreme value (distribution) (distribución) general de valores extremos
GR2M	: Génie Rural à 2 paramètres mensuel Ingeniería rural de 2 parámetros mensual

IDW	: Inverse Distance Weighted interpolation Interpolación ponderada por distancia inversa
IED	: IDW with External Drift IDW con deriva externa
KED	: Kriging with External Drift Kriging con deriva externa
KGE	: Kling-Gupta efficiency Índice de eficiencia de Kling-Gupta
MIDAGRI	: Ministerio de Desarrollo Agrario y Riego
NE	: (distancia) no euclídea
OK	: Ordinary Kriging Kriging ordinario
PISCO	: Peruvian Interpolated data of SENAMHI's Climatological and hydrological Observations Data peruana interpolada de observaciones climatológicas e hidrológicas de SENAMHI
RMSE	: Root Mean Square Error Raíz del error cuadrático medio
SENAMHI	: Servicio Nacional de Meteorología e Hidrología
SNIRH	: Sistema Nacional de Información de Recursos Hídricos
SRTM	: Shuttle Radar Topography Mission Misión Topográfica de Radar Shuttler
TRMM	: Tropical Rainfall Measuring Mission Misión de Medición de Lluvias Tropicales
VR	: vector regional
WMO	: World Meteorological Organization Organización Meteorológica Mundial

CAPITULO I: INTRODUCCIÓN

1.1 GENERALIDADES

La precipitación es la principal variable climática que se considera para el diseño y desarrollo de obras de ingeniería. En particular es fundamental para un adecuado modelamiento hidrológico, diseño hidráulico y gestión de los recursos hídricos. Sin embargo, en nuestro país la red de estaciones no es lo suficientemente densa para ser representativa en todo lugar, por ello se hace necesario usar métodos de interpolación adecuados o productos de percepción remota para suplir esta falta de información.

Es posible decir así que la precipitación es una de las principales fuentes de incertidumbre dentro de las variables de diseño y, sin embargo, tiene una muy significativa influencia en los fenómenos que permite modelar. Es así que en regiones montañosas como la del país se ha buscado relacionar la precipitación con otras variables espaciales, entre ellas la elevación, para poder estimarla así con mayor precisión.

Para Perú se han desarrollado investigaciones con el fin de hacer una mejor estimación de la distribución espacial. En ese sentido, Lavado y Fernández (2014) y Chamorro (2015) comparan distintos métodos de interpolación, obteniendo un mejor resultado en métodos que consideraban la elevación. Además, productos de percepción remota son usados por Aybar et al. (2020) para el desarrollo del producto grillado nacional de precipitación PISCO. Asimismo, Navarro (2019) y Mourre et al. (2016) hacen uso de estos productos para estimar la distribución espacial de la precipitación.

En zonas de características montañosas se han propuesto distintos métodos para estimar la precipitación. Valery et al. (2010) usa un gradiente altitudinal optimizado y su metodología es usada en Perú por Rau et al. (2019) para la vertiente del Pacífico. Asimismo, en zonas cercanas a los Alpes, Gottardi et al. (2012) y Frei (2014) usan distancias no euclídeas en la interpolación de precipitación y temperatura respectivamente.

1.2 PROBLEMÁTICA

La escasa distribución de estaciones en muchas zonas del país limita la adecuada estimación de la precipitación. Debemos considerar que la Organización Mundial de Meteorología recomienda al menos 1 estación cada 250 km² para zonas montañosas (World Meteorological Organization, 2008), una densidad mayor a la recomendada para otro tipo de zonas, pero esto es muy difícil de conseguir en la práctica.

En ese sentido, las zonas montañosas, que requieren una mayor densidad de estaciones debido a su variabilidad espacial, son las que por su inaccesibilidad se encuentran escasamente instrumentadas (Beniston et al., 1997; Condom et al., 2011). Esta falta de información se muestra aún más grave, si consideramos que en la vertiente del Pacífico es en las zonas altas donde se generan los mayores volúmenes de precipitación y por ello tienen una gran influencia en la estimación de caudales, oferta hídrica y eventos extremos.

Es importante notar que los Andes son un factor determinante en el desarrollo de la precipitación en nuestra región. Su orografía accidentada y de fuerte pendiente conduce a muchos de los procesos propios de esta región relacionados a la precipitación (Insel et al., 2010). Es así que estas características especiales, hacen complicado el uso de métodos tradicionales de interpolación, y ponen de manifiesto la necesidad de estudiar a profundidad metodologías que podrían ser adecuadas al caso indicado.

1.3 OBJETIVOS

1.3.1 Objetivo General

Evaluar el uso de métodos de interpolación espacial basados en distancias no euclídeas en la estimación de la precipitación mensual.

1.3.2 Objetivos Específicos.

- Identificar una distancia no euclídea que permita representar adecuadamente la conectividad entre dos ubicaciones.
- Comparar el desempeño de los métodos de interpolación propuestos con otras metodologías.

- Evaluar el uso de las interpolaciones propuestas en modelamiento hidrológico.

1.4 ANTECEDENTES INVESTIGATIVOS

Las técnicas para estimar espacialmente la precipitación han estado evolucionando constantemente debido al acceso a más datos, productos, y capacidad de procesamiento, es así que en el caso de zonas montañosas y en particular andinas se han evaluado distintas metodologías en investigaciones previas.

Es posible apoyarse en productos de percepción remota para la estimación de la precipitación. En la cuenca alta del río Piura se han utilizado datos de percepción remota de la Misión de Medición de Precipitaciones Tropicales (TRMM, por sus siglas en inglés) para distribuir espacialmente la precipitación a una resolución de 1km^2 entre los años 2000 y 2010 mediante el uso de redes neuronales. (Navarro, 2019).

En la Cordillera Blanca se ha evaluado la performance del método de interpolación de Kriging con deriva externa (KED, Kriging with external drift) frente al producto TRMM y a las salidas del modelo atmosférico WRF, mostrándose que los tres son capaces de representar los patrones espaciales de precipitación a escala anual. Sin embargo, estos productos no eran eficientes para estimar la precipitación a escalas cortas de tiempo, recomendándose el uso de nuevos métodos. (Mourre et al., 2016)

La relación entre elevación y precipitación ha resultado bastante útil para realizar mejores estimaciones. Así, por ejemplo, se han usado diferentes métodos de interpolación, para estimar la precipitación en las cuencas de Rímac, Chillón, Lurín y la parte alta del Mantaro, donde hubo una mejora significativa con los métodos que consideraban la elevación como variable auxiliar (Lavado y Fernández, 2014).

En cuencas de alta montaña de Francia y Suecia, se ha estimado la precipitación usando gradientes altitudinales con una interpolación de distancia inversa ponderada (IDW, por sus siglas en inglés) lo que permitió representar con mayor precisión esta variable, cerrar el balance de aguas en cuencas donde no se podía y optimizar el modelamiento hidrológico (Valéry et al., 2010).

Esta última metodología es también aplicada en la vertiente del Pacífico de Perú por Rau (2019) con fines de modelamiento hidrológico. Además, es evaluada en la región Norte Chico de Chile, donde muestra una mejora significativa frente a otros métodos de interpolación determinísticos y los coeficientes de escorrentía se hacen más realistas. (Bourgin, 2010)

En Finlandia se han examinado diferentes métodos para realizar la interpolación espacial de precipitación y temperatura a paso mensual para un periodo de tiempo de 30 años. Entre los métodos usados, el método de Kriging con deriva externa (KED, Kriging with external drift) y el modelo aditivo generalizado con Kriging residual fueron los que mostraron mejores resultados. Eligiéndose finalmente el KED para la interpolación final. Además de la elevación, se consideraron variables auxiliares adicionales como el porcentaje de mar y el porcentaje de lagos. (Aalto et al., 2013)

En una región de los Alpes, con 440 estaciones, se han comparado los métodos de regresión lineal y KED para la interpolación de precipitación diaria. De entre ambos métodos el KED mostró menores errores en la interpolación. La elevación mostró ser un factor importante en la mayor precisión de la interpolación al compararlo por ejemplo con el Kriging ordinario. Si bien se probaron otras variables auxiliares, no resultaron ser significativas. Se concluyó que el uso de solo una variable auxiliar con Kriging sería más efectivo que el desarrollo de un modelo de regresión con múltiples variables auxiliares. (Masson y Frei, 2014)

La elevación en general es la variable auxiliar preferida para realizar interpolaciones de alta montaña, sin embargo, el difícil relieve de estas zonas hace que la distancia más corta entre dos estaciones no siempre refleje adecuadamente el grado de conexión entre ellas. En ese sentido se han aplicado distancias no euclídeas para realizar la interpolación de temperatura en Suiza. Esta distancia se puede considerar como el recorrido más corto considerando los accidentes topográficos como montañas o valles. (Frei, 2014)

CAPITULO II: MARCO TEÓRICO Y CONCEPTUAL

2.1. MARCO TEÓRICO

2.1.1 Métodos de interpolación determinísticos

El problema general que intentan resolver los distintos métodos de interpolación puede expresarse como: dadas n ubicaciones o puntos $s_1, s_2, s_3, \dots, s_n$ con observaciones de una variable aleatoria, se desea estimar el valor de esta variable en una ubicación o punto sin medición s_0 .

Los métodos determinísticos hacen uso de ecuaciones matemáticas que predicen un único valor en cada punto de interpolación, sin consideraciones estadísticas o estimaciones del error posible. Muchos de ellos se basan en el principio de la similitud, que supone que los puntos más cercanos espacialmente tienen más relación entre sí.

En las siguientes secciones se hará referencia en general a los datos de las estaciones como ubicaciones conocidas o ubicaciones muestreadas, y a los puntos a estimar como ubicaciones desconocidas.

2.1.1.1 *Polígonos de Thiessen*

Este método propuesto por Thiessen (1911) es probablemente el más simple y uno de los más antiguos métodos de interpolación. La metodología asigna al sitio a estimar el valor del dato más cercano. Es decir, se asigna a cada ubicación desconocida el valor de la ubicación conocida más cercana.

Como resultado se genera una superficie dividida en polígonos, cada uno correspondiente a una ubicación muestreada. Geométricamente cada polígono se genera uniendo los puntos medios de las líneas que conectan las ubicaciones muestreadas. En la Figura N° 1(a) se muestra la configuración característica de esta metodología.

Es conocido que el método de Thiessen no reproduce adecuadamente los patrones de precipitación para zonas montañosas con complicada topografía (Bourgin, 2010). Varios autores han notado una baja performance de este método con respecto a otros (Goovaerts, 2000; Mair y Fares, 2011) en este tipo de zonas;

sin embargo, considerando su aún amplio uso se tendrá en cuenta con fines comparativos.

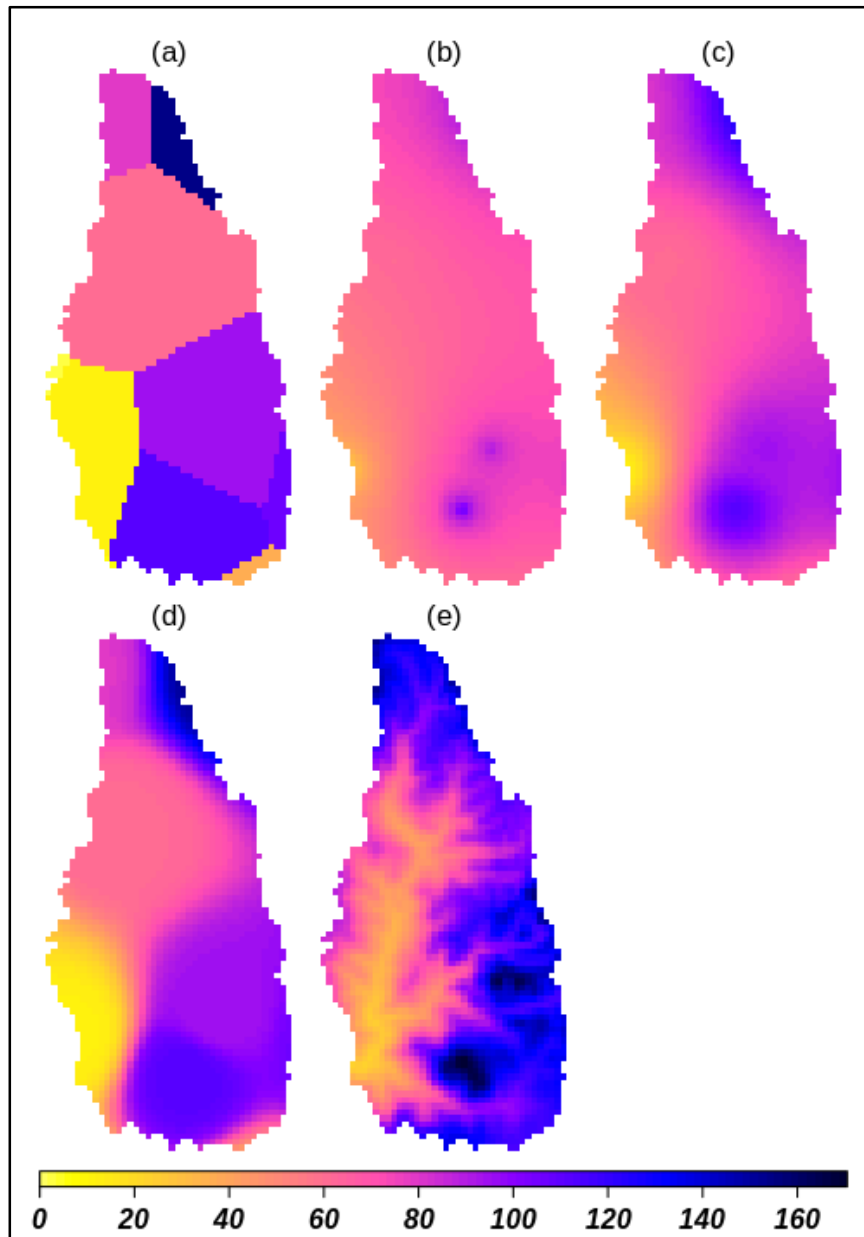


Figura N° 1 Interpolación mediante distintos métodos. Polígonos de Thiessen (a), IDW $p = 1$ (b), IDW $p = 2$ (c), IDW $p = 5$ (d) e IDW con corrección por altitud (e)

Fuente: Elaboración propia

2.1.1.2 IDW

El método IDW (Interpolación ponderada por distancia inversa, Inverse Distance Weighted interpolation) usa una media ponderada de datos cercanos como

estimación para la ubicación desconocida. En esta ponderación cada ubicación muestreada considerada tiene un peso inversamente proporcional a la distancia a la ubicación desconocida elevada a una potencia.

$$\hat{P} = \frac{\sum P_i/d_i^p}{\sum 1/d_i^p} \quad (2-1)$$

Donde:

\hat{P} : Valor a estimar en la ubicación desconocida.

P_i : Valor en la ubicación conocida i .

d_i : Distancia desde la ubicación desconocida a la ubicación conocida i .

p : Potencia de la distancia.

En general una mayor potencia produce una mayor influencia de las estaciones cercanas, pareciéndose más a la metodología de los Polígonos de Thiessen. Mientras que una potencia baja tiende a considerar ubicaciones cercanas y lejanas por igual. Es posible calibrar este parámetro para obtener mejores resultados (Chen y Liu, 2012). Una configuración típica de esta metodología se muestra en la Figura N° 1.

2.1.1.3 IDW con corrección de gradiente altitudinal

La relación existente entre elevación y precipitación es especialmente fuerte para pasos de tiempo mensuales y anuales (Bárdossy y Pegram, 2013). Esta relación puede ser simplificada mediante el uso de gradientes altitudinales que establecen aumentos de precipitación dado un aumento de la elevación entre ubicaciones. Valery et al. (2010) establece un factor de corrección multiplicativo que considera el gradiente altitudinal, así se establece:

$$\hat{P} = \frac{1}{\sum w_i} \sum w_i P_i \times e^{\theta(z-z_i)} \quad (2-2)$$

Donde:

\hat{P} : Valor a estimar en la ubicación desconocida.

P_i : Valor en la ubicación conocida i .

θ : parámetro relacionado al gradiente altitudinal.

w_i : Peso de la ubicación conocida i . Se estima como en el método IDW

$$(w_i = 1/d_i^p)$$

- z : Elevación en la ubicación a estimar
 z_i : Elevación en la ubicación conocida i

Asimismo, el parámetro θ se puede calibrar a partir de los datos disponibles. Rau et al. (2019) usó un valor de $\theta = 4 \times 10^{-4} m^{-1}$ para la vertiente del Pacífico, que representa aproximadamente un incremento de 20 % en la precipitación local para 500 m de diferencia en elevación. Esta metodología es también usada en la región Norte Chico de Chile por Bourgin (2010).

En la Figura N° 2 se muestra un esquema de la relación espacial que este método plantea. Asimismo, en la Figura N° 1(e) se muestra una interpolación típica usando esta metodología.

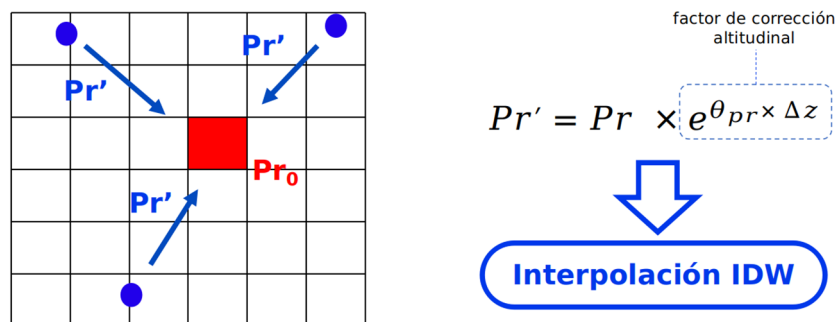


Figura N° 2. Esquema de interpolación mediante metodología propuesta por Valery (2010)

Fuente: Elaboración propia

2.1.2 Métodos de interpolación geoestadísticos

Este tipo de técnicas utilizan las relaciones estadísticas entre las ubicaciones con mediciones. Estos métodos cuantifican la correlación espacial y predicen no solo el valor estimado en las ubicaciones desconocidas sino también evalúan su incertidumbre.

Si bien es cierto existen otras metodologías, las de más amplio uso y que se consideran estándar son el método Kriging y sus derivados. En las secciones posteriores se incluye el fundamento teórico de estas metodologías.

2.1.2.1 Estimación de la correlación espacial

En geoestadística la correlación especial es modelada mediante el variograma. El variograma relaciona las semivarianzas como una función de las distancias. Por

cuestiones prácticas se usa el nombre de variograma como sinónimo del de semivariograma (Bachmaier y Backes, 2011). Por lo tanto, se seguirá esta convención en el presente texto.

Para estimar la correlación espacial a partir de datos observados es necesario hacer suposiciones de estacionariedad. Con estacionariedad nos referimos a que ciertas propiedades de una variable regionalizada sean invariantes respecto a cualquier traslación espacial. Para usar ello, se debe asumir algunas premisas o suposiciones.

Se habla de estacionariedad de segundo orden o débil cuando el valor esperado de la variable regionalizada es constante para cualquier ubicación y además la covarianza para dos ubicaciones solo depende de la distancia entre estas.

Otra forma comúnmente usada es la estacionariedad intrínseca que define dos enunciados:

1. El valor esperado para la variable regionalizada es constante sobre todo el dominio.
2. La varianza de la diferencia de la variable regionalizada para dos diferentes ubicaciones depende solo de la separación entre estas.

Considerando la variable regionalizada Z dependiente de la ubicación s , estos dos enunciados se pueden expresar matemáticamente como:

$$E(Z(s)) = \mu \quad (2-3)$$

$$\gamma(\mathbf{h}) = \frac{1}{2}E(Z(s) - Z(s + \mathbf{h}))^2 \quad (2-4)$$

Con esta última ecuación se puede considerar la semivarianza γ como función únicamente de la distancia h , mas no de la ubicación s .

En ese sentido para estimar la semivarianza experimentalmente, se pueden tomar múltiples pares $\{Z(s_i), Z(s_j)\}$ los cuales tengan una separación idéntica o muy cercana $h = s_i - s_j$ y estimar la correlación a partir de estos. Si además se asume isotropía, es decir que la semivarianza es independiente de la dirección, se puede reemplazar el vector \mathbf{h} por su longitud $h = \|\mathbf{h}\|$.

De esta manera se puede definir el variograma empírico como:

$$\gamma'(h) = \frac{1}{2N_h} \sum_{|s_i - s_j|=h} [Z(s_i) - Z(s_j)]^2 \quad (2-5)$$

En el caso de puntos de observación irregularmente espaciados, se puede adoptar la consideración de agrupar valores bajo un mismo h considerando una vecindad cercana. Dos puntos s_i y s_j se considerarán separados a una distancia h si su separación está en el rango $h - \epsilon \leq |s_i - s_j| \leq h + \epsilon$ para un ϵ establecido. Esta condición significa:

$$|s_i - s_j| \leq h \pm \epsilon \quad (2-6)$$

En la Figura N° 3 se observa un variograma empírico estimado usando la aproximación descrita, también se muestra la nube de todas las semivarianzas a partir de las cuales se estima.

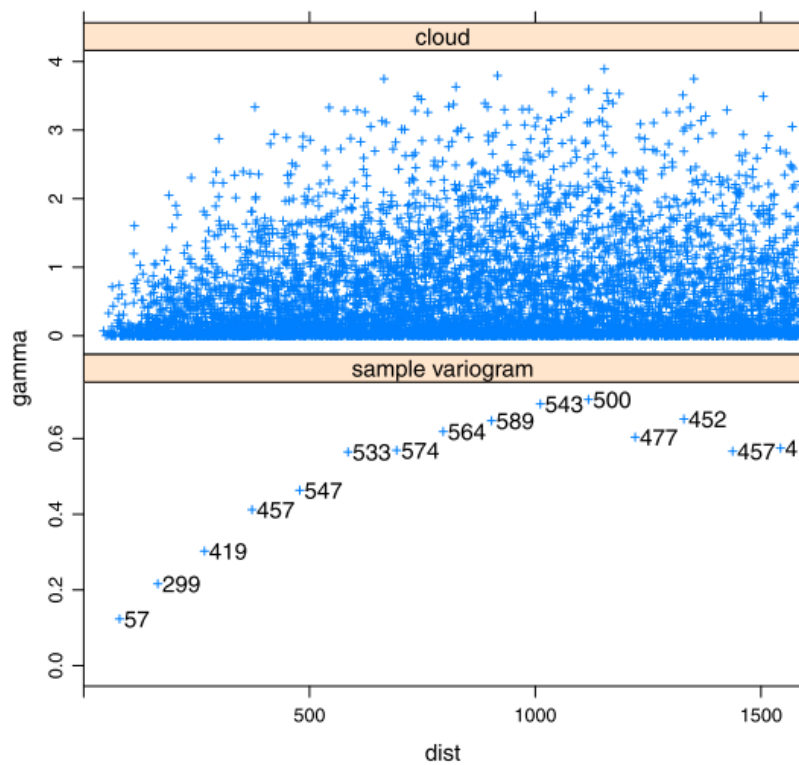


Figura N° 3 Nube de semivarianzas (arriba) y variograma experimental o empírico (abajo).

Fuente: Bivand et al. (2013)

Variogramas teóricos

Una vez conocido el variograma experimental, se hace necesario modelar esta relación empírica para estimar la semivarianza en todo el rango continuo de distancias. Para ello se hacen uso de modelos parametrizados de variogramas o variogramas teóricos, los cuales pueden ser ajustados mediante sus parámetros para representar de manera más precisa al variograma experimental.

Existen tres parámetros comunes a ciertos modelos de variograma que se muestran en la Figura N° 4. El primero de ellos es el efecto pepita, denotado por C_0 , representa una discontinuidad puntual del variograma en el origen, que puede deberse a errores propios de la medición o a una variación espacial a una escala más pequeña a la distancia entre cualquiera dos puntos de la muestra (Bickel et al., 2007). Por otro lado, la meseta, denotada por $C_0 + C_1$, representa la cota superior del variograma, es decir el límite del semivarianza cuando la distancia h tiende al infinito. Finalmente el rango a representa la distancia a partir de la que dos observaciones se consideran independientes, se puede interpretar como la zona de influencia (Giraldo, 2002).

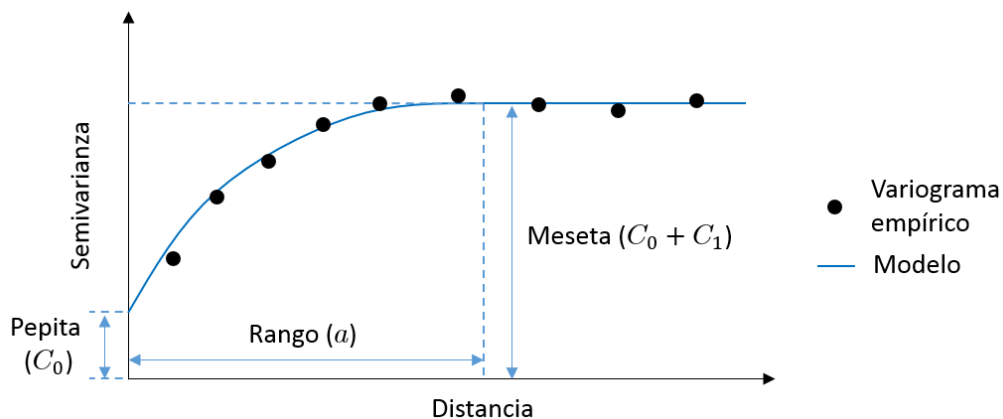


Figura N° 4 Componentes típicos de un variograma: pepita, meseta y rango.

Fuente: Elaboración propia

Entre los variogramas teóricos más usuales tenemos los mostrados en la Figura N° 5 y se detallan a continuación.

Variograma pepita puro. Representa la carencia de correlación espacial entre las observaciones de la variable aleatoria (Giraldo, 2002).

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ C_0 & h > 0 \end{cases}, \text{ donde } C_0 > 0 \quad (2-7)$$

Variograma esférico. El crecimiento rápido en el origen va decreciendo hasta volverse nulo a partir del rango a .

$$\gamma(h) = \begin{cases} C_0 + C_1 \left[\frac{3}{2} \left(\frac{h}{a} \right) - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right] & h \leq a \\ C_0 + C_1 & h > a \end{cases} \quad (2-8)$$

Variograma exponencial. A pesar de que la función $\gamma(h)$ tiende a $C_0 + C_1$ cuando la distancia h tiende al infinito, no existe una distancia en particular a partir de la cual las variables aleatorias sean independientes. Sin embargo, en este caso el rango se estima como $3a$ y representa el valor en el cual se alcanza el 95% de la meseta (Bivand et al., 2013).

$$\gamma(h) = C_0 + C_1 \left[1 - \exp \left(-\frac{h}{a} \right) \right] \quad (2-9)$$

Variograma gaussiano. De forma similar al modelo exponencial, la dependencia espacial se desvanece solo a una distancia que tiende a infinito (Giraldo, 2002). De forma práctica se considera el rango como $\sqrt{3}a$ (Bivand et al., 2013).

$$\gamma(h) = C_0 + C_1 \left[1 - \exp \left(-\frac{h^2}{a^2} \right) \right] \quad (2-10)$$

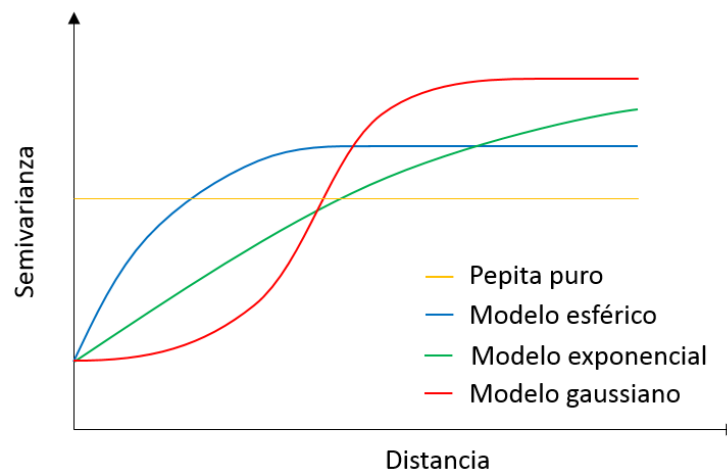


Figura N° 5 Ejemplos de variogramas teóricos

Fuente: Elaboración propia

2.1.2.2 Kriging ordinario

Los conceptos de la iniciales de la geoestadística y el método de Kriging, hoy de amplio uso, fueron definidos formalmente por Matheron (1963) a partir de los trabajos prácticos de Krige (1951), en honor al cual el método lleva su nombre.

La idea básica del kriging es predecir el valor en una ubicación desconocida, calculando una media ponderada de los valores en ubicaciones cercanas conocidas, donde los pesos óptimos de la ponderación se calculan en base a la distancia a la ubicación de interpolación y el grado de autocorrelación espacial asociada a esta (Isaaks, 1989).

El método de Kriging ordinario considera el siguiente modelo (Singh y Verma, 2019):

$$Z(\mathbf{s}) = \mu + \varepsilon(\mathbf{s}) \quad (2-11)$$

Donde $Z(\mathbf{s})$ es el valor en la ubicación \mathbf{s} , μ es la media y $\varepsilon(\mathbf{s})$ representa el error en la ubicación \mathbf{s} .

De esta forma, el modelo está basado en una media μ constante para la data (no hay tendencia) y errores aleatorios $\varepsilon(\mathbf{s})$ con dependencia espacial. Se asume que el proceso aleatorio $\varepsilon(\mathbf{s})$ es estacionario.

En términos prácticos, la predicción se forma como una combinación lineal ponderada (promedio ponderado) de la data:

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^N w_i Z(\mathbf{s}_i) \quad (2-12)$$

$$\sum_{i=1}^N w_i = 1 \quad (2-13)$$

Donde:

$\hat{Z}(\mathbf{s}_0)$: representa la estimación en la ubicación \mathbf{s}_0

$Z(\mathbf{s}_i)$: representa el valor conocido en la ubicación \mathbf{s}_i

w_i : son los pesos de la combinación lineal ponderada.

Para calcular los pesos w_i se tienen en cuenta los objetivos de sesgo nulo (considera la suposición de estacionariedad) y mínima varianza de los errores de

la estimación, de esta forma se pueden calcular matricialmente como un sistema de ecuaciones. Esta aproximación se puede encontrar desarrollada por Webster y Oliver (2007).

$$\begin{bmatrix} \gamma(s_1, s_1) & \gamma(s_1, s_2) & \cdots & \gamma(s_1, s_N) & 1 \\ \gamma(s_2, s_1) & \gamma(s_2, s_2) & \cdots & \gamma(s_2, s_N) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(s_N, s_1) & \gamma(s_N, s_2) & \cdots & \gamma(s_N, s_N) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \\ \Psi \end{bmatrix} = \begin{bmatrix} \gamma(s_1, s_0) \\ \gamma(s_2, s_0) \\ \vdots \\ \gamma(s_N, s_0) \\ 1 \end{bmatrix}$$

$$\mathbf{A} \mathbf{w} = \mathbf{b} \quad (2-14)$$

Donde $\gamma(s_i, s_j)$ representa la semivarianza entre las ubicaciones s_i y s_j y se estima en función de la distancia h entre ambas ubicaciones considerando un variograma modelado a partir de un variograma teórico como los descritos en la sección 2.1.2.1. Así mismo es importante notar que el elemento Ψ representa el multiplicador de Lagrange que es usado para formular la minimización de la varianza.

Así en base al sistema matricial se puede resolver y calcular los pesos.

$$\mathbf{w} = \mathbf{A}^T \mathbf{b} \quad (2-15)$$

Asimismo, se puede calcular la varianza de la estimación o varianza de Kriging mediante:

$$\sigma^2_{OK} = \mathbf{b}^T \mathbf{w} \quad (2-16)$$

Esta varianza que es una estimación de $Var(\hat{Z}(s_0) - Z(s_0))$ depende de la ubicación de las observaciones y el modelo de variograma, mas no en los datos mismos, lo cual puede conllevar a estimaciones no satisfactorias como es discutido por Heuvelink y Pebesma (2002).

Además, en forma similar a como se han deducido las soluciones y sistema de ecuaciones en base a las semivarianzas se puede hacer ello tomando las covarianzas, esta aproximación es desarrollada por Giraldo (2002) y otros autores.

En el presente texto se usarán las siglas OK para denotar al método de Kriging ordinario por sus siglas en inglés (Ordinary Kriging).

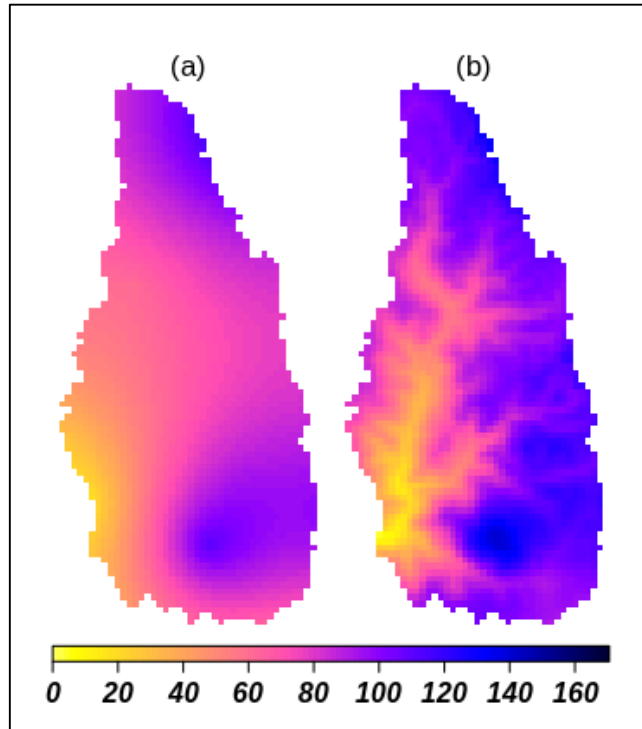


Figura N° 6 Interpolación usando el método de Kriging ordinario (a) y Kriging con deriva externa (b)
Fuente: Elaboración propia

2.1.2.3 Kriging con deriva externa

Hasta ahora se ha asumido para el método de Kriging que se cumple la estacionariedad de la variable regionalizada. Pero es natural que en ciertos casos no se cumpla con esta condición y más bien la variable responde a una tendencia. Por ejemplo, en el caso de la precipitación se puede encontrar una tendencia que responde a la elevación. Para tratar este tipo de tendencia se puede descomponer la variable $Z(x)$ en una componente determinística y una estocástica estacionaria de media cero (Giraldo, 2002). Así tenemos:

$$Z(s) = \mu(s) + \varepsilon(s) \quad (2-17)$$

Si bien es cierto se pueden considerar varias variables auxiliares que permitan modelar la tendencia, en el presente texto trabajaremos el caso de una sola variable.

De forma similar el método de Kriging ordinario se puede considerar a la predicción como una combinación lineal ponderada.

$$\hat{Z}(s_0) = \sum_{i=1}^N w_i Z(s_i) \quad (2-18)$$

La estimación de los pesos es bastante similar a la forma empleada en el Kriging ordinario, considerando ahora la información de la variable auxiliar f_1 , que se ve reflejada en el incremento del número de ecuaciones del sistema.

$$\begin{bmatrix} \gamma(s_1, s_1) & \gamma(s_1, s_2) & \cdots & \gamma(s_1, s_N) & 1 & f_1(s_1) \\ \gamma(s_2, s_1) & \gamma(s_2, s_2) & \cdots & \gamma(s_2, s_N) & 1 & f_1(s_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \gamma(s_N, s_1) & \gamma(s_N, s_2) & \cdots & \gamma(s_N, s_N) & 1 & f_1(s_N) \\ 1 & 1 & \cdots & 1 & 0 & 0 \\ f_1(s_1) & f_1(s_2) & \cdots & f_1(s_N) & 0 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \\ \Psi_0 \\ \Psi_1 \end{bmatrix} = \begin{bmatrix} \gamma(s_1, s_0) \\ \gamma(s_2, s_0) \\ \vdots \\ \gamma(s_N, s_0) \\ 1 \\ f_1(s_0) \end{bmatrix}$$

$$\mathbf{A} \mathbf{w} = \mathbf{b} \quad (2-19)$$

Como en el Kriging ordinario la matriz es invertida y se pueden obtener los pesos y los multiplicadores de Lagrange.

$$\mathbf{w} = \mathbf{A}^T \mathbf{b} \quad (2-20)$$

Y además se puede calcular la varianza de Kriging:

$$\sigma^2_{KED} = \mathbf{b}^T \mathbf{w} \quad (2-21)$$

Adicionalmente a la forma matricial previamente descrita, es usual solucionar el mismo problema de hallar los pesos óptimos analizando covarianzas en lugar de semivarianzas, para lo cual, con lo cual se forma un sistema matricial de similares características. Esta aproximación es normalmente usada con fines de eficiencia computacional (Hengl et al., 2003)

Es importante notar que los métodos de Kriging confían en el variograma. Estadísticamente la validez del variograma depende de los valores medidos. Mientras mayor sea la cantidad de datos, más válido será el variograma. Por otro lado, la validez de un análisis local será menos precisa si la cantidad de datos disponible es pequeña (Singh y Verma, 2019).

Otro método usado para realizar estimaciones considerando una variable auxiliar es el Co-Kriging. Sin embargo el método de Kriging con deriva externa normalmente tiene una mejor performance que el Co-Kriging (Bárdossy y Pegram, 2013; Majani, 2007). Además, el método de Kriging con deriva externa es

ampliamente usado y suele proveer mejores resultados en comparación a otros métodos en zonas donde existe dependencia altitudinal de la precipitación (Aalto et al., 2013; Lloyd, 2005; Masson y Frei, 2014; Pardo-Igúzquiza, 1998).

En la Figura N° 6 se muestra una comparación entre el método de Kriging ordinario y el método de Kriging con deriva externa, se usa la misma escala de colores que la Figura N° 1 para que los gráficos sean comparables con los de los métodos determinísticos.

El método de Kriging con deriva externa es denotado como KED por sus siglas en inglés (Kriging with External Drift), de ahora en adelante esta será la forma utilizada.

2.1.3 Validación cruzada

En algunas áreas es común separar los datos recopilados en dos partes: un grupo de datos para la calibración y otro para la evaluación o validación. Esto, por ejemplo, es realizado en modelos hidrológicos. Es así como, en situaciones donde se tiene una gran cantidad de información y el tamaño de la data de calibración es suficiente para que el modelo sea casi invariante ante la incorporación de nuevos datos, esta estrategia es muy conveniente. En los métodos de interpolación tratados, no se dan estas condiciones, ya que, si se considera la variabilidad espacial, existen pocos datos útiles para un área determinada, es así que tratar con un gran conjunto de datos para la validación puede resultar en pérdida de información valiosa para la predicción.

En por ello que se suele preferir usar otra estrategia. Se emplea un tipo especial de validación cruzada conocida como la técnica Jack-knife, donde cada ubicación observada se considera como una ubicación sin medición y con el resto de las observaciones se procede a realizar la estimación en esa ubicación.

De esta forma se dispone de dos registros, uno de observaciones y otro de simulaciones, ambos son comparados mediante algún índice de desempeño, que indica qué tan precisa ha sido la predicción. Esta aproximación permite, de esta forma, realizar la comparación entre diferentes métodos.

2.1.4 Índices de desempeño

Estos índices se usan en conjunto con la validación cruzada y permiten representar la bondad de ajuste de las predicciones o simulaciones con respecto a los valores observados. Se detallarán algunos de los más usados para este fin. Asimismo, es posible usarlos para evaluar el desempeño de modelos hidrológicos, que se detallarán en 2.1.6.

2.1.4.1 Eficiencia de Nash-Sutcliffe

Indica la relación entre la varianza residual o ruido comparada con la varianza de los datos medidos (Nash y Sutcliffe, 1970). Representa la relación de la variabilidad de los datos simulados con respecto a los observados.

$$Nash = 1 - \frac{\sum_{t=1}^n (Sim_t - Obs_t)^2}{\sum_{t=1}^n (Sim_t - \overline{Obs})^2} \quad (2-22)$$

Donde Sim_t y Obs_t representan los valores simulados y observados en el tiempo t respectivamente y \overline{Obs} representa la media de los valores observados. Este índice puede variar de $-\infty$ a 1, representando 1 su óptimo valor.

El índice de Nash-Sutcliffe se considera como una generalización del coeficiente de determinación R^2 . La ecuación que describe al R^2 es similar a la del Nash. Sin embargo, el R^2 es usado generalmente en el ámbito de regresiones lineales ($Obs = a \times Sim + b$), y en este caso el Nash solo considera la evaluación de un caso particular donde $a = 1$ y $b = 0$ (Legates y McCabe Jr., 1999). Es así que se pueden obtener valores altos de R^2 incluso si las magnitudes difieren considerablemente. Además, estos índices tienen interpretaciones muy distintas (Ritter y Muñoz-Carpena, 2013). Sin embargo, si se deja de lado la componente de regresión lineal, el Nash y el R^2 son equivalentes. Para evitar este tipo de ambigüedades es preferible trabajar con el índice de Nash - Sutcliffe.

2.1.4.2 Raíz del error medio cuadrático

Expresa la desviación estándar del error de predicción. Representa la magnitud de la desviación de los valores simulados respecto a los observados (Cabrera, 2009).

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Sim_t - Obs_t)^2} \quad (2-23)$$

Puede variar de 0 a ∞ , siendo su valor óptimo es 0, la ventaja de este índice es que representa el error en las mismas unidades de los datos.

2.1.4.3 Coeficiente de correlación de Pearson

Representa el ajuste lineal entre las series de datos. Asimismo, muestra la covariabilidad entre valores simulados y observados sin penalización por sesgo (Wang et al., 2011).

$$CC = r = \frac{\sum_{t=1}^n (Sim_t - \overline{Sim})(Obs_t - \overline{Obs})}{\sqrt{\sum_{t=1}^n (Sim_t - \overline{Sim})^2} \sqrt{\sum_{t=1}^n (Obs_t - \overline{Obs})^2}} \quad (2-24)$$

Este índice varía entre -1 y 1, representando 1 la mejor performance.

2.1.4.4 Error de sesgo

Representa la diferencia relativa entre el volumen total simulado y observado. Mide la tendencia promedio de los valores simulada a ser mayores o menores que los datos observados (Gupta et al., 1999).

$$Bias = \frac{\sum_{t=1}^n (Sim_t - Obs_t)}{\sum_{t=1}^n Obs_t} \quad (2-25)$$

Suele representarse como porcentaje y su valor óptimo es 0.

2.1.4.5 Eficiencia de Kling-Gupta

Este criterio propuesto por Gupta et al. (2009) es un indicador que facilita el análisis global basado en diferentes componentes (correlación, sesgo y variabilidad) para propósitos de modelamiento hidrológico.

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (2-26)$$

Donde los subcriterios son los siguientes: r es el coeficiente de correlación, $\alpha = \sigma_{sim}/\sigma_{obs}$ es la relación entre las desviaciones estándar, y $\beta = \mu_{sim}/\mu_{obs}$ es la relación entre las medias. La mejor performance del índice es representada por el valor de 1.

Asimismo, una discusión sobre las ventajas con respecto a la Eficiencia de Nash-Sutcliffe en modelamiento hidrológico puede encontrarse en Gupta et al. (2009). Además, Kling et al. (2012) proponen una versión revisada de este índice para asegurar que los ratios usados como subcriterios no estén correlacionados.

2.1.5 Algunas aproximaciones de distancias no euclídeas

La distancia puede ser definida como una medida cuantitativa de la lejanía o proximidad entre dos puntos. Es así como, en los métodos de interpolación previamente descritos, la distancia representa esto. Sin embargo, en este caso no se consideran otros elementos de la topografía, como accidentes geográficos o diferencias en la elevación que pueden modificar esta proximidad.

La distancia euclídea entre dos puntos equivale a la longitud del segmento de línea que los une. Es así que en muchas disciplinas científicas las distancias euclídeas pueden describir inadecuadamente relaciones de proximidad entre datos espaciales (Curriero, 2006).

2.1.5.1 Distancia de perfil topográfico

Esta distancia se considera como el recorrido a lo largo del perfil topográfico de la línea recta entre dos ubicaciones (Ver Figura N° 7). Este enfoque es usado por Gottardi et al. (2012), que además propone la penalización de los incrementos verticales multiplicándolos por un coeficiente. Las pendientes son tratadas de manera diferente de acuerdo a su orientación con el objetivo de penalizar las crestas montañosas más que los valles. La ventaja de esta distancia sobre la distancia euclídea es que toma en consideración el relieve.

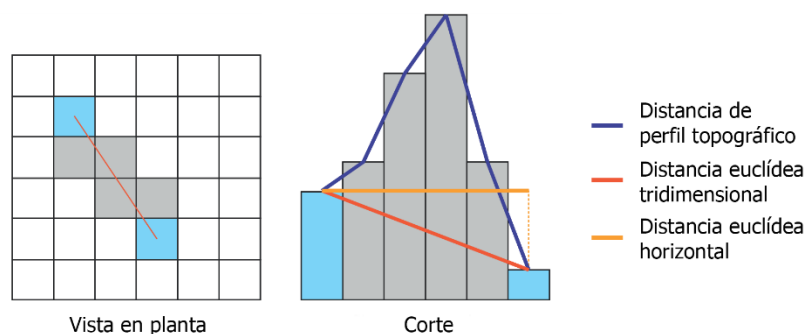


Figura N° 7 Distancia de perfil topográfico en comparación a distancias euclídeas.

Fuente: Adaptado de (Gottardi et al., 2012)

2.1.5.2 Distancia generalizada más corta

Frei (2014) propone una distancia generalizada para realizar la interpolación de residuales de temperatura, con el fin de considerar la influencia de obstáculos topográficos, así como de la diferencia entre valles y picos. La distancia generalizada $D_\lambda(s_1, s_2)$ entre las ubicaciones s_1 y s_2 es definida como:

$$D_\lambda(s_1, s_2) = \min_{\Gamma(s_1, s_2)} \left(\int_{\Gamma} \sqrt{\left(\frac{\partial \Gamma_x}{\partial l}\right)^2 + \left(\frac{\partial \Gamma_y}{\partial l}\right)^2 + \left(\lambda \frac{\partial \Gamma_z}{\partial l}\right)^2} dl \right) \quad (2-27)$$

Esta distancia es definida como la mínima longitud de todas las posibles conexiones entre dos ubicaciones, donde las conexiones solo están permitidas sobre la superficie topográfica. Las trayectorias Γ se consideran como funciones vectoriales tridimensionales $\Gamma = (\Gamma_x, \Gamma_y, \Gamma_z)^T$ parametrizadas por l . Además, se considera un coeficiente λ el peso relativo de los incrementos verticales de distancia con respecto a los horizontales.

Un mayor λ penalizará más fuertemente los recorridos verticales y como resultado, las obstrucciones montañosas entre dos ubicaciones generarán una mayor distancia. Un caso particular ocurre cuando $\lambda = 0$, donde la distancia generalizada se reduce a la distancia euclídea. Una comparación de este coeficiente se observa en la Figura N° 8 donde un menor λ refleja una menor influencia topográfica.

Asimismo, se permiten trayectorias en la atmósfera libre, siempre que esto sea posible. Esto genera conexiones cortas entre laderas de valles, que serían altamente penalizadas si solo se considerase trayectorias en la superficie topográfica, las cuales tendrían que seguir el fondo del valle. Esto permitiría un mejor intercambio de información entre ubicaciones que pueden tener un intercambio de masas de aire. Esto se ve reflejado en la Figura N° 8, donde la estación ubicada en la montaña se ve conectada con otras laderas o picos, mientras que la estación del fondo del valle no presenta estas conexiones debido a su confinamiento.

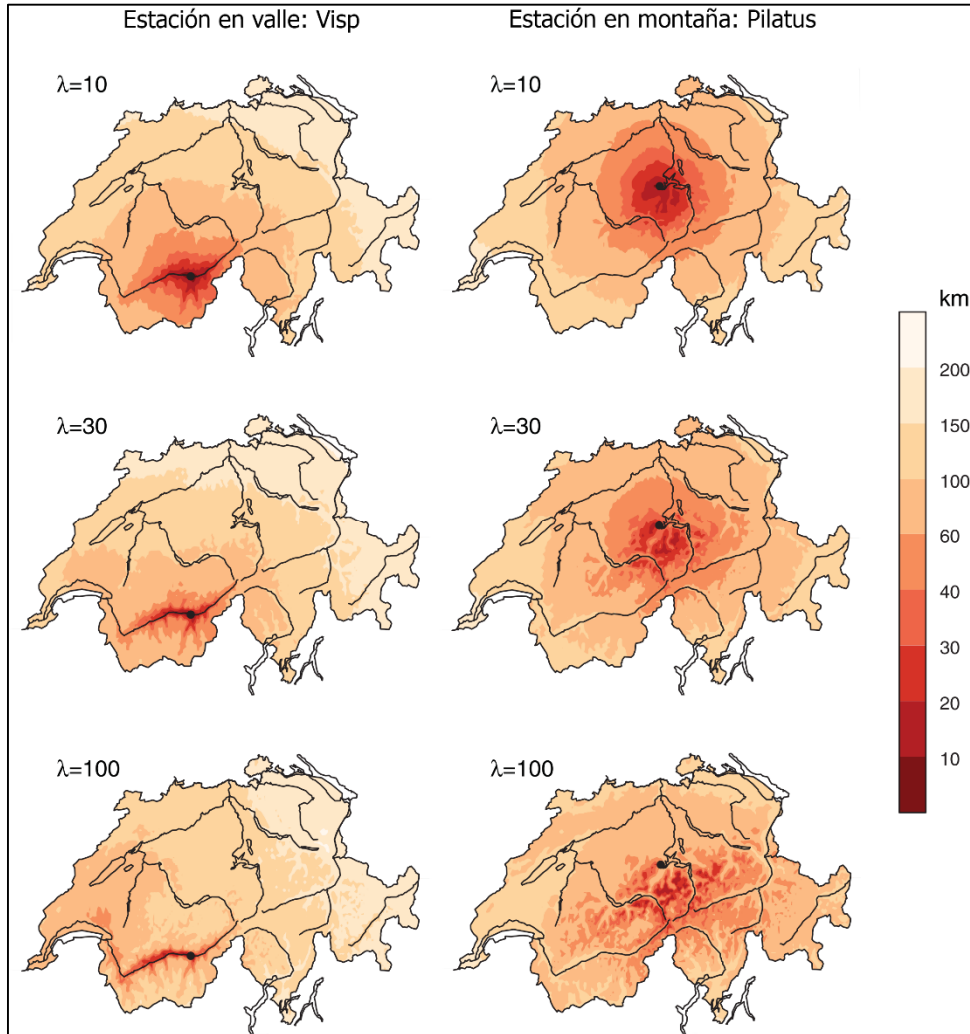


Figura N° 8 Distancia generalizada más corta D_λ para diferentes coeficientes λ desde una estación en valle (izquierda) y una en montaña (derecha)

Fuente: Frei (2014)

2.1.6 Modelos hidrológicos

Un modelo es una representación simplificada de un sistema del mundo real y los diferentes procesos que gobiernan su operación (Hingray et al., 2014). Estos pueden ser representaciones físicas, sin embargo, es más frecuente que sean modelos matemáticos. En un sentido amplio se puede considerar a los modelos hidrológicos como aquéllos que representan algún fenómeno hidrológico. En nuestro caso en particular los consideraremos como modelos matemáticos precipitación-escorrentía de una cuenca.

Es así que los modelos hidrológicos tratan de representar el ciclo hidrológico dentro de una cuenca. Para ello se debería contar con las mediciones de cada

uno de sus subprocesos, sin embargo, esto en la práctica es muy difícil, por lo que existe un grado de incertidumbre en la estimación de las variables hidrológicas.

Los modelos tienen parámetros, los cuales son medidas numéricas de una propiedad o característica que es constante bajo ciertas condiciones (Weather et al., 2008). También poseen entradas y salidas, la principal entrada será la precipitación y la principal salida la esorrentía.

Se habla de un modelo agregado cuando sus parámetros, entradas y salidas están espacialmente promediados y se representan por un único valor para toda la cuenca. El modelo será distribuido si sus parámetros, entradas y salidas varían espacialmente. Asimismo, un modelo semidistribuido considerará una representación agregada para subcuencas individuales.

2.1.6.1 Calibración y validación

La data disponible se divide en dos grupos independientes: uno para la etapa de calibración y otro para la validación. Es recomendable que ambos grupos tengan datos representativos de las diferentes condiciones climáticas presentes en la data.

Calibración

En el proceso de calibración, se estiman los parámetros que no pueden ser evaluados directamente con datos medidos. En principio, existen 3 métodos de calibración aplicables:

- a. Prueba y error, un ajuste de parámetros manual
- b. Calibración automática, una optimización numérica de los parámetros.
- c. Una combinación de las 2 primeras.

Si bien es cierto se puede realizar una evaluación visual de la bondad de ajuste entre caudales simulados y observados en el método de prueba y error, lo usual es considerar un índice de desempeño para realizar una evaluación menos subjetiva. Es así que en la calibración automática se busca optimizar este índice de desempeño por medio de la variación de parámetros. Algunos ejemplos de posibles índices de desempeño de detallaron en la sección 2.1.4.

Validación

Con el objetivo de establecer si un modelo calibrado puede ser considerado para un uso subsecuente, este debe ser probado con datos diferentes a los usados para la calibración. La forma de evaluar su aceptabilidad es por medio del mismo índice de desempeño.

2.1.6.2 Generalidades del modelo GR2M

El modelo GR2M en su última versión fue desarrollado por Mouelhi et al. (2006). El modelo GR2M es un modelo hidrológico agregado a paso mensual de 2 parámetros. Este modelo es ampliamente usado debido a su alta eficiencia en uso de recursos y su aproximación semiempírica y ha demostrado tener una buena performance en comparación con otros modelos a paso mensual similares (Rau et al., 2019)

El modelo tiene 2 parámetros que se pueden optimizar:

X_1 : capacidad del reservorio de producción (mm)

X_2 : coeficiente de intercambio subterráneo (adimensional)

El modelo considera como datos de entrada la precipitación P y evapotranspiración potencial E , que puede ser estimada considerando la temperatura. Estas variables se representan en la Figura N° 9. Asimismo, el caudal simulado Q se considera como variable de salida y F representa el flujo de intercambio no atmosférico, que puede ser de ganancia o pérdida.

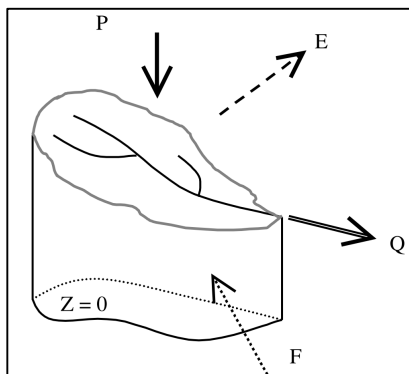


Figura N° 9 Esquema de la cuenca y variables del modelo GR2M

Fuente: Mouelhi (2006)

2.1.6.3 Descripción del modelo GR2M

Un esquema de la estructura del modelo se muestra en la Figura N° 10. P_k representa la precipitación mensual en el mes k y ET la evapotranspiración en el mismo mes.

Las ecuaciones que regulan el funcionamiento del modelo se describen a continuación y se presentan de forma similar a (Perrin et al., 2007).

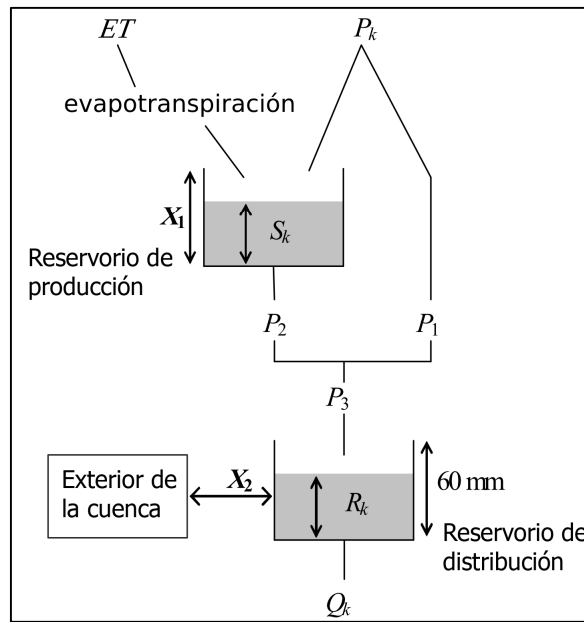


Figura N° 10 Esquema de modelo GR2M
Fuente: Adaptado de (Perrin et al., 2007)

Producción

La función de producción del modelo trabaja sobre un reservorio que representa la humedad del suelo. Una parte P_s de la precipitación P_k será añadida al contenido inicial del reservorio S_k .

$$P_s = \frac{X_1 \left(1 - \left(\frac{S_k}{X_1} \right)^2 \right) \tanh \left(\frac{P_k}{X_1} \right)}{1 + \frac{S_k}{X_1} \tanh \left(\frac{P_k}{X_1} \right)} \quad (2-28)$$

El parámetro X_1 , la capacidad del reservorio, es positivo y se expresa en mm. La precipitación en exceso, P_1 , está dada por:

$$P_1 = P - P_s \quad (2-29)$$

Y el contenido del reservorio se actualiza como:

$$S' = S_k - P_s \quad (2-30)$$

Debido a la evapotranspiración, una cantidad E_s es tomada del reservorio:

$$E_s = \frac{S' \left(2 - \frac{S'}{X_1} \right) \tanh \left(\frac{E}{X_1} \right)}{1 + \left(1 - \frac{S'}{X_1} \right) \tanh \left(\frac{E}{X_1} \right)} \quad (2-31)$$

Luego, el nivel S' se convierte en S'' :

$$S'' = S' - E_s \quad (2-32)$$

Percolación

El reservorio de producción que contiene la humedad del suelo se vacía parcialmente en una cantidad correspondiente a la percolación P_2 .

$$P_2 = S'' \left[1 - \left(1 + \left(\frac{S''}{X_1} \right)^3 \right)^{-1/3} \right] \quad (2-33)$$

Y el nivel S_{k+1} , que se usará para el cálculo del mes siguiente, está dado por:

$$S_{k+1} = S'' - P_2 \quad (2-34)$$

Distribución e intercambio exterior profundo

La cantidad de agua total P_3 que llega al reservorio de distribución será:

$$P_3 = P_1 + P_2 \quad (2-35)$$

El nivel R_k en el reservorio de distribución se convierte en:

$$R' = R_k + P_3 \quad (2-36)$$

Debido a que existe un intercambio subterráneo de agua, que puede ser agua que sale o entra a la cuenca o simplemente agua que no discurre como escorrentía superficial, se considera una cantidad F .

$$F = (X_2 - 1)R' \quad (2-37)$$

El parámetro X_2 es positivo y adimensional. El nivel en el reservorio se convierte en:

$$R'' = X_2 R' \quad (2-38)$$

El reservorio de distribución, de capacidad fijada igual a 60 mm, se vacía parcialmente para generar la caudal simulado Q_k , según la siguiente ecuación:

$$Q_k = \frac{R''^2}{R'' + 60} \quad (2-39)$$

El contenido del reservorio es actualizado:

$$R_{k+1} = R'' + Q_k \quad (2-40)$$

2.2 MARCO CONCEPTUAL

2.2.1 Precipitación

La precipitación se define como todo tipo de partículas de agua, sólidas o líquidas, que caen desde la atmósfera y llegan al suelo (Ahrens y Henson, 2018). Esta variable se mide normalmente en mm. Además, la precipitación es la variable hidrológica clave que permite relacionar la atmósfera con los procesos superficiales (Weather et al., 2008). Es así que la precisión de las estimaciones de respuesta hidrológica no puede ser mejor que la precisión de las evaluaciones de precipitación (Fekete et al., 2004).

La medición precisa de la precipitación es fundamental para el análisis de procesos hidrológicos cuantitativos como el pronóstico de inundaciones en tiempo real y la calibración y validación de modelos hidrológicos (McMillan et al., 2011). En ese sentido, entre los tipos de medición de precipitación encontramos las mediciones puntuales por medio de estaciones pluviométricas y las mediciones en áreas por medio de radares y satélites. Sin embargo, debido a los errores asociados a las técnicas de medición con radar y especialmente con satélites estas requieren una corrección a partir de observaciones de estaciones pluviométricas (Dingman, 2015). Asimismo, las observaciones en estaciones pluviométrica son susceptibles a diferentes fuentes de error como la presencia de

obstáculos a corrientes de aire alrededor del instrumento, la forma y orientación del dispositivo, posibles pérdidas por evaporación viento y salpicadura, así como errores inherentes a la operación de la estación (Hingray et al., 2014).

La formación de la precipitación requiere la condensación del vapor de agua atmosférico. Asimismo, la saturación es una condición esencial para el proceso de condensación. Pero la saturación en sí misma no es condición suficiente para la condensación, que usualmente requiere la presencia de núcleos de condensación alrededor de los cuales se forman gotas de agua o cristales (Musy y Higy, 2010). Las características de la precipitación dependen fuertemente en las condiciones atmosféricas de circulación y las características de las masas de aire. Por ejemplo, la precipitación convectiva resulta de la condensación de masas de aire húmedo inestables que se enfrían a medida que ascienden debido a su inestabilidad. Otro caso es el de la precipitación frontal que se produce cuando una masa de aire frío y una masa de aire cálido colisionan. Asimismo, la precipitación también depende de características de la superficie terrestre, como la orografía, el contraste entre el océano/continente y la cobertura. Este es el caso de la precipitación orográfica relacionada a la presencia de montañas que obligan a las masas de aire a ascender produciendo la condensación (Hingray et al., 2014).

2.2.2 Temperatura

Es el nivel de calor o frío de una sustancia, como es medida por termómetro. Es además una medida de la energía cinética o velocidad promedio de los átomos presentes en la sustancia (Ahrens y Henson, 2018). La medición de esta variable se realiza en grados centígrados (°C).

2.2.3 Evapotranspiración

La tropósfera, la capa de la atmósfera más cercana a la superficie, contiene en todo momento una cierta cantidad de agua en estado gaseoso que viene de:

- Evaporación de superficies abiertas de agua (océanos, mares, lagos, ríos), suelos desnudos y de superficies cubiertas por nieve o hielo.
- Transpiración de la vegetación que libera agua a la atmósfera.
- Evapotranspiración de suelo cubierto por vegetación.

La evapotranspiración es el término combinado que representa la transpiración y la evaporación del agua que ocurre en un ambiente vegetal. Los dos procesos son combinados en un solo término debido a la frecuente dificultad práctica de diferenciarlos (Musy y Higy, 2010). Además, esta variable al igual que la precipitación se representa en mm.

2.2.4 Caudal de escorrentía superficial

La escorrentía superficial representa la porción del suministro agua a la superficie que no es absorbida por el suelo ni se acumula en su superficie, sino que fluye pendiente abajo (Hillel et al., 2005). En ese sentido, los caudales de escorrentía superficial, o llamados solamente caudales en la presente tesis, son las mediciones del volumen escorrentía superficial que pasa por una sección en un intervalo de tiempo. Esta sección corresponde normalmente a la de un río o canal y los caudales son medidos usualmente en L^3/s .

2.2.5 Variabilidad espacial

Muchas variables meteorológicas como la precipitación, evapotranspiración y temperatura varían espacialmente sobre regiones geográficas que constituyen volúmenes de control para análisis hidrológicos. La descripción de la variabilidad espacial se basa en mediciones puntuales hechas en estaciones. En el caso de la precipitación la variabilidad espacial para una región será mayor para valores horarios y diarios en comparación con valores mensuales o anuales. Asimismo, una región con precipitación convectiva y una topografía variable tenderá a tener relativamente mayor variabilidad (Dingman, 2015).

2.2.6 Interpolación espacial

La interpolación espacial, también conocida como predicción espacial, es el proceso de la estimación de precipitación en los puntos de una grilla que cubra una región, a partir de valores medidos. Una vez que los valores son estimados para todos los puntos de la grilla, se puede calcular el promedio espacial para un área determinada. Todos los métodos de interpolación están basados en supuestos como el que la superficie a ser modelada es continua (sin huecos o saltos). Asimismo, es asumido que todos los puntos de la grilla tienen alguna relación con los valores medidos; esta relación es modelada por las funciones de interpolación (Dingman, 2015).

2.2.7 Distancia no euclídea

Las distancias usadas en geografía son medidas de la separación entre dos objetos geográficos (L'Hostis y Alain, 2020).

La distancia euclídea se define como la longitud del segmento de línea que une dos puntos. En el ámbito geográfico, la distancia entre dos puntos de la Tierra se calcula utilizando la longitud del segmento que los une en un gran círculo (great circle distance), es decir, la distancia más corta sobre una esfera está a lo largo de la trayectoria de un círculo cuyo centro es el centro de la esfera (Carter, 2002). Considerando esto, en la presente investigación se consideran como distancias no euclídeas aquellas que utilizan una definición diferente.

CAPITULO III: ÁREA DE ESTUDIO E INFORMACIÓN

En el presente capítulo se describe el área de estudio, así como la información básica usada para el desarrollo de la presente tesis. En líneas generales, se consideró como zona de estudio la zona andina peruana y tener una base de datos de calidad suficiente para realizar un correcto análisis. Es importante notar que los procedimientos descritos en este y los siguientes capítulos fueron desarrollados con el lenguaje de programación R (R Core Team, 2020).

3.1 FUENTES DE INFORMACIÓN

3.1.1 Topografía

Para esta información se usó el modelo de elevación digital SRTM (Shuttle Radar Topography Mission, Misión topográfica Radar Shuttle) en su versión 3, la cual ofrece una resolución aproximada de 90 y 30 metros en sus formatos de 3 y 1 segundo de arco respectivamente. Esta información se encuentra disponible en <https://earthdata.nasa.gov> y permitirá tener una cobertura de la elevación a nivel nacional (Figura N° 11) y será usada como variable auxiliar en los métodos de interpolación, para lo cual es agregada a una resolución de grilla de 0.02 °. <https://earthdata.nasa.gov/>

3.1.2 Hidrometeorología

La información de precipitación principalmente y caudales en ciertas estaciones fue obtenida del Sistema Nacional de Información de Recursos Hídricos (SNIRH) administrado por la Autoridad Nacional del Agua (ANA), donde se encuentra a disposición del usuario información hidrometeorológica de diferentes fuentes, tales como SENAMHI, ANA, diferentes Juntas de Usuarios, proyectos especiales, etc.

3.2 ÁREA DE ESTUDIO

Si bien es cierto la metodología de distancias no euclídeas es aplicable en cualquier zona, su finalidad es principalmente mejorar la precisión de interpolación en zonas de alta montaña donde la topografía es complicada y normalmente existe una mayor incertidumbre. Es por ello que se escoge la zona andina peruana como área de estudio.

Además de exploraciones previas, se observó que en cuencas de la vertiente del Pacífico existen dos zonas que se diferencian claramente por su gradiente altitudinal. Teniendo ello en consideración se ha tomado la región con una elevación mayor a 2500 msnm como objeto de estudio. En la zona de la vertiente del Atlántico se observó además una muy baja densidad de estaciones en la zona nororiental, considerando además que la vegetación puede ser un indicador de diferentes condiciones climáticas o regímenes de precipitación (Brovkin, 2002; Navarro, 2019) se consideró el mapa global de bosques de Hansen et al. (2013) para separar la región de estudio de la zona influenciada por el bosque tropical amazónico.

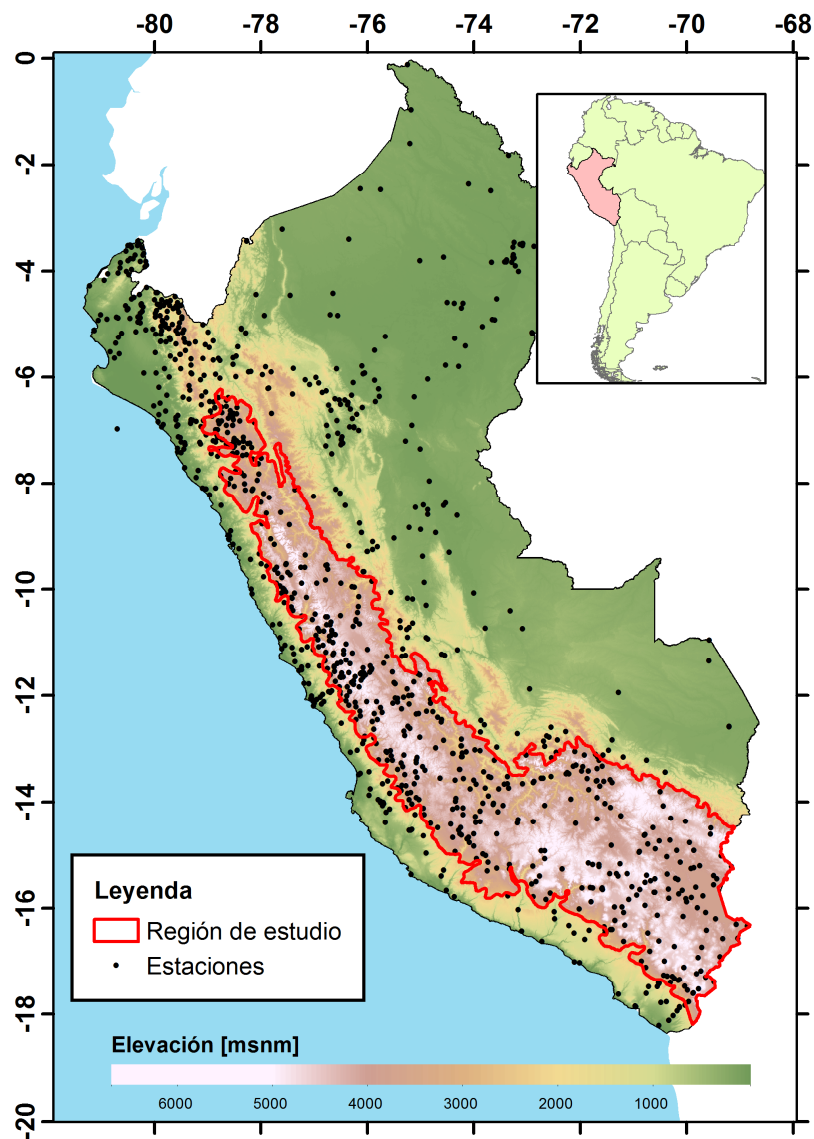


Figura N° 11. Área de estudio, mapa de elevaciones SRTM y red de estaciones

Fuente: Elaboración propia

Como zona de análisis se ha considerado el área continua de mayor área, donde se encontraron estaciones para realizar un análisis adecuado. Asimismo, se aplicó un suavizamiento a los límites de esta área con el fin de tener un área lo más continua posible y considerar los valles profundos cercanos a su frontera. Esta zona se muestra delimitada en la Figura N° 11 y cubre una superficie de 325 980 km².

3.3 INFORMACIÓN DE PRECIPITACIÓN

Los datos de precipitación son la base esencial de esta investigación. Es importante tomar un periodo adecuado de análisis considerando la disponibilidad de información que tenemos. En la Figura N° 12 se observa la disponibilidad de estaciones para el periodo seleccionado, observándose que para la zona de estudio existen cerca de 200 estaciones disponibles en cualquier mes. Se observó en la data que los primeros años, anteriores a 1970, presentaban estaciones que solo funcionaron en ese periodo y otras que no guardaban consistencia con los datos posteriores. Considerando esto, se eligió el periodo que va de 1970 a 2017 para ejecutar el análisis de los datos de precipitación.

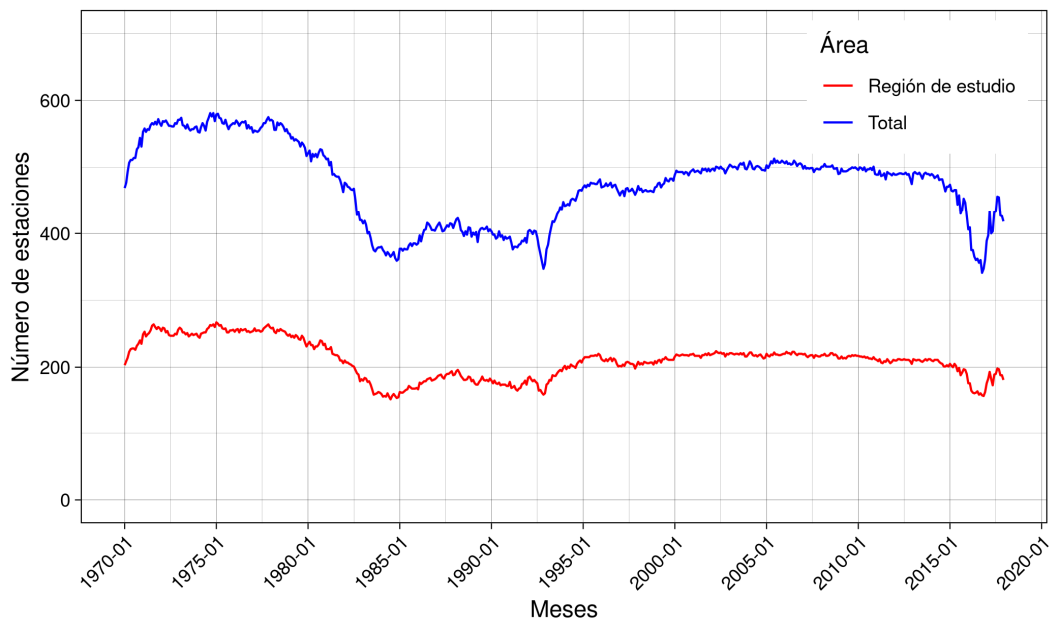


Figura N° 12. Disponibilidad en el tiempo de estaciones de la base de datos

Fuente: Elaboración propia

Si bien es cierto que la información meteorológica de la base de datos usada proviene de diferentes instituciones, en su mayor parte las estaciones son del

SENAMHI. De esta última fuente la mayoría de sus datos originales se encontraron a paso de tiempo de cada 12 horas. Con respecto a otras fuentes, la información estaba en formatos diarios y mensuales. Los datos con paso de tiempo diferente al mensual fueron procesados para obtener este tipo de información y todas las estaciones pasaron un control de calidad que se describe posteriormente.

Los datos de precipitación fuera del área de estudio no han sido ignorados completamente. Si bien es cierto que la precisión de la interpolación es analizada solo en la región seleccionada. Las estaciones en regiones exteriores cercanas también son útiles para estimar la precipitación en ubicaciones dentro del límite de la región de estudio, pero cercanas a su frontera. Es por ello que la base de datos será tratada en su conjunto para los análisis posteriores, solo separándose las estaciones dentro de la región de estudio para fines relacionados a la evaluación de la interpolación.

Asimismo, en la Figura N° 11 se observa que la red de estaciones disponibles no es uniforme espacialmente. Existen zonas que están mucho mejor muestreadas que otras, lo que influirá indirectamente en los resultados y precisión de la interpolación. En general se ve una falta de estaciones en zonas cercanas a la selva, que es especialmente notoria en la zona noreste de la región de estudio. Asimismo, la zona de la vertiente del Pacífico es la que concentra una mayor densidad de estaciones.

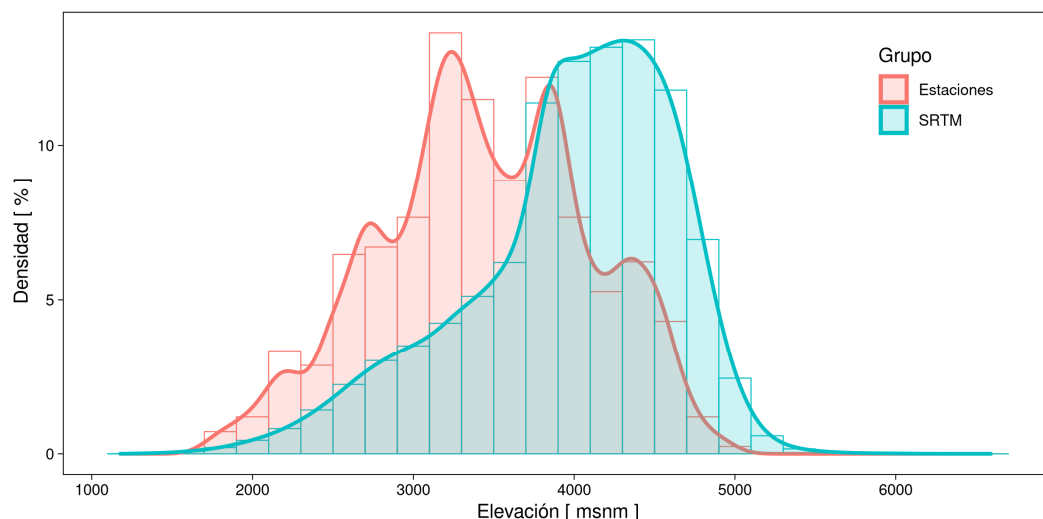


Figura N° 13. Densidad de elevaciones de estaciones y DEM en el área de estudio

Fuente: Elaboración propia

Se debe considerar además que las estaciones tienden a concentrarse en zonas de menor elevación o de valle, donde generalmente están los poblados. Es así que las zonas altas de montañas tienen una menor representación. Este contraste es apreciable en la región de estudio. En la Figura N° 13 se observa que las estaciones tienen una mayor concentración entre los 3000 y 4000 msnm, mientras que las altitudes del área de estudio se ubican más densamente entre los 3800 y 4600 msnm aproximadamente. Es así como existe un desfase en la representación de elevaciones de las estaciones y las zonas altas están subrepresentadas.

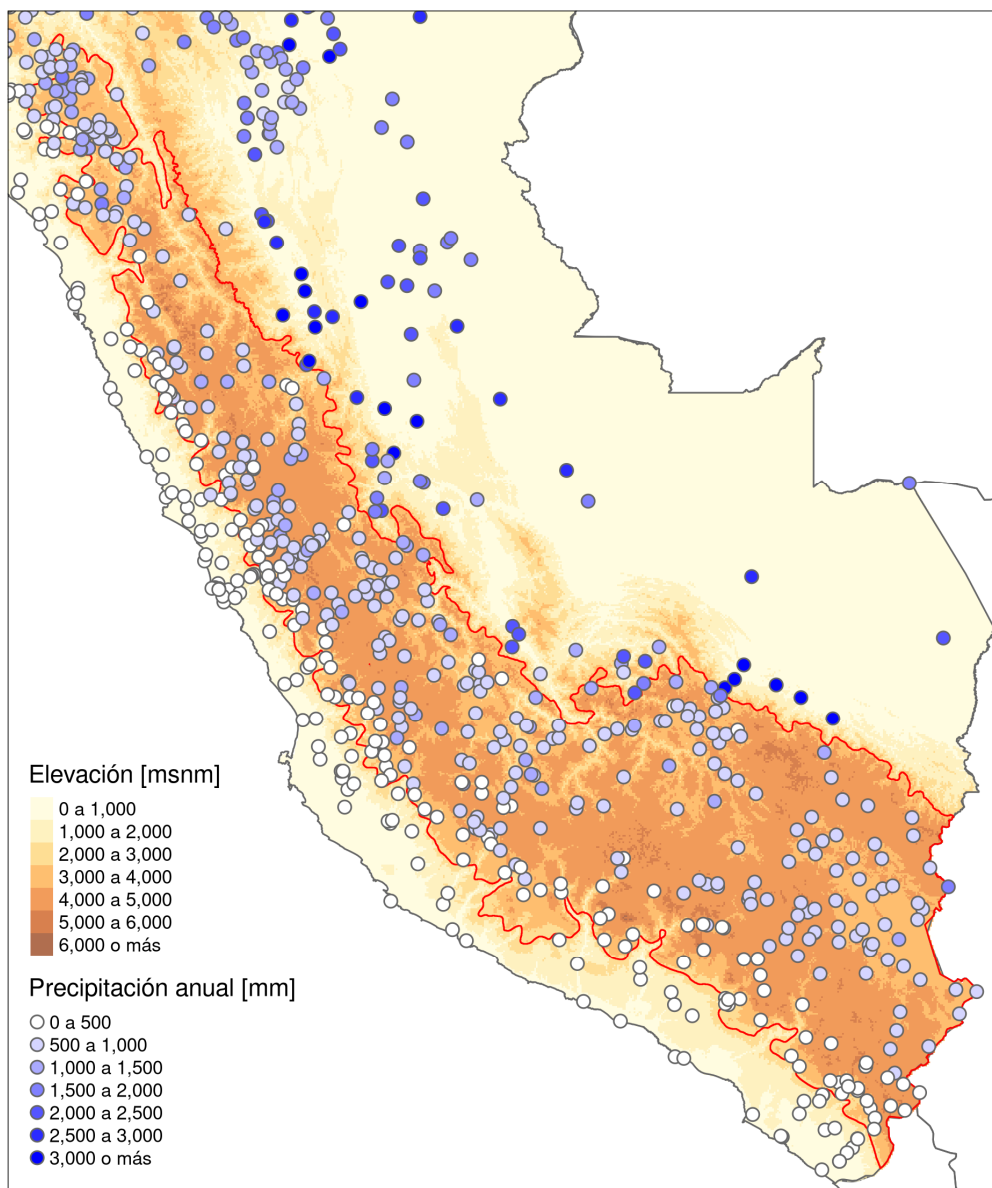


Figura N° 14. Precipitación media anual en estaciones

Fuente: Elaboración propia

En la Figura N° 14 y Figura N° 15 se representan respectivamente la precipitación media anual y el número de años de registro. Se observa que las mayores precipitaciones se registran en zonas cercanas a la selva amazónica.

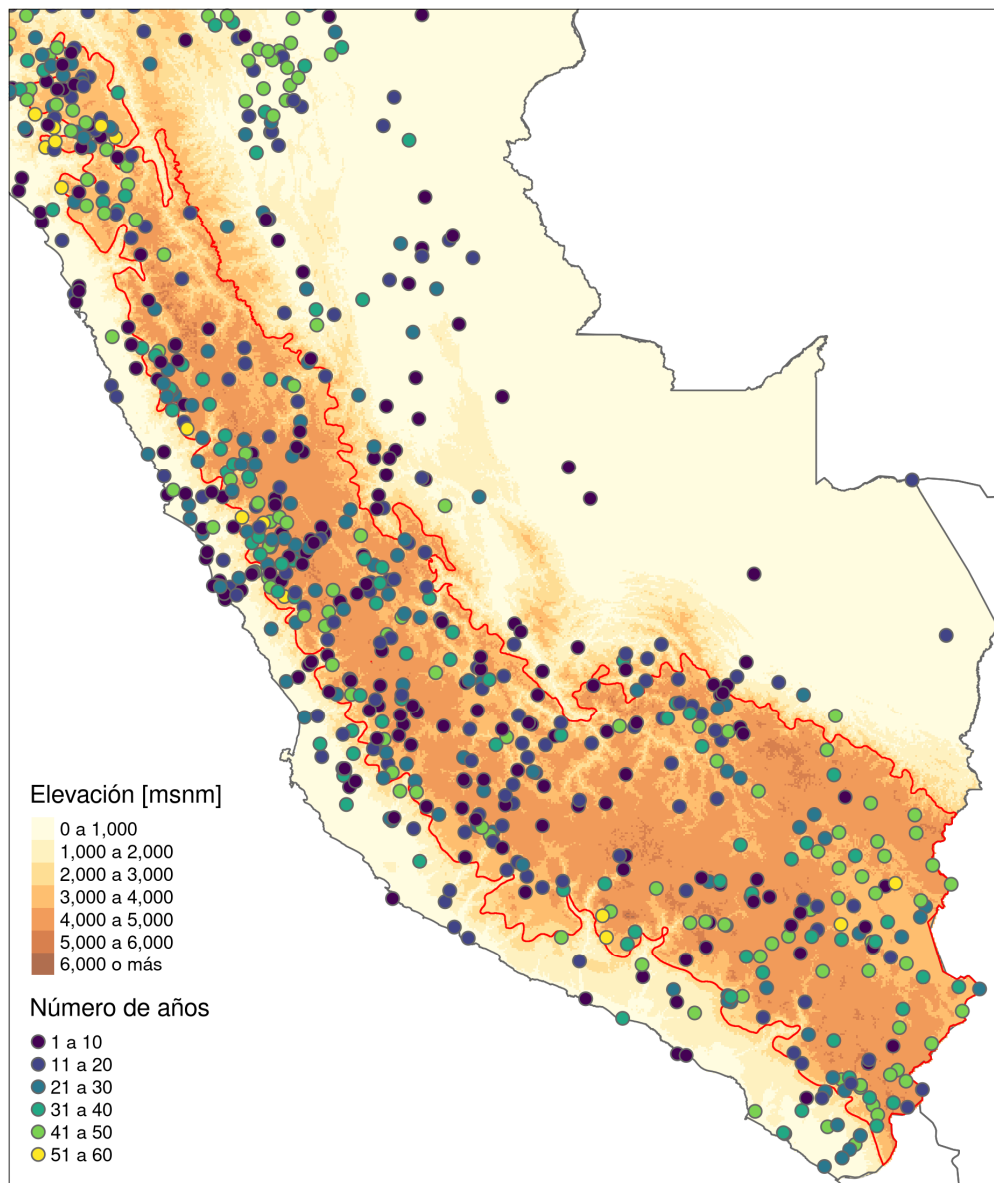


Figura N° 15. Número de años de registro en estaciones

Fuente: Elaboración propia

3.4 ANÁLISIS DE LA INFORMACIÓN HIDROMETEOROLÓGICA

La validación de los datos de precipitación es una etapa fundamental para asegurar filtrar los posibles errores en la data y que estos no generen cálculos erróneos en la evaluación de la interpolación. Por otro lado, no existe un procedimiento estandarizado para el control de calidad de datos de precipitación (Branisavljević et al., 2009), es así que se propone una serie de controles basados en investigaciones previas desarrolladas en el país.

3.4.1 Valores extremos

Se ha llegado a encontrar en la data valores muy por encima del resto en sus respectivas estaciones. Para este control se consideraron las estaciones con precipitaciones diarias o cada 12 horas. Se realizó una inspección visual de las series de tiempo agrupadas por cuenca, y se identificaron aquellas estaciones que presentaran valores extremos sospechosos. En estos casos, se siguió un análisis similar al descrito por Aybar et al. (2020) y Keller et al. (2015). Un límite de 200 años de periodo de retorno fue usado para identificar estos extremos. La precipitación límite con este periodo de retorno fue calculada para cada estación, considerando la distribución Gumbel, y la distribución de valores extremos generalizada (GEV por sus siglas en inglés), de la cual la distribución Gumbel es un caso particular, dependiendo de cada caso.

Considerando estos valores extremos, se tomaron como válidos si cumplían al menos una de 2 condiciones con respecto a las estaciones más cercanas dentro de un radio de 20 km. Si al menos una de estas estaciones presentaba un valor de por lo menos el 50 % del valor evaluado o si al menos la mitad de ellas presentaban valores mayores al 20 %, el valor extremo era considerado válido. Además, considerando que la densidad de estaciones no es igual en todo el territorio, si el número de estaciones cercanas era menor a 4 y no se cumplían las condiciones previas, se permitía realizar la evaluación en un radio de 40 km.

Considerando los valores extremos que no cumplían las condiciones previas, se realizaba una última comparación con las estaciones más cercanas para ver si estos órdenes de magnitud se podían presentar en estas otras. Finalmente, en caso de no cumplir ninguna de estas condiciones, el dato extremo era descartado y se consideraba como desconocido.

En la Figura N° 16 se representa el límite de la precipitación a 200 años de periodo de retorno con la línea roja. En las estaciones mostradas los valores no cumplen las condiciones establecidas y son filtrados. Asimismo, esta forma de filtrado permitió identificar valores claramente erróneos correspondientes a valores 999, cuya simbología está relacionada a datos faltantes. En total se filtraron datos de 25 estaciones.

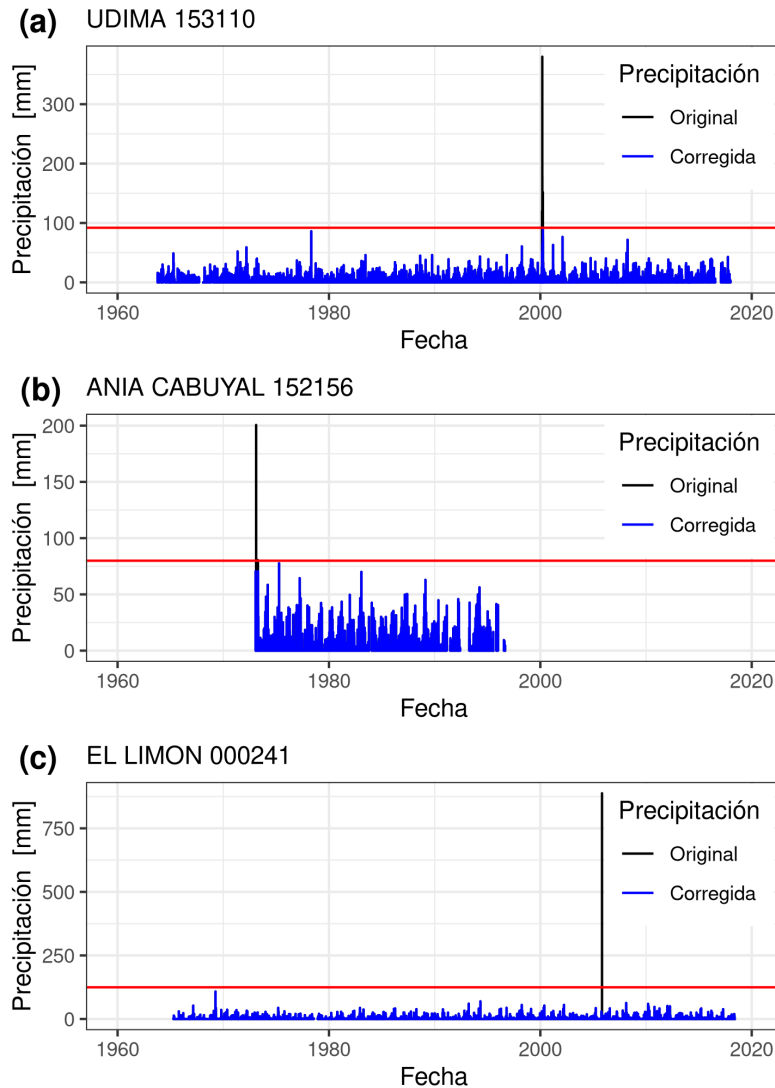


Figura N° 16. Ejemplos de filtrado de estaciones con valores extremos

Fuente: Elaboración propia

3.4.2 Quiebres y segmentos inconsistentes obvios

Para esta etapa se siguió una identificación visual basada en el agrupamiento por cuencas. Para poder identificar estos segmentos, los periodos sospechosos fueron comparados con el resto del registro de la estación y con las series de las estaciones cercanas. Se pudieron identificar con más detalle los límites de estos segmentos por medio de gráficos interactivos de las series de tiempo de las estaciones y se consideraron tramos limitados por periodos sin datos de darse el caso. Finalmente, los segmentos inconsistentes identificados fueron eliminados de la data. En la Figura N° 17 se muestran ejemplos de ello.

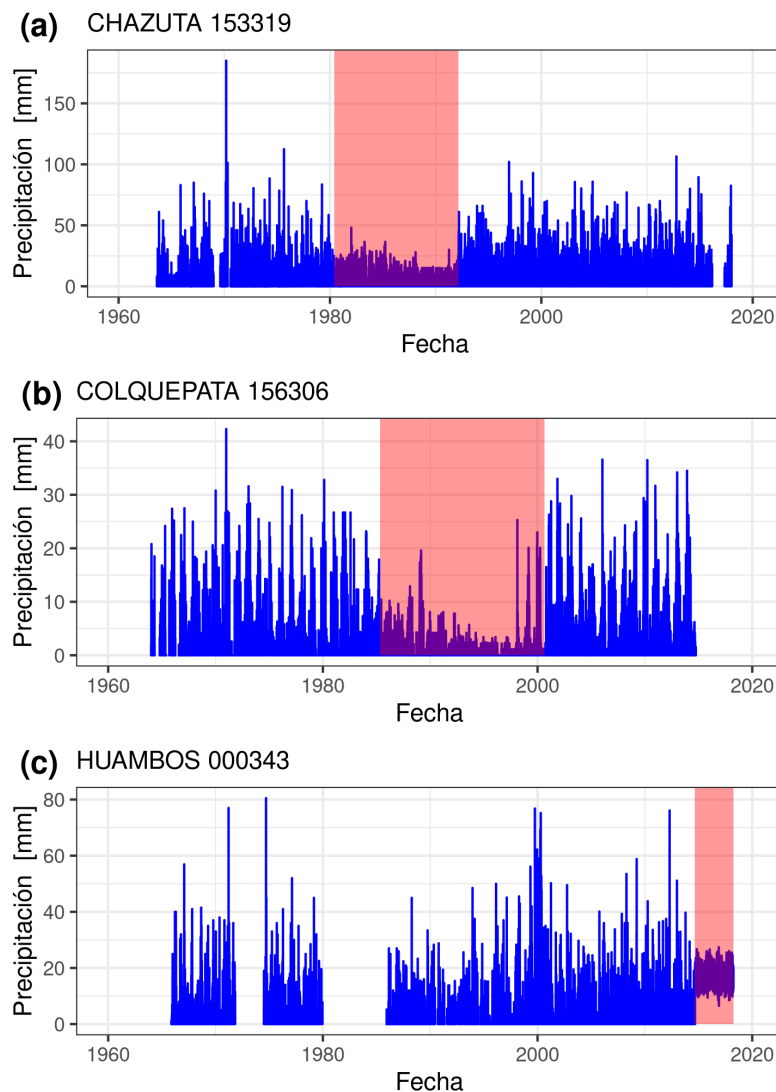


Figura N° 17. Ejemplos de filtrado de estaciones con segmentos inconsistentes

Fuente: Elaboración propia

3.4.3 Homogeneidad de los datos

La finalidad de esta etapa es evaluar la homogeneidad de los datos e identificar información inconsistente producto de problemas como microclimas, instrumentación o variaciones en tiempo y posición. La metodología usada es la de Vector Regional (VR), que se considera frecuentemente en otras investigaciones (Espinoza et al., 2009; Rau et al., 2017). Si bien es cierto existen variantes del vector regional, la más usada es la propuesta por Brunet-Moret (1979), que es el enfoque seguido en esta tesis.

La metodología del Vector Regional considera el principio de la proporcionalidad de precipitaciones anuales entre estaciones vecinas representados como índices de precipitación que caracterizan el patrón de precipitación en un área determinada (Rau et al., 2017)

Dentro de la zona de análisis, se deben calcular los valores de los índices pluviométricos anuales Z_i y las precipitaciones medias extendidas P_j para cada estación. Esto se consigue por medio de la resolución aproximada de un sistema de ecuaciones, que es producto de la minimización por mínimos cuadrados de la siguiente expresión.

$$\sum_{i=1}^N \sum_{j=1}^M \left(\frac{P_{ij}}{P_j} - Z_i \right)^2 \quad (3-1)$$

Donde N representa el número de años y M el número de estaciones. Asimismo, i y j representan los índices de los años y estaciones respectivamente, de donde P_{ij} representa la precipitación anual en la estación j y año i . El vector regional que representa el régimen de la zona corresponde a los índices Z_i .

Una vez calculados los respectivos índices, se puede evaluar la pertenencia de una determinada estación a la zona por medio de dos índices:

- el coeficiente de correlación r entre el vector regional y los índices pluviométricos anuales de las estaciones
- la desviación estándar de las diferencias entre los índices pluviométricos de las estaciones y el vector regional.

Rau (2017) sugiere que la desviación estándar debe estar por debajo de 0.4 y que el coeficiente de correlación debe ser mayor que 0.7. Un valor mínimo de 0.5 puede llegar a ser aceptable para zonas de la costa baja debido a las condiciones secas y que generan precipitaciones muy bajas o nulas. Estos serán los límites que se usarán para identificar la pertenencia de las estaciones a un grupo. Asimismo, se permite un límite de 0.5 en la correlación solo para zonas menores a los 2000 msnm, esto engloba también indirectamente a la selva, en la que se considera aceptable debido a la baja densidad de estaciones. Se han probado otros posibles criterios para definir la pertenencia a una zona con resultados también aceptables, entre ellos el índice de Nash-Sutcliffe, sin embargo, sus resultados no han sido utilizados. Otro posible criterio podría ser el de los límites multiplicativos inferior y superior definidos por Brunet-Moret (1979).

Se suele agrupar las estaciones por zonas climáticas, de tal forma que sea más probable su pertenencia a ellas. Sin embargo, las estaciones de las fronteras de estas zonas climáticas suelen presentar índices menores, lo que dificulta su pertenencia a una zona climática al compararlo con el vector regional (Espinoza et al., 2009). Es por ello que se plantea realizar una zonificación dinámica, donde se analiza la vecindad cercana de cada estación como área para encontrar un vector regional, de esta forma se evita que estaciones en las fronteras de posibles zonas climáticas se vean perjudicadas.

Estas vecindades de análisis han sido fijadas al rango de 6 a 9 estaciones más cercanas. En cada una de ellas se calcula el vector regional, y se van eliminando estaciones que no cumplan los criterios de pertenencia establecidos. De esta forma, se tienen 2 posibilidades, una donde se obtiene para esta vecindad un grupo de estaciones consistentes entre ellas y la otra donde el grupo queda vacío.

Finalmente, se seleccionan solo los grupos de al menos 4 estaciones consistentes. Solo las estaciones que pertenecen a alguno de estos grupos son consideradas para el estudio.

Los coeficientes de correlación correspondientes a los grupos hallados para cada estación se muestran en la Figura N° 18, donde se observa que para la región de estudio gran parte de las estaciones presentan un r mayor a 0.8.

Para este proceso se consideraron solo las estaciones con al menos 5 años de datos, para poder comparar de manera realista sus índices con los del vector

regional estimado. El número total de estaciones seleccionadas es de 698, de las cuales 382 pertenecen a la región de estudio. El listado de estas últimas estaciones se muestra en el Anexo 1.

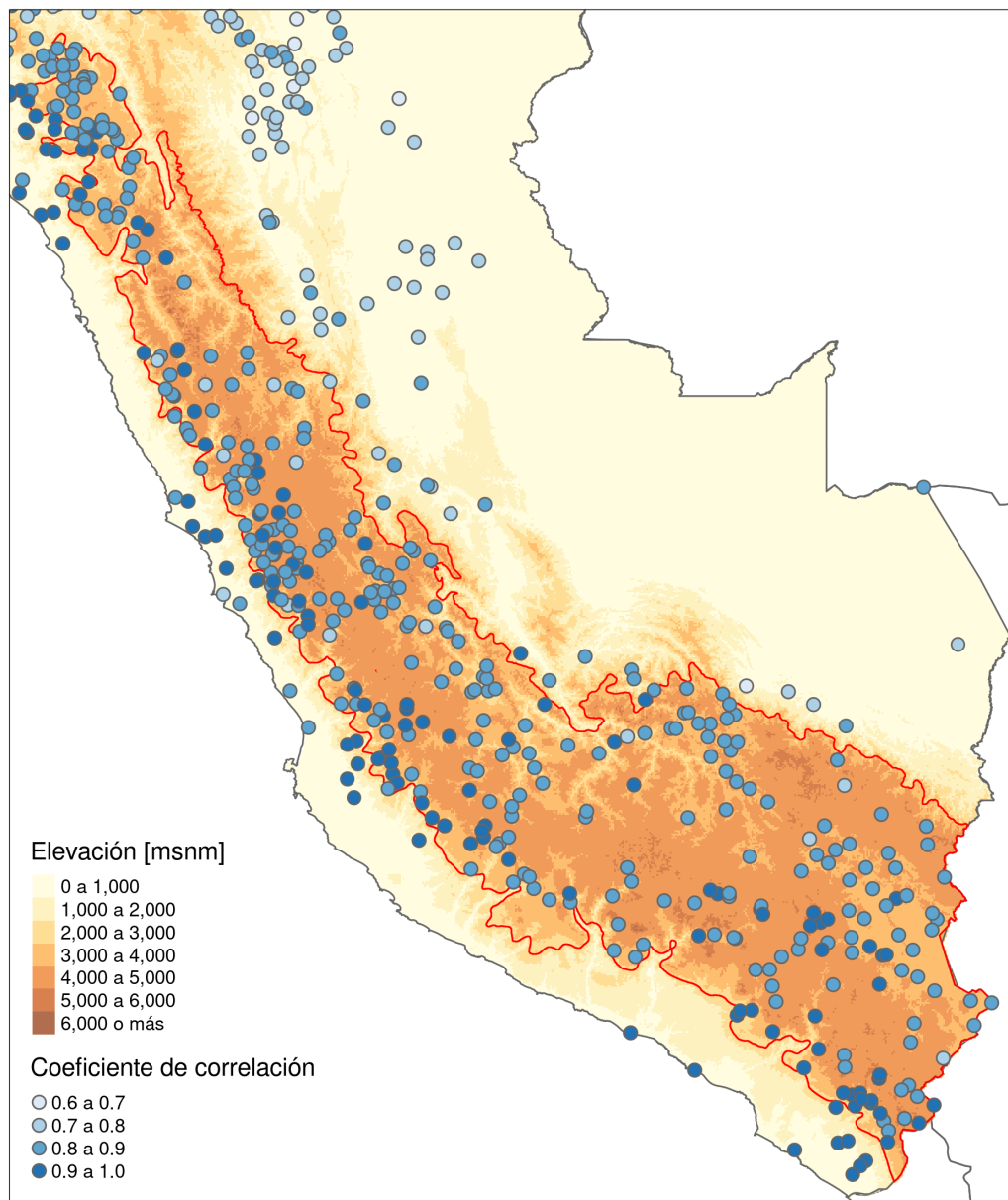


Figura N° 18. Coeficiente de correlación para estaciones seleccionadas

Fuente: Elaboración propia

CAPITULO IV: MÉTODOS Y DESARROLLO

4.1 ZONAS DE ANÁLISIS

Considerando la extensión del área de estudio mostrada en el capítulo anterior, se realizó una división en subregiones, para que las condiciones locales se puedan ver reflejadas en la interpolación y realizar así un análisis más preciso. Estas zonas se delimitaron considerando cuencas hidrográficas contiguas y la división por vertientes. Asimismo, se buscó contar con un adecuado número de estaciones en cada una. Estas zonas se presentan en la Figura N° 19.

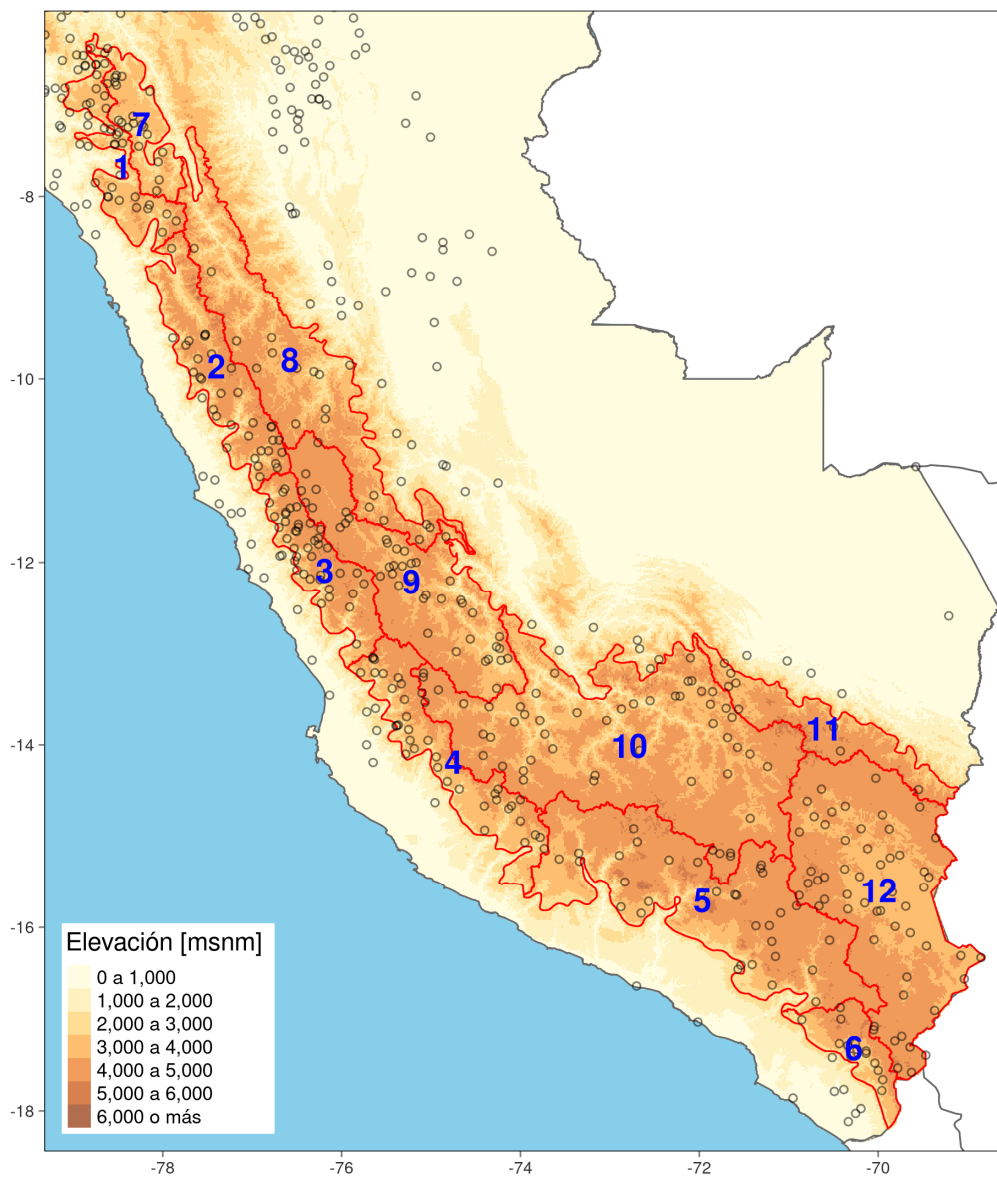


Figura N° 19. Zonas de análisis y estaciones

Fuente: Elaboración propia

Información adicional de cada zona se presenta en Tabla N° 1. Las zonas de la 1 a la 6 corresponden a la vertiente del Pacífico, mientras que las zonas de la 7 a la 11 a la vertiente del Atlántico. La parte peruana de la vertiente del Titicaca se corresponde con la zona 12.

Tabla N° 1. Datos adicionales de zonas de análisis

Fuente: Elaboración propia

Zona	Elevación [msnm]			Precip. anual media [mm]	Número de estaciones	Área [10 ³ km ²]
	Máx.	Mín.	Media			
1	5019	1646	3304.02	968.84	27	10.01
2	6738	1352	3904.98	607.35	32	18.02
3	5866	1445	4007.49	522.38	39	13.20
4	5266	1376	3641.86	433.68	37	18.36
5	6403	1118	4141.00	426.00	30	48.17
6	5797	1752	3854.92	157.75	14	7.52
7	4702	1396	3236.05	881.64	21	9.06
8	6423	671	3770.13	831.09	19	35.95
9	5668	1165	4026.81	789.39	42	31.91
10	6222	1111	4015.68	777.27	50	71.21
11	6312	1841	4190.60	878.82	8	13.38
12	5815	3806	4226.29	703.88	53	49.19

En los procesos de interpolación se usaron las estaciones dentro de cada una de las subregiones, además de las que estuvieran en un radio de 40 km de estas. Se consideraron estas estaciones cercanas con el fin de que el proceso de estimación de precipitación en las ubicaciones próximas a los límites no se vea perjudicado.

4.2 DISTANCIAS NO EUCLÍDEAS PROPUESTAS

El fin de las distancias no euclídeas aquí propuestas es representar con mayor exactitud la “proximidad espacial” entre dos ubicaciones. En particular, accidentes geográficos como montañas o elevaciones topográficas son tomados en cuenta y pueden generar una mayor separación entre dos ubicaciones que pueden parecer próximas.

En ese sentido, la distancia propuesta por Frei (2014) cumple este objetivo y además se han obtenido buenos resultados obtenidos para la interpolación de temperatura en Suiza. Sin embargo, el autor menciona que el procedimiento para calcular esta distancia puede ser computacionalmente pesado. Es por ello que aquí se proponen otras dos aproximaciones más sencillas a esta distancia, que también es evaluada.

Un parámetro que es fundamental y que comparten las tres aproximaciones, es el coeficiente λ . Este coeficiente, como se observa en la ecuación (2-27) penaliza los desplazamientos verticales. Un $\lambda = 0$ sería equivalente a la distancia euclídea convencional para todos los casos, mientras que un mayor λ representa un mayor distanciamiento entre zonas de distinta elevación y una mayor cercanía relativa entre zonas con elevaciones similares.

En tal sentido, un mayor λ también significaría mayores distancias ante obstrucciones topográficas. De igual forma, esto resulta en una mayor influencia del relieve topográfico. La forma particular de usar este coeficiente se detallará en la subsiguiente descripción de cada una de las distancias propuestas.

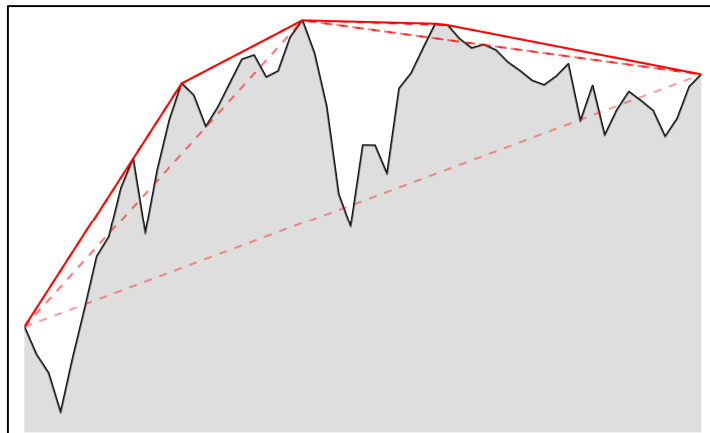


Figura N° 20. Distancia mínima por sobre el perfil longitudinal

Fuente: Elaboración propia

4.2.1 Distancia no euclídea 1 (NE 1)

Esta distancia se basa en el recorrido por sobre el perfil topográfico entre las 2 ubicaciones. En este caso, se extrae el perfil topográfico del modelo de elevación y se calcula la distancia del recorrido mínimo por sobre este perfil. En el cálculo de distancias parciales se considera el coeficiente λ que multiplica a las

componentes verticales. Esta aproximación muestra similitudes con la forma de calcular la distancia descrita por Gottardi et al. (2012)

En la Figura N° 20 se observa esta distancia mínima. Para su cálculo se sigue un proceso iterativo, partiendo de un trazo que es la línea más corta, sin considerar el perfil topográfico, que conecta la ubicación inicial y final, se agregan ubicaciones de paso o vértices a esta línea. Estos, puntos se agregan uno en cada etapa, tomando para ello el punto del perfil que esté más alejado del trazo en el paso anterior. El proceso se detiene cuando el trazo está por sobre cualquier punto del perfil longitudinal.

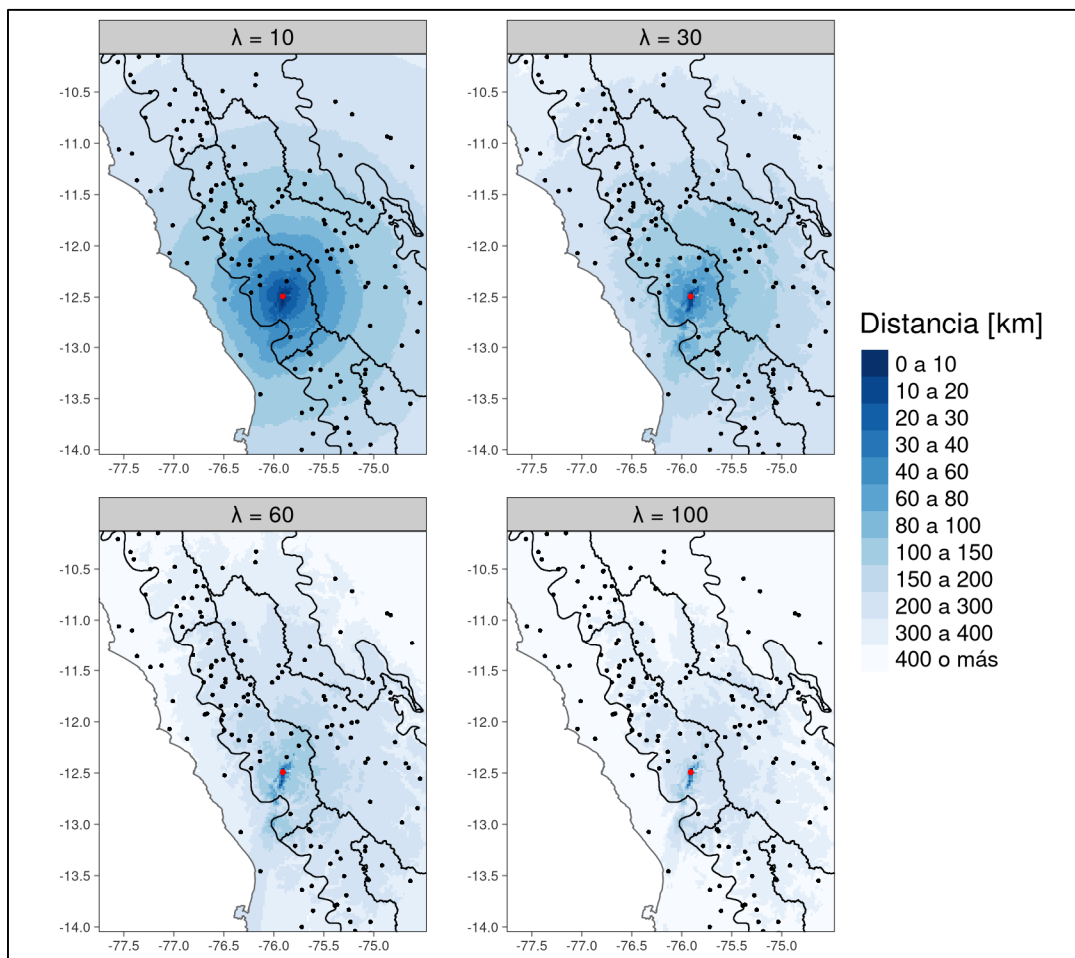


Figura N° 21. Distancia no euclídea 1 (NE 1) desde estación en valle (Yauyos)

Fuente: Elaboración propia

Esta distancia se presenta desde una ubicación localizada en un valle (Figura N° 21) y una ubicada en una zona montañosa (Figura N° 22). Se observa el distinto comportamiento para λ altos en cada caso, que responde a las condiciones

topográficas propias de cada zona. En general, ambas ilustraciones muestran que menores distancias corresponden a menores obstrucciones, en particular en la línea geodésica que uniría dos ubicaciones y que corresponde a la proyección del perfil longitudinal. Es así que la zona del valle se muestra aislada mientras que la ubicación en montaña parece tener una mayor conectividad.

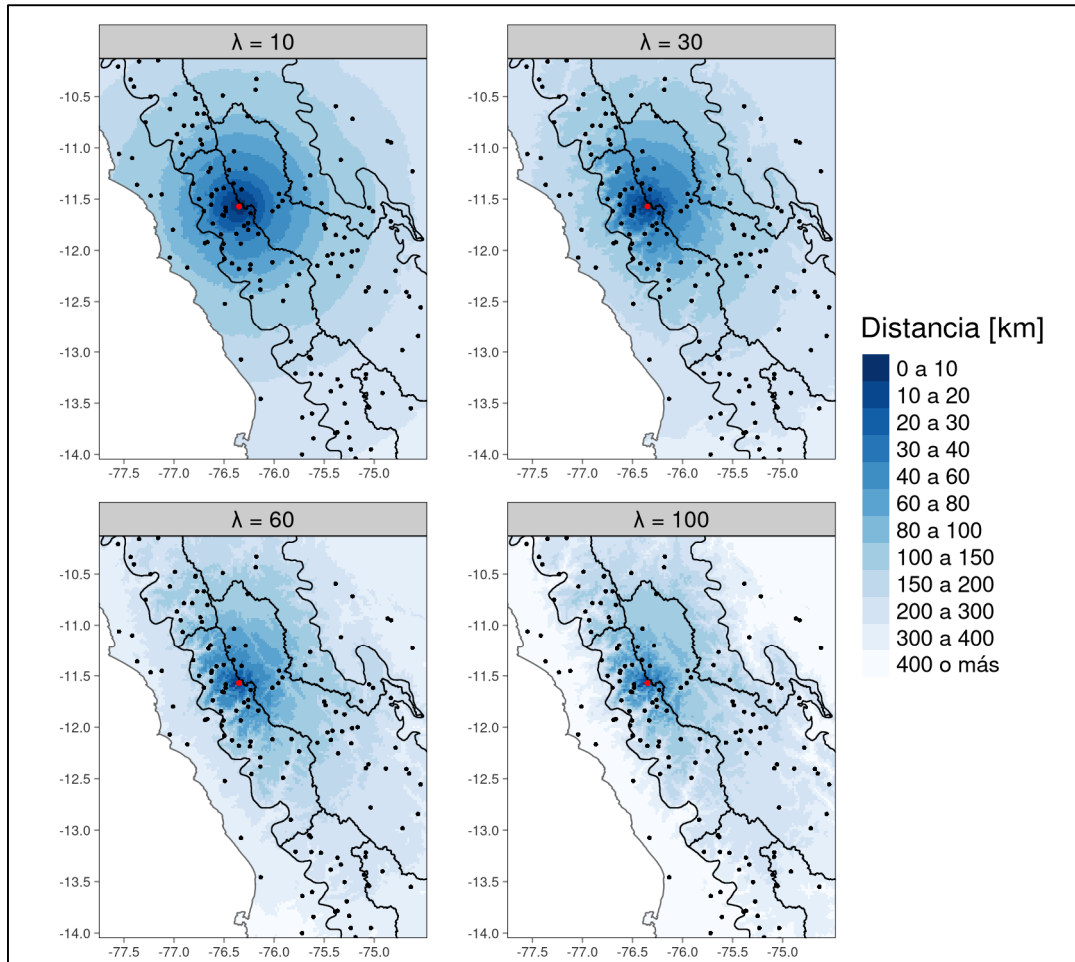


Figura N° 22. Distancia no euclídea 1 (NE 1) desde estación montañosa (Milloc)

Fuente: Elaboración propia

4.2.2 Distancia no euclídea 2 (NE 2)

Esta distancia sigue la aproximación propuesta por Frei (2014). Esta distancia es equivalente a la del menor recorrido entre dos ubicaciones, considerando desplazamientos sobre la superficie topográfica y en la atmósfera libre donde sea posible. Además, la componente vertical de los desplazamientos es multiplicada por un coeficiente λ , con el fin de representar la influencia de esta componente con respecto a la horizontal.

El cálculo de esta distancia generalizada puede ser resuelto transformándolo a un problema del camino más corto o recorrido mínimo propio de la teoría de grafos. En ese sentido las estaciones y celdas representan los vértices del grafo y las conexiones permitidas las aristas. Asimismo, debido al alto costo computacional que representa la estimación de este tipo de distancias (Frei, 2014), se procedió a hacer uso del clúster del Laboratorio Nacional de Hidráulica. El proceso de cálculo de esta distancia se realiza en 2 etapas.

4.2.2.1 Primera etapa

En esta etapa se calculan las posibles conexiones entre todas las ubicaciones. Estas ubicaciones pueden ser celdas o las propias estaciones.

Dos ubicaciones son consideradas conectadas si no presentan obstrucciones en su perfil longitudinal, es decir si es posible unir las directamente mediante la línea más corta. Para la búsqueda de posibles conexiones se ha considerado un radio alrededor de cada celda de 40 km como espacio de búsqueda. Estas conexiones representan los posibles desplazamientos en la atmósfera libre y permiten conectar o representar la cercanía de, por ejemplo, dos ubicaciones de elevaciones cercanas en un mismo valle. Asimismo, las celdas vecinas en las 8 direcciones son consideradas conectadas a la celda central. La lista de conexiones es procesada previamente y almacenada para los procesos posteriores.

4.2.2.2 Segunda etapa

En la segunda etapa se calcula la matriz de adyacencia, esta matriz almacena las distancias o pesos de cada conexión. Por ejemplo, la entrada 30 en la columna i y fila j representa que existe una conexión entre las ubicaciones i y j con una distancia de 30 km. Es decir, cada conexión tiene asociado una distancia, que es calculada en base a la componente horizontal y la vertical multiplicada por el coeficiente λ . En ese sentido se calcula una matriz de adyacencia para cada λ que se evalúa.

Finalmente, la distancia asociada a dos ubicaciones se calcula encontrando el recorrido con la distancia mínima entre ambas. Para ello, se debe resolver el problema del camino más corto considerando todas las conexiones y sus pesos

respectivos. Esto se resuelve aplicando el algoritmo de Dijkstra (1959) sobre la matriz de adyacencia.

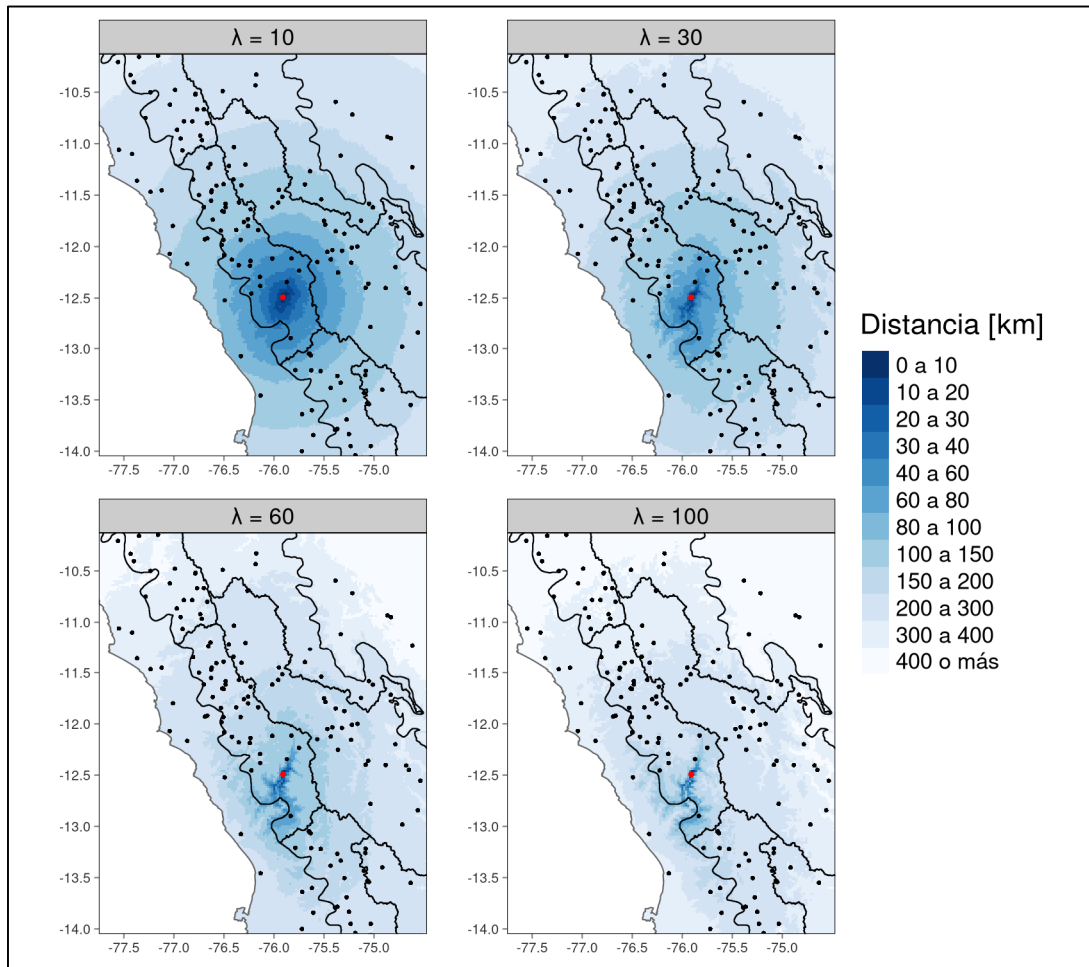


Figura N° 23. Distancia no euclídea 2 (NE 2) desde estación en valle (Yauyos)

Fuente: Elaboración propia

En la Figura N° 23, la distancia desde la estación Yauyos, ubicada en valle, muestra a mayores λ una mayor cercanía relativa entre ubicaciones dentro de la zona de valle cercana. Asimismo, también se ven favorecidas zonas con elevaciones cercanas y la forma del relieve topográfico es más marcada.

En comparación con la distancia NE 1 representada en la Figura N° 21, se aprecian menores magnitudes. Esto es razonable considerando que el perfil longitudinal entre dos ubicaciones es solo uno de los múltiples caminos posibles evaluados para calcular la presente distancia, es decir, se puede considerar la distancia NE 1 como un caso particular o simplificado de la distancia NE 2. Otra

consecuencia con λ altos es que las menores distancias en la aproximación NE 1 tienden a concentrarse en el eje longitudinal del valle, mientras que en el presente caso las distancias menores y la capacidad de compartir información también se presentan en las zonas bajas adyacentes.

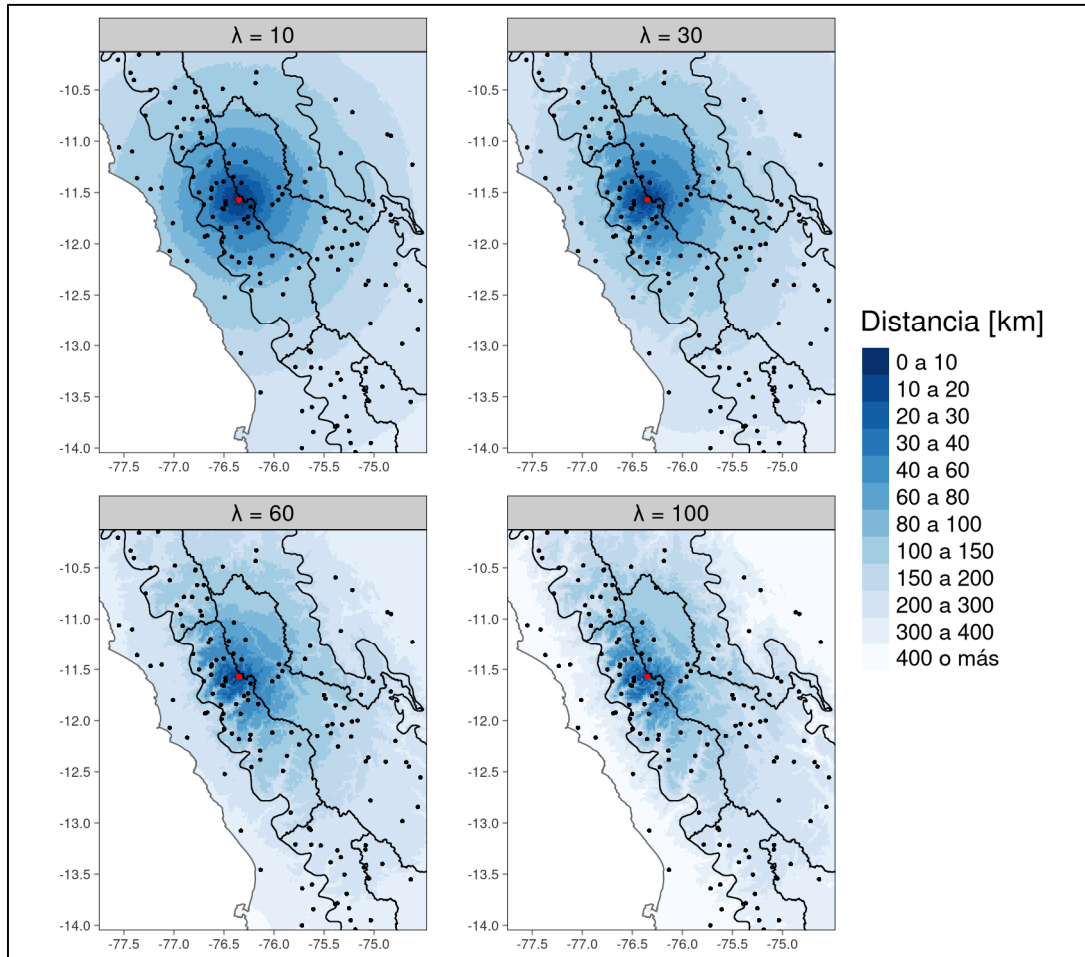


Figura N° 24. Distancia no euclídea 2 (NE 2) desde estación montañosa (Milloc)

Fuente: Elaboración propia

La distancia desde una estación de montaña, representada en la Figura N° 24, muestra las distancias menores y mayor conectividad entre zonas montañosas, especialmente para altos valores de λ . Asimismo, se puede apreciar una estratificación dependiente de la elevación.

4.2.3 Distancia no euclídea 3 (NE 3)

Esta distancia generalizada puede ser considerada una simplificación de la anterior. El procedimiento de cálculo es similar, salvo en el cálculo de conexiones.

En esta aproximación solo los recorridos en el terreno son permitidos, es decir, las conexiones a través de la atmósfera libre no son posibles. Para el cálculo de conexiones se consideran solo los vecinos en las 8 direcciones de cada ubicación.

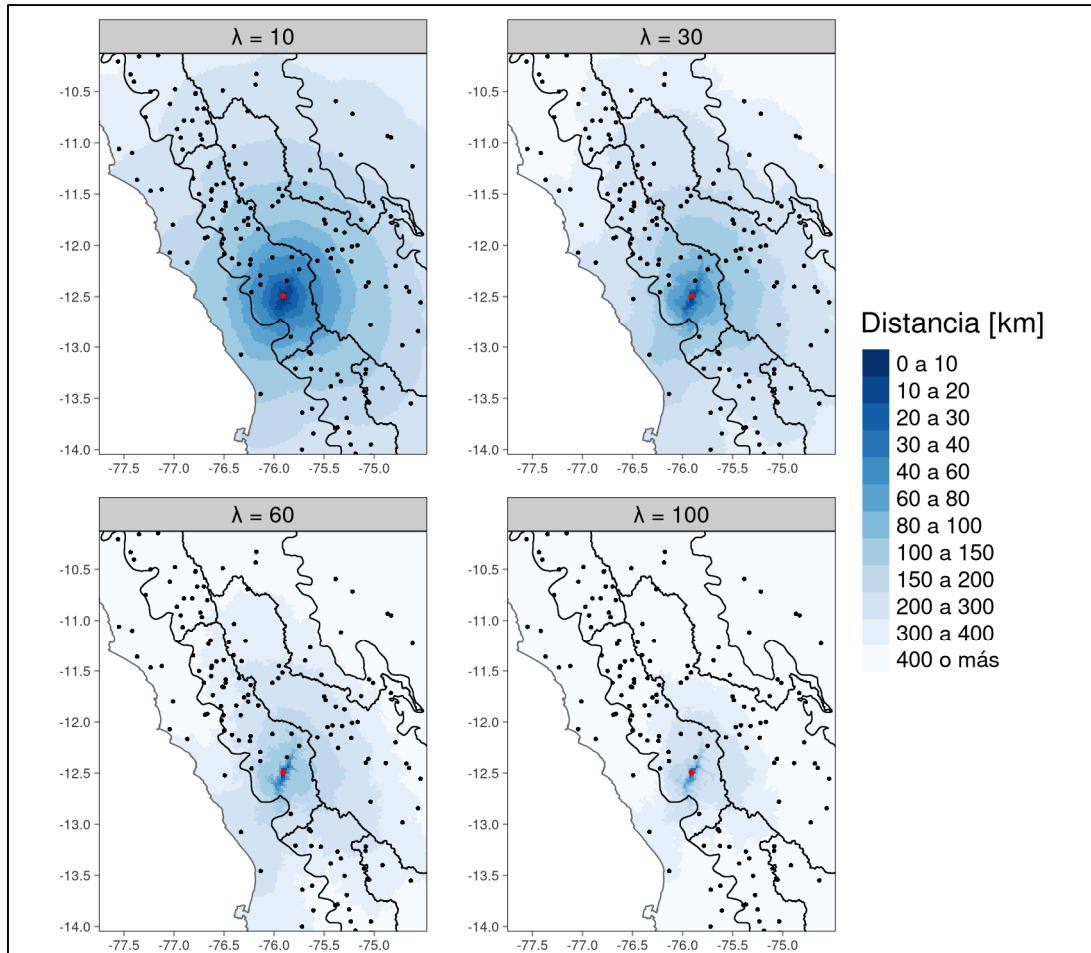


Figura N° 25. Distancia no euclídea 3 (NE 3) desde estación en valle (Yauyos)

Fuente: Elaboración propia

En la Figura N° 25 y Figura N° 26 se muestra el cálculo de esta distancia desde una estación en valle y una montañosa respectivamente. Esta distancia presenta magnitudes mayores a las anteriores debido a que los desplazamientos y conexiones solo están permitidos siguiendo la superficie topográfica. Además, como consecuencia de ello, las superficies con un mismo rango de distancia son continuas, no se presentan saltos, que permitían conectar zonas de elevaciones similares separadas en las aproximaciones anteriores.

Si bien es cierto que las distancias aquí mostradas resultan mucho mayores, esto no es necesariamente malo, dado que los pesos de las distintas estaciones sobre

una ubicación específica son dependientes de la relación relativa de magnitud entre las distancias desde estas estaciones.

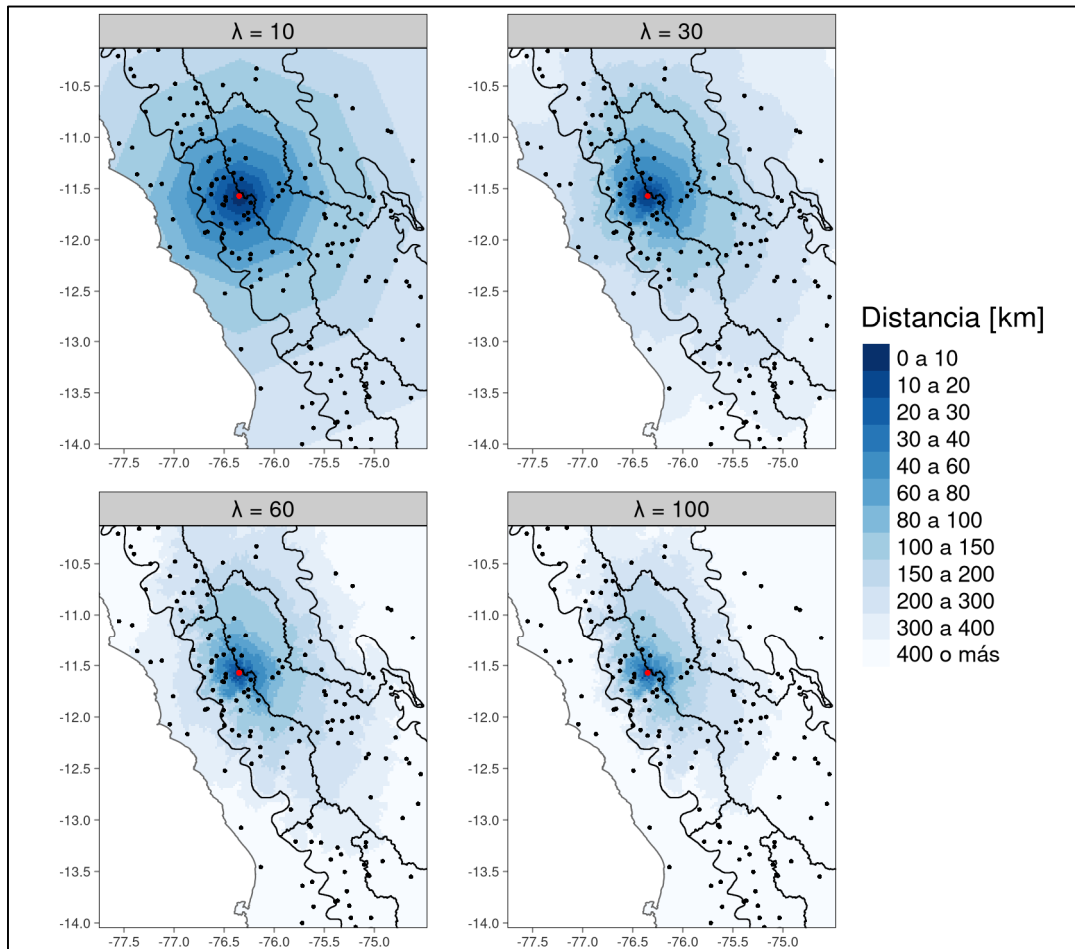


Figura N° 26. Distancia no euclídea 3 (NE 3) desde estación montañosa (Milloc)

Fuente: Elaboración propia

4.3 PROCESO DE INTERPOLACIÓN

El proceso que se detalla se realiza sobre cada una de las zonas de estudio. Además, el cálculo, para cada λ de interés, de la matriz de distancias entre las estaciones y todas las ubicaciones a interpolar (por ejemplo, las celdas de un raster), es un paso previo para el proceso de interpolación.

Primeramente, se asume que existe una tendencia subyacente al conjunto de datos, que puede explicar en cierta medida su comportamiento. Esta tendencia es representada mediante un modelo lineal dependiente de la elevación, y para estimar sus parámetros se realiza una regresión lineal. Es decir, se tiene que:

$$P_L(z) = \beta_0 + \alpha z \quad (4-1)$$

Donde $P_L(z)$ representa la precipitación que depende de la elevación z y los parámetros a ajustar son β_0 y α .

Con el fin de representar con mayor precisión las condiciones locales de diferentes ubicaciones, se realiza una interpolación de residuales, que sigue un esquema similar al de IDW. Los residuales en cada estación R_i son calculados como la diferencia entre los valores observados y los valores estimados de la tendencia subyacente (4-1). Se puede expresar el cálculo del residual sobre una ubicación a estimar $R(s_0)$ mediante:

$$R(s_0) = \frac{1}{\sum w_i} \sum w_i R_i \quad (4-2)$$

$$\text{con } w_i = 1/(d_{\lambda,i})^p$$

Donde w_i representa el peso de cada estación, $d_{\lambda,i}$ es la distancia no euclídea desde la estación i calculada con el coeficiente λ y p es la potencia a la que se eleva. La forma de seleccionar λ para cada mes se describirá más adelante. La potencia p se calibra de forma global, considerando como posibles valores 1, 1.5, 2.0 y 2.5. Además, las estaciones usadas sobre la ubicación s_0 son las que se encuentran cercanas, en particular, se toman las que están en un radio de distancia convencional de 150 km. Este radio será también usado para la comparación con otros métodos basados en IDW.

En base a estos dos procesos se calculan los campos $P(s)$ y $R(s)$ sobre todas las ubicaciones a estimar (por ejemplo, las celdas de un raster), lo cual se ve aplicado en la Figura N° 27. Es importante notar que valores negativos no son permitidos para la precipitación, cualquier valor negativo resultante es transformado en cero. Finalmente, y considerando esto, la estimación resultante estará dada por la suma de ambas variables:

$$\hat{P}(s_0) = P_L(s_0) + R(s_0) \quad (4-3)$$

4.3.1.1 Selección de λ

Al igual que el proceso de regresión lineal, el coeficiente λ es calculado de forma independiente para cada paso de tiempo. Esto permite una mejor representación de las condiciones climáticas particulares de cada mes. Como se observa en la Figura N° 27 se calcula el RMSE para una secuencia de λ entre 0 y 300. Este RMSE, que representa una medida del error, se calcula sobre el conjunto de valores observados y valores estimados en las estaciones en el mes. Para calcular el valor estimado sobre una estación se usa solo el resto de las estaciones. El λ que presenta el menor RMSE, es seleccionado como el representativo del mes.

4.3.1.2 Ejemplos

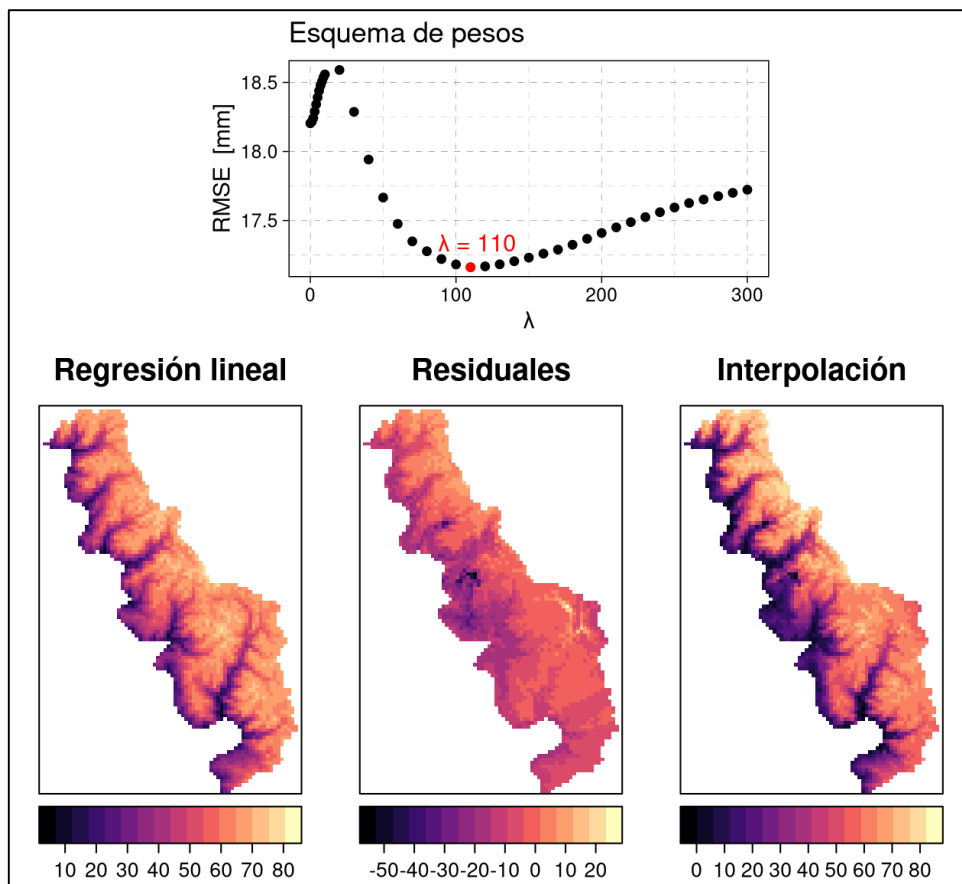


Figura N° 27. Proceso de interpolación en zona 3 usando distancias NE 2 (1982-12)

Fuente: Elaboración propia

Los procesos descritos previamente son representados en su conjunto en la Figura N° 27, donde se representa un mes de temporada húmeda. En este caso particular el modelo lineal representa una mayor precipitación a mayor altura y es

muy cercano al resultado final. Asimismo, se optimiza un λ de magnitud media lo que se refleja en el campo de residuales. Esto contrasta con el campo de residuales de la Figura N° 28 que se asemeja más a una interpolación convencional IDW como consecuencia del bajo λ calibrado. Por otro lado, el comportamiento del campo de residuales de la Figura N° 29 es opuesto, dado que se observa una mayor estratificación que está relacionada a su mayor λ .

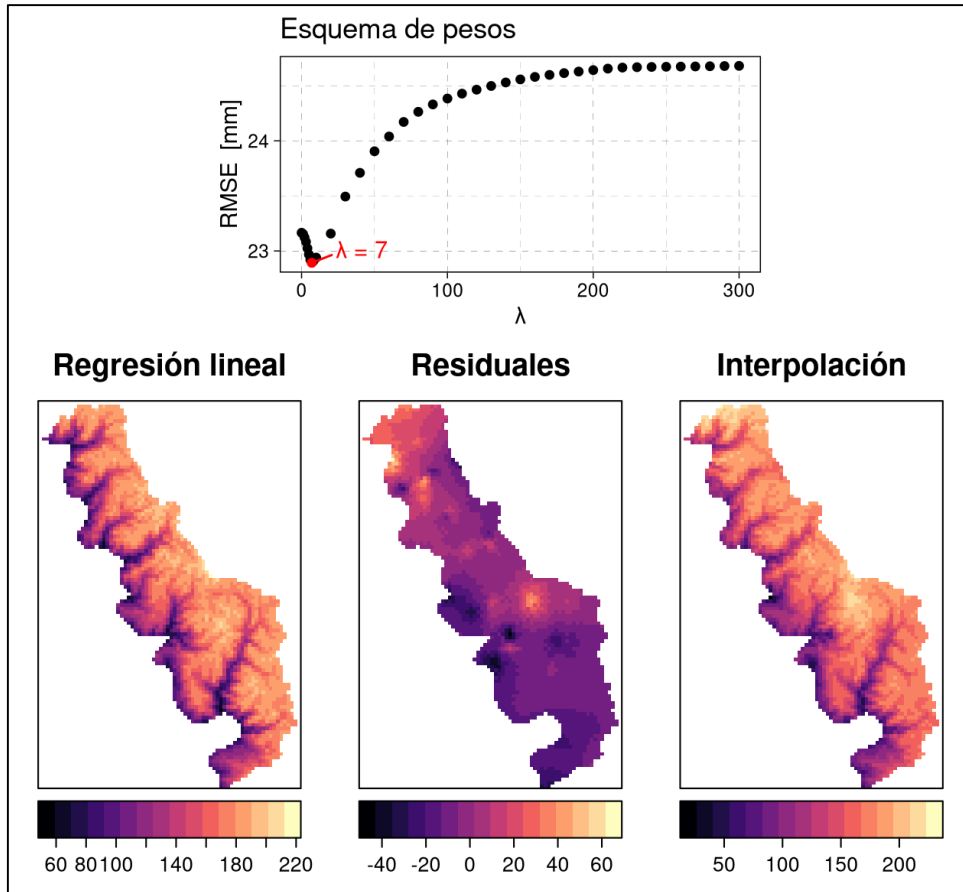


Figura N° 28. Proceso de interpolación en zona 3 usando distancias NE 2 (2009-01)

Fuente: Elaboración propia

La Figura N° 28 representa también un mes de época húmeda donde el modelo lineal presenta un gradiente mayor al caso anterior. Asimismo, considerando las bajas magnitudes de los residuales en relación con los otros subprocesos, el modelo lineal parece explicar bastante bien la precipitación de este mes en particular. Finalmente, en la Figura N° 29 se presenta, a diferencia de las anteriores, un mes de época seca. En este mes en particular, el λ se hace máximo.

Además, se observa la estratificación característica, que se ve especialmente marcada en la parte norte.

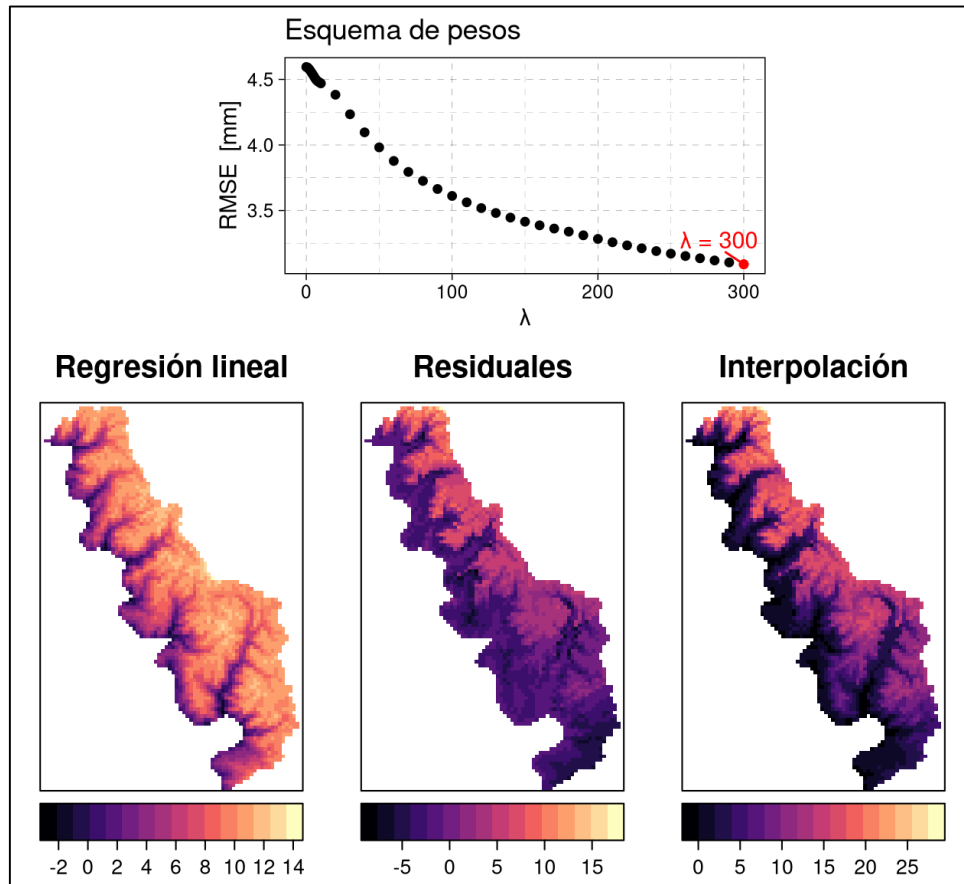


Figura N° 29. Proceso de interpolación en zona 3 usando distancias NE 2 (1980-08)

Fuente: Elaboración propia

4.3.1.3 Similitud con otras aproximaciones

La metodología de interpolación previamente descrita sigue la propuesta por Frei (2014), sin embargo, se han modificado y simplificado ciertos procesos. El modelo o tendencia subyacente originalmente correspondía originalmente a un perfil no lineal con más parámetros que respondía a las características particulares de la distribución de temperatura en Suiza, en cambio, en el presente trabajo se ha usado un modelo lineal con solo dos parámetros. Esta selección responde al objetivo de dejar el proceso lo más sencillo y con menos parámetros posible, dado que la densidad de estaciones que se usarían para calibrarlos es menor. Sin embargo, si se usara un área más amplia con tendencias marcadas, sería

conveniente poder usar un modelo con dos pendientes que permitiría representar mejor el comportamiento global de la precipitación.

Por otro lado, como se ha mencionado anteriormente, el uso de un $\lambda = 0$ es equivalente al uso de distancias convencionales. En este caso particular se ha notado que el procedimiento propuesto es similar al método IED (IDW with External Drift) descrito por Ruelland (2020). Este método usa la interpolación IDW de residuales sobre un modelo lineal de tendencia. Dada esta similitud y que también se han observado buenos resultados, se usa también este método para la comparación en las siguientes secciones.

4.4 EVALUACIÓN DE PERFORMANCE

4.4.1 Validación cruzada

Con la finalidad de cuantificar el error y la performance del método propuesto y sus variantes con cada una de las distancias no euclídeas, se usa un proceso de validación cruzada. Como se detalla en la sección 2.1.3, en este proceso se interpola sobre la ubicación de una estación, considerando solo el resto de las estaciones. Repitiendo este proceso para cada estación, se obtienen series simuladas que son comparadas con las observadas por medio de índices de desempeño.

Para la presente aplicación, se usarán la raíz del error medio cuadrático ($RMSE$), la eficiencia de Nash-Sutcliffe ($Nash$) y el coeficiente de correlación (r). Utilizar estos tres índices permite evaluar diferentes aspectos del desempeño de los métodos: el $RMSE$ permitirá obtener una estimación del error en unidades de precipitación, el r es una medida de la correlación lineal y el $Nash$ compara la relación de la varianza residual con la observada y permite cuantificar la precisión en una escala única donde 1 es su valor máximo y óptimo.

En ese sentido, una mejor performance representará un menor $RMSE$ y unos mayores valores de $Nash$ y r .

4.4.2 Comparación con otros métodos

Los resultados de la validación cruzada sirven para comparar distintos métodos de interpolación. Además de los resultados del método descrito considerando las

distintas propuestas de distancia, también se comparan estos con otros métodos de distintas características.

Entre los métodos clásicos a comparar están el de inverso de la distancia (IDW), kriging ordinario (OK) y kriging con deriva externa (KED). Los dos últimos son métodos geoestadísticos y el último considera como variable auxiliar la elevación.

Asimismo, son objeto de comparación otros métodos determinísticos basados en el de IDW que usan como dato la elevación. En particular se usan el método de IDW con corrección de gradiente altitudinal (Valery) propuesto por Valéry et al. (2010) y el método de IDW con deriva externa (IED) descrito por Ruelland (2020). Este último método, como se detalló anteriormente, es de interés dado que representa un caso particular ($\lambda = 0$) del método en estudio.

Cada uno de los métodos a evaluar tiene parámetros propios que son ajustados a los datos de entrada. En forma particular, la potencia p de la distancia, presente en los métodos basados en IDW, puede tomar los valores 1, 1.5, 2 o 2.5 y se calibra en base al *RMSE* como un parámetro global que no depende del mes.

Es importante recalcar que siguiendo el procedimiento de validación cruzada la interpolación sobre la ubicación de cada estación es un proceso independiente. Por ejemplo, en el caso del método KED, los variogramas son calculados independientemente para la interpolación sobre la ubicación de cada estación (Figura N° 30). Se observa que para los casos de validación cruzada (a), donde se elimina una estación, los parámetros del variograma pueden llegar a variar con respecto al caso (b) donde no se elimina ninguna.

Finalmente, dado el alto costo computacional que representan algunos procesos, en particular el cálculo de la distancia NE2, se vio por conveniente realizar el preprocesamiento de las listas de conexiones para cada zona e implementar los procesos de interpolación espacial, algoritmos de validación cruzada y comparación de metodologías con técnicas de computación paralela que permitieron obtener resultados y hacer varios ensayos de forma rápida en el clúster del Laboratorio Nacional de Hidráulica.

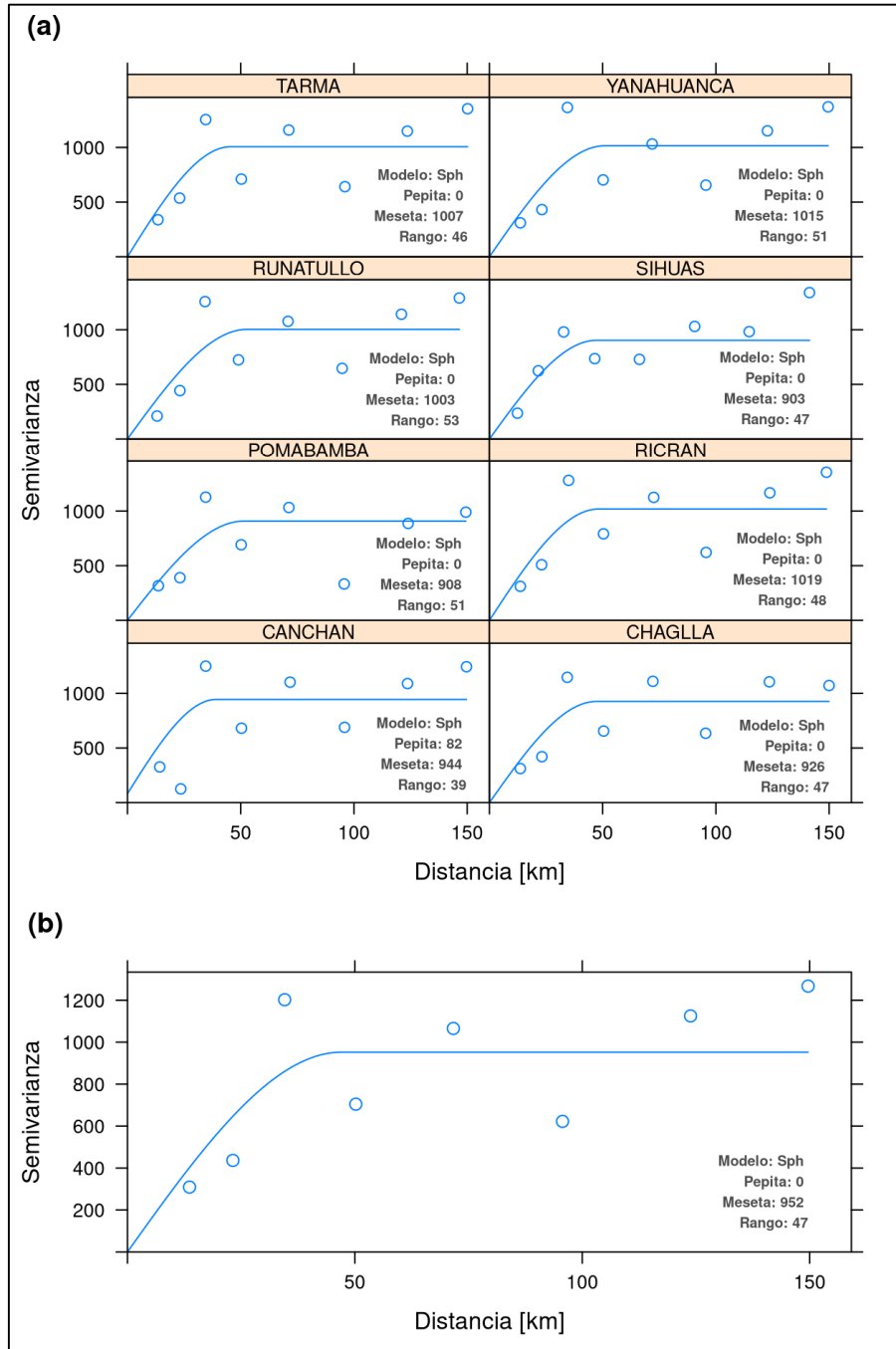


Figura N° 30. Cálculo de variogramas en noviembre de 2009 para KED en zona 8: en base a validación cruzada para algunas estaciones (a) y considerando todos los datos (b)

Fuente: Elaboración propia

4.5 UNA APLICACIÓN EN MODELAMIENTO HIDROLÓGICO

Adicionalmente, se han aplicado y comparado los distintos métodos de interpolación de precipitación en el modelamiento hidrológico de algunas cuencas específicas. Estas cuencas, mostradas en la Figura N° 31, corresponden a la del Caplina y Obrajillo. La cuenca Caplina posee una superficie de 536.1 km² y cuenta con la estación hidrométrica Calientes (periodo 2005-2017). Asimismo, la cuenca Obrajillo tiene un área de 363.5 km² y fue delimitada a partir de la estación hidrométrica del mismo nombre (periodo 1997-2017).

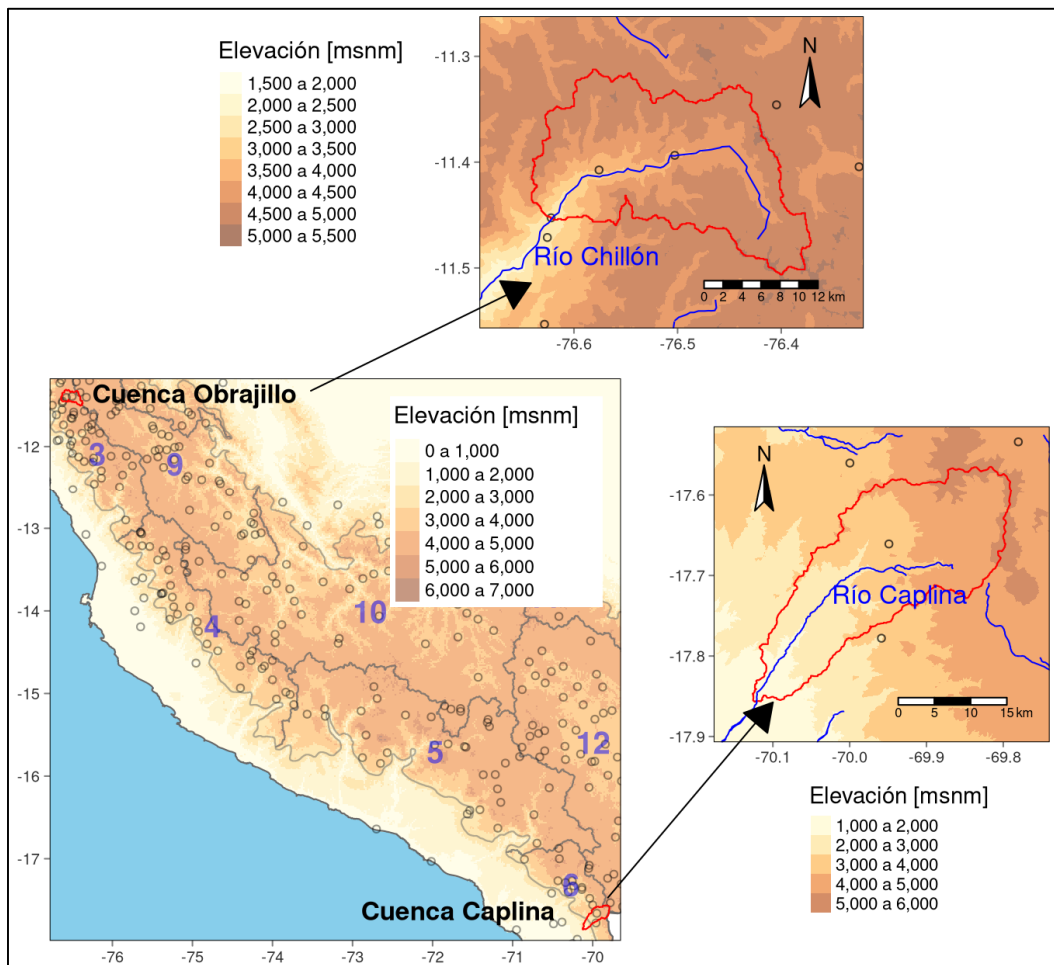


Figura N° 31. Localización de cuencas de estudio

Fuente: Elaboración propia

El modelo hidrológico que se va a usar es el GR2M, que es un modelo agregado y parsimonioso, que cuenta con 2 parámetros (X_1 y X_2) para calibrar. Adicionalmente a los datos de precipitación, el modelo requiere la

evapotranspiración potencial (*ETP*), que se puede calcular en base a la temperatura media. Asimismo, el paquete airGR desarrollado por Coron et al. (2017), en el lenguaje R, fue usado para calibrar el modelo y obtener las series simuladas.

Si bien es cierto se puede usar el método propuesto con distancias no euclídeas para interpolar los datos de las estaciones de temperatura, como fue usado en investigaciones anteriores, se ha seleccionado el método KED por ser de amplio uso. Esta interpolación es realizada resolución espacial de 0.02° , la misma que la de la precipitación:

Con los datos grillados de temperatura es posible calcular la evapotranspiración potencial, como se muestra en la Figura N° 32 para un mes de época seca y uno de época húmeda. La evapotranspiración potencial es calculada mediante el método de Oudin et al. (2005) que fue desarrollado especialmente para aplicaciones hidrológicas. Este método depende de la temperatura media diaria T (en $^\circ\text{C}$) y de la radiación solar extraterrestre R_e (en $\text{MJ}/\text{m}^2/\text{d}$), que puede ser estimada a partir de la latitud y el número de día juliano, y se expresa como:

$$\begin{cases} ETP = \frac{R_e T + 5}{\lambda \rho 100}, & \text{si } T + 5 > 0 \\ ETP = 0 & \text{, en caso contrario} \end{cases} \quad (4-4)$$

Donde *ETP* es la evapotranspiración potencial (mm/d), λ representa el flujo de calor latente (2.45 MJ/kg) y ρ es la densidad del agua (kg/m^3)

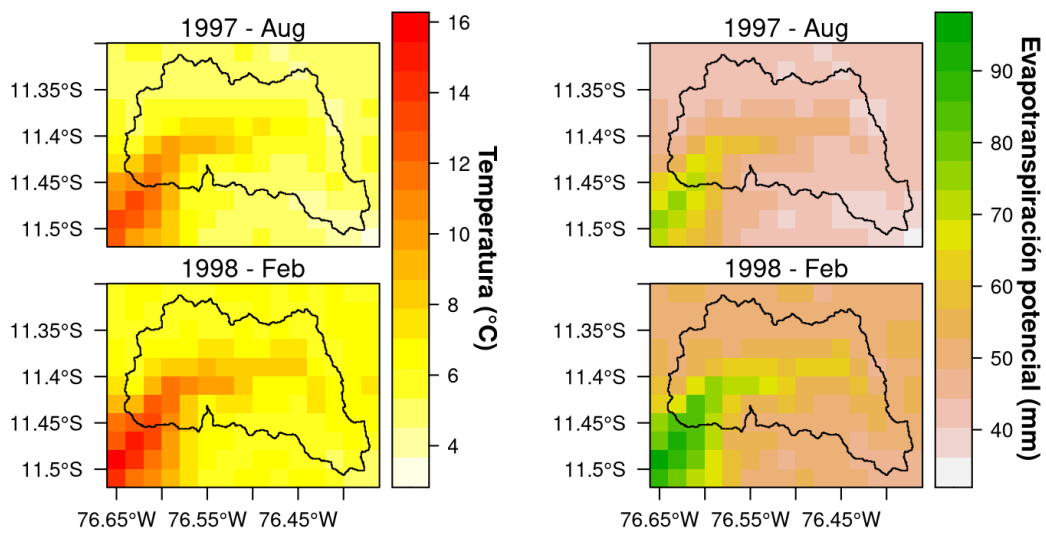


Figura N° 32. Interpolación de temperatura y cálculo de evapotranspiración potencial - cuenca Obrajillo

Fuente: Elaboración propia

Contando con los datos grillados, se extraen las series de tiempo de *ETP* y de precipitación medias para cada cuenca. Finalmente, para cada caso, se divide la serie de caudales observados en un periodo para la calibración y en otro que será usado de forma independiente para la validación. Adicionalmente se considera un periodo previo usado como etapa de calentamiento del modelo, con el fin de que los volúmenes de reservorios y otros procesos internos del modelo se establezcan antes del periodo de calibración. Estos periodos se detallan en la Tabla N° 2.

Tabla N° 2. Periodos de calentamiento, calibración y validación

Fuente: Elaboración propia

Periodo	Cuenca Caplina	Cuenca Obrajillo
Calentamiento	2000-2004	1994-1996
Calibración	2005-2011	1997-2009
Validación	2012-2017	2008-2017

El índice de *Nash* es usado para calibrar y evaluar la performance en la validación. Este proceso se realiza considerando todas las posibles series de precipitación que se obtienen de los métodos de interpolación a comparar.

CAPITULO V: ANÁLISIS Y DISCUSIÓN DE RESULTADOS

5.1 EVALUACIÓN DE DESEMPEÑO

El método de interpolación propuesto, considerando las diferentes aproximaciones de distancias no euclídeas, ha sido evaluado por medio de la validación cruzada. En la Tabla N° 3 se presentan los índices de desempeño para las zonas de la vertiente del Pacífico. Estos son calculados como la media de los índices de cada estación en la zona. Una primera comparación muestra que el uso de distancias no euclídeas (casos NE1, NE2 y NE3 descritos en la sección 4.2) muestra una significativa mejora con respecto a la aplicación del mismo método, pero sin considerar estas distancias (caso IED).

Tabla N° 3. Índices de desempeño medios para interpolación propuesta considerando opciones de distancias no euclídeas – Vertiente del Pacífico (zonas de la 1 a la 6)

Fuente: Elaboración propia

Zona	Índice	IED	NE1	NE2	NE3
1	RMSE	47.11	46.54	46.04	46.47
	Nash	0.65	0.66	0.66	0.66
	r	0.87	0.87	0.87	0.87
2	RMSE	36.08	34.82	34.21	34.95
	Nash	0.39	0.43	0.46	0.43
	r	0.85	0.86	0.87	0.86
3	RMSE	25.62	25.13	24.96	25.41
	Nash	0.77	0.78	0.79	0.78
	r	0.90	0.91	0.91	0.91
4	RMSE	29.83	29.25	29.04	29.45
	Nash	0.66	0.67	0.68	0.68
	r	0.89	0.90	0.90	0.89
5	RMSE	25.12	24.54	24.55	24.78
	Nash	0.75	0.76	0.76	0.75
	r	0.90	0.91	0.91	0.91
6	RMSE	16.16	15.09	15.25	15.68
	Nash	0.58	0.63	0.63	0.61
	r	0.91	0.91	0.92	0.91

Como se observa, en la mayoría de las zonas el caso NE2 muestra una mejor performance. Esto no es tan evidente en las zonas ubicadas más al sur, específicamente las zonas 5 y 6. En la zona 5 el caso NE1 muestra ligeramente una mejor performance en el *RMSE* y similar al NE2 en los otros índices; mientras que en la zona 6 el caso NE1 tiene un mejor *RMSE*, pero se ve superado por el caso NE2 en el *r*. Por otro lado, en el resto de las zonas (de la 1 a la 4) el caso NE2 muestra los menores *RMSE* y los mejores *Nash* y *r*, estos dos últimos junto con los casos NE1 y NE3 en ciertas zonas.

Tabla N° 4. Índices de desempeño medios para interpolación propuesta considerando opciones de distancias no euclídeas – Vertiente del Atlántico y Titicaca (zonas de la 7 a la 12)

Fuente: Elaboración propia

Zona	Índice	IED	NE1	NE2	NE3
7	RMSE	34.86	32.36	32.36	32.65
	Nash	0.67	0.72	0.72	0.72
	r	0.88	0.89	0.89	0.88
8	RMSE	35.25	33.31	32.88	34.47
	Nash	0.52	0.56	0.57	0.54
	r	0.84	0.85	0.85	0.84
9	RMSE	32.93	32.13	32.08	32.44
	Nash	0.66	0.68	0.68	0.67
	r	0.87	0.87	0.87	0.87
10	RMSE	34.68	34.16	34.11	34.17
	Nash	0.73	0.74	0.74	0.74
	r	0.90	0.90	0.90	0.90
11	RMSE	53.31	47.94	45.11	50.21
	Nash	0.34	0.43	0.53	0.42
	r	0.76	0.80	0.81	0.77
12	RMSE	27.46	27.08	26.99	27.12
	Nash	0.81	0.82	0.82	0.82
	r	0.92	0.93	0.93	0.92

De forma similar, la Tabla N° 4 muestra la performance del uso de las tres distancias no euclídeas propuestas y del caso donde no se usan. Es así como se aprecia una significativa mejora con el uso de distancias no euclídeas frente al caso IED de distancias convencionales. Esta mejora es particularmente significativa en las zonas 7,8 y 11, y puede deberse a que comparten en común

una topografía con abruptos cambios de elevación, donde los cañones del Marañón les confieren esta característica a las primeras dos zonas y la parte oriental de la cordillera de Carabaya que desciende hasta la amazonía a la tercera.

Asimismo, se observa que en la mayoría de las zonas el caso NE2 muestra la mejor performance, a excepción de la zona 7 donde el caso NE1 muestra unos índices equivalentes. Es importante notar también que el caso NE1 muestra unos mejores desempeños al caso NE3 y en la mayoría de zonas cercanos al caso NE2, siendo así una posible alternativa al este último dado su menor costo computacional.

En la Figura N° 33 se muestra la distribución de índices de desempeño para el caso NE2, que ha sido el que mejores resultados mostró. Se puede notar un menor *RMSE* en la zona 6, como era de esperarse por ser la menos lluviosa. Además, se observa una amplia incertidumbre en términos de *r* para la zona 11, que puede estar relacionado a tener magnitudes de precipitación bastante elevadas en estaciones cercanas. Información similar para los otros casos, se muestra en la Figura N° 34, Figura N° 35 y Figura N° 36.

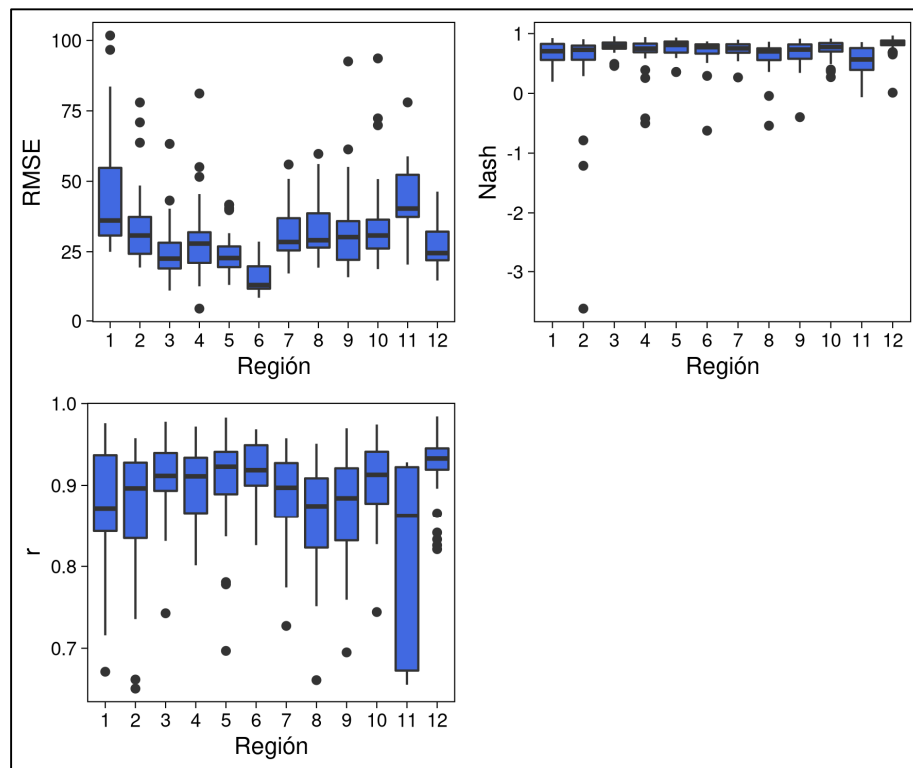


Figura N° 33. Diagrama de cajas de distribución de índices de desempeño por zonas.

Fuente: Elaboración propia

5.2 COMPARACIÓN CON OTROS MÉTODOS

En la Tabla N° 5 y en la Tabla N° 6 se muestran la performance media del caso NE2, que ha mostrado los mejores resultados en el anterior análisis, en comparación a las de otros métodos de interpolación.

De la Tabla N° 5 se observa que en general el método NE2 muestra una mejor performance que los otros métodos, esto es particularmente claro en el caso de la zona 2. Además, se puede notar que en la zona 1,4 y 6 el método de Valery muestra un mejor *Nash* que el NE2, mientras en todas las demás zonas e índices el método NE2 siempre obtiene los mejores indicadores.

Es también importante resaltar que los métodos que consideran la elevación muestran mejores índices que los métodos que no lo hacen, esto es particularmente significativo, por ejemplo, en los *Nash* de las zonas 2, 4 y 6.

Tabla N° 5. Índices de desempeño medios para métodos de interpolación de comparación –
Vertiente del Pacífico (zonas de la 1 a la 6)
Fuente: Elaboración propia

Zona	Índice	NE2	IDW	Valery	OK	KED
1	RMSE	46.04	47.63	46.40	48.57	48.91
	Nash	0.66	0.65	0.67	0.63	0.61
	r	0.87	0.86	0.87	0.85	0.86
2	RMSE	34.21	38.52	36.83	39.01	37.10
	Nash	0.46	0.12	0.35	0.10	0.37
	r	0.87	0.85	0.85	0.84	0.84
3	RMSE	24.96	29.60	26.25	28.48	26.38
	Nash	0.79	0.69	0.77	0.71	0.76
	r	0.91	0.90	0.90	0.89	0.90
4	RMSE	29.04	35.97	30.17	37.10	30.89
	Nash	0.68	0.37	0.69	0.28	0.62
	r	0.90	0.88	0.89	0.87	0.88
5	RMSE	24.55	25.82	25.22	25.70	25.64
	Nash	0.76	0.73	0.75	0.74	0.75
	r	0.91	0.90	0.90	0.89	0.90
6	RMSE	15.25	19.59	16.09	19.52	17.37
	Nash	0.63	0.06	0.65	0.22	0.50
	r	0.92	0.90	0.91	0.88	0.89

De la Tabla N° 6, se observa que el método NE2 muestra mejores índices en las zonas 7, 8, 9 y 12; donde solo se ve superado en el *Nash* de la zona 8 y 12 por el método de Valery e IDW respectivamente. Sin embargo, un comportamiento diferente es el que se observa en las zonas 10 y 11, donde los métodos de IDW y Valery muestran los mejores resultados respectivamente. Es particularmente clara esta diferencia en la zona de 11 donde el método de Valery supera ampliamente a los demás en cuanto al *RMSE* y *Nash*, que podría deberse a sus condiciones climáticas especiales debido a que se encuentra muy próxima a la localidad de Quincemil en la selva alta que es considerada la más lluviosa del país (MIDAGRI, s. f.).

De la misma forma al anterior grupo de zonas se presenta una mejora significativa en los métodos que si consideran la elevación en la mayoría de las zonas. Sin embargo, este no es el caso de las zonas 10 y 12, donde parece hasta cierto punto contraproducente.

Tabla N° 6. Índices de desempeño medios para métodos de interpolación de comparación –
Vertiente del Atlántico y Titicaca (zonas de la 7 a la 12)

Fuente: Elaboración propia

Zona	Índice	NE2	IDW	Valery	OK	KED
7	RMSE	32.36	36.28	34.99	36.95	34.84
	Nash	0.72	0.65	0.67	0.65	0.69
	r	0.89	0.88	0.88	0.87	0.87
8	RMSE	32.88	35.52	33.80	36.67	35.94
	Nash	0.57	0.52	0.59	0.47	0.51
	r	0.85	0.85	0.85	0.83	0.82
9	RMSE	32.08	33.42	33.04	34.31	33.74
	Nash	0.68	0.65	0.66	0.62	0.64
	r	0.87	0.87	0.87	0.86	0.86
10	RMSE	34.11	33.67	33.73	35.44	35.82
	Nash	0.74	0.74	0.74	0.71	0.70
	r	0.90	0.91	0.90	0.89	0.89
11	RMSE	45.11	43.20	38.18	53.07	50.95
	Nash	0.53	0.58	0.67	0.38	0.42
	r	0.81	0.85	0.87	0.81	0.79
12	RMSE	26.99	27.19	27.39	27.98	27.80
	Nash	0.82	0.83	0.82	0.82	0.81
	r	0.93	0.92	0.92	0.91	0.92

En la Figura N° 34, Figura N° 35 y Figura N° 36 se presentan diagramas de caja de los índices por zona y método, que, a diferencia de la media de los índices previamente usada, permiten visualizar además la distribución de estos índices que representa información interesante para comparar los distintos métodos.

Se puede notar que en varias zonas 2, 4 y 6 los métodos que consideran la elevación permiten explicar mejor las estaciones que generan *Nash* excesivamente bajos, es decir, los menores *Nash* tienden a alejarse de los valores mínimos extremos. Esto es particularmente notorio en la zona 4, donde los métodos IDW y OK muestran desempeños muy pobres en comparación a los otros, y el método de Valery muestra mayores mínimos. Sucede lo opuesto a lo anterior en la zona 12, donde los métodos IDW y OK parecen explicar mejor estos extremos en términos de *Nash*.

Con respecto a los otros índices (*RMSE* y *r*) las distribuciones tienden a ser más uniformes y presentar menos extremos. El coeficiente de correlación *r*, por ejemplo, presenta en todos los casos valores altos y muchas veces menos puntos atípicos en comparación a los otros índices.

Los casos que usan distancias no euclídeas muestran en general similitud en sus distribuciones, y también con el método IED como era de esperarse dada su estructura base similar. En la zona 11, sin embargo, todos los métodos tienden a presentar performances muy variables entre ellos. En esta zona el método de Valery presenta una distribución que elimina extremos presentes en otros métodos y una mejor mediana.

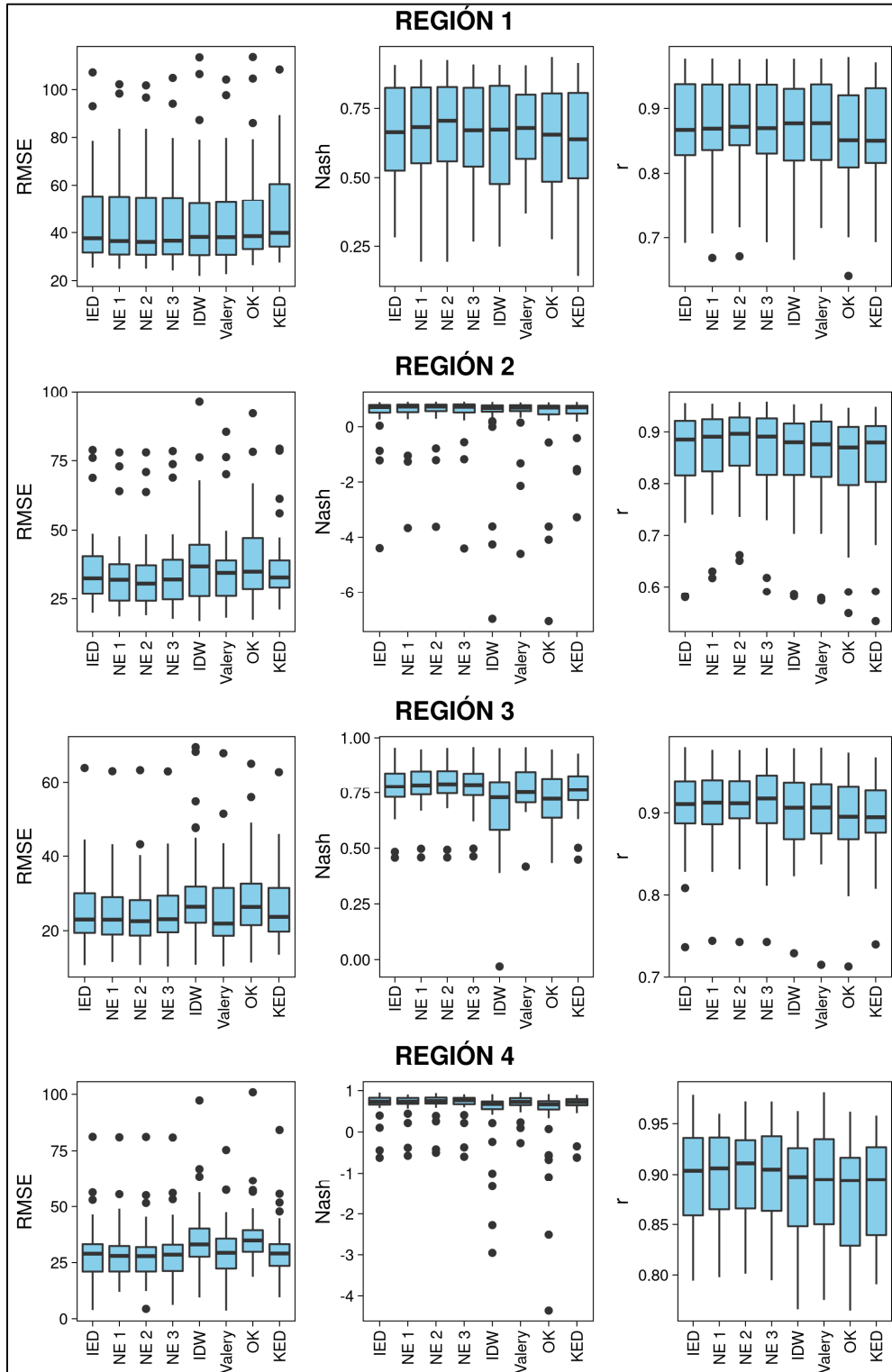


Figura N° 34. Diagrama de cajas de índices de desempeño para los métodos de interpolación evaluados (zonas de la 1 a la 4)

Fuente: Elaboración propia

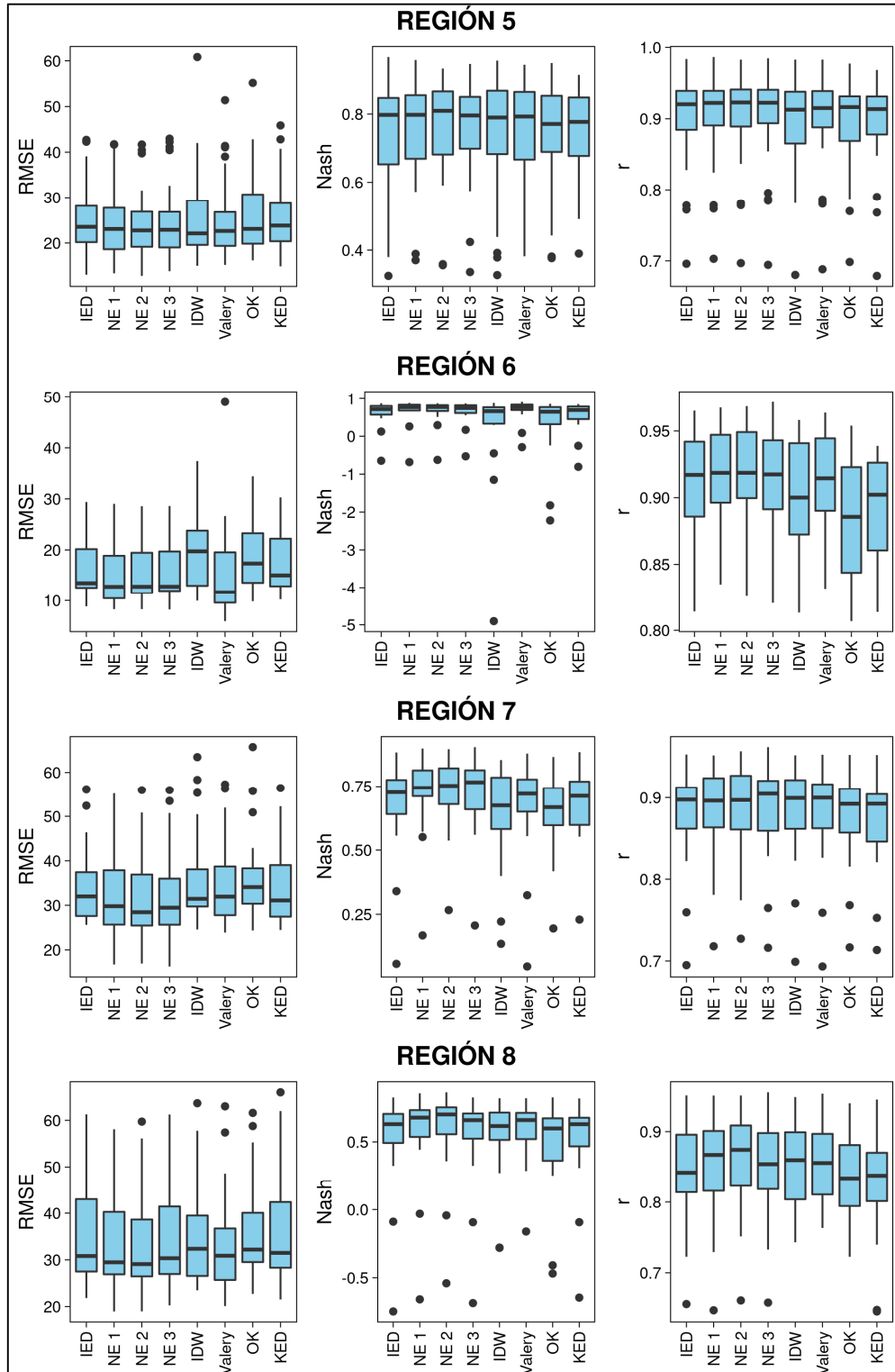


Figura N° 35. Diagrama de cajas de índices de desempeño para los métodos de interpolación evaluados (zonas de la 5 a la 8)

Fuente: Elaboración propia

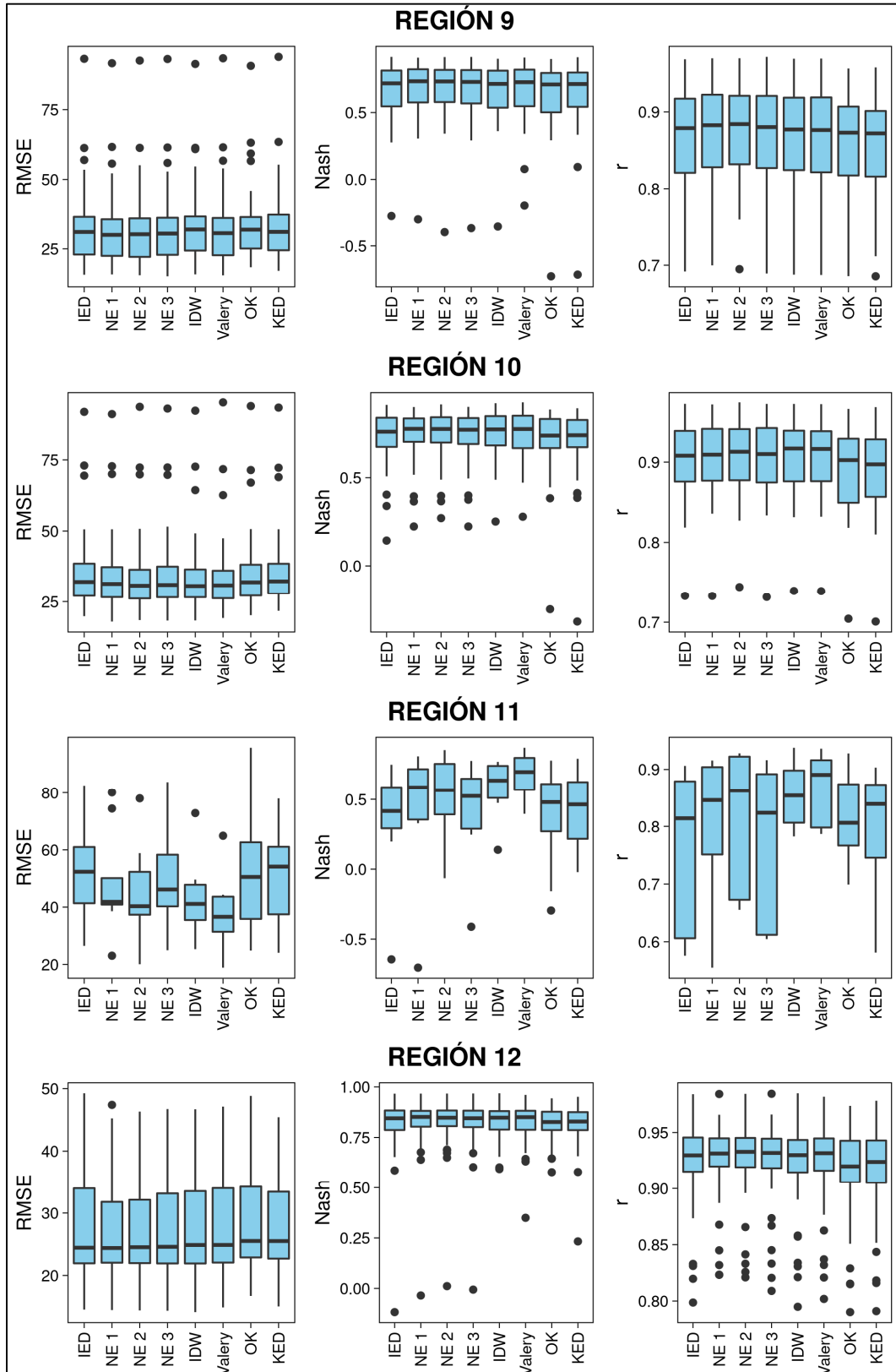


Figura N° 36. Diagrama de cajas de índices de desempeño para los métodos de interpolación evaluados (zonas de la 9 a la 12)

Fuente: Elaboración propia

5.3 EJEMPLOS

Se considera para la presentación gráfica de los mapas de ejemplo solo el caso NE2 de entre las opciones de distancias no euclídeas, por haber mostrado mejores resultados (de ahora en adelante nombrado solo como NE), y se lo compara con los otros métodos de interpolación con los que se comparó.

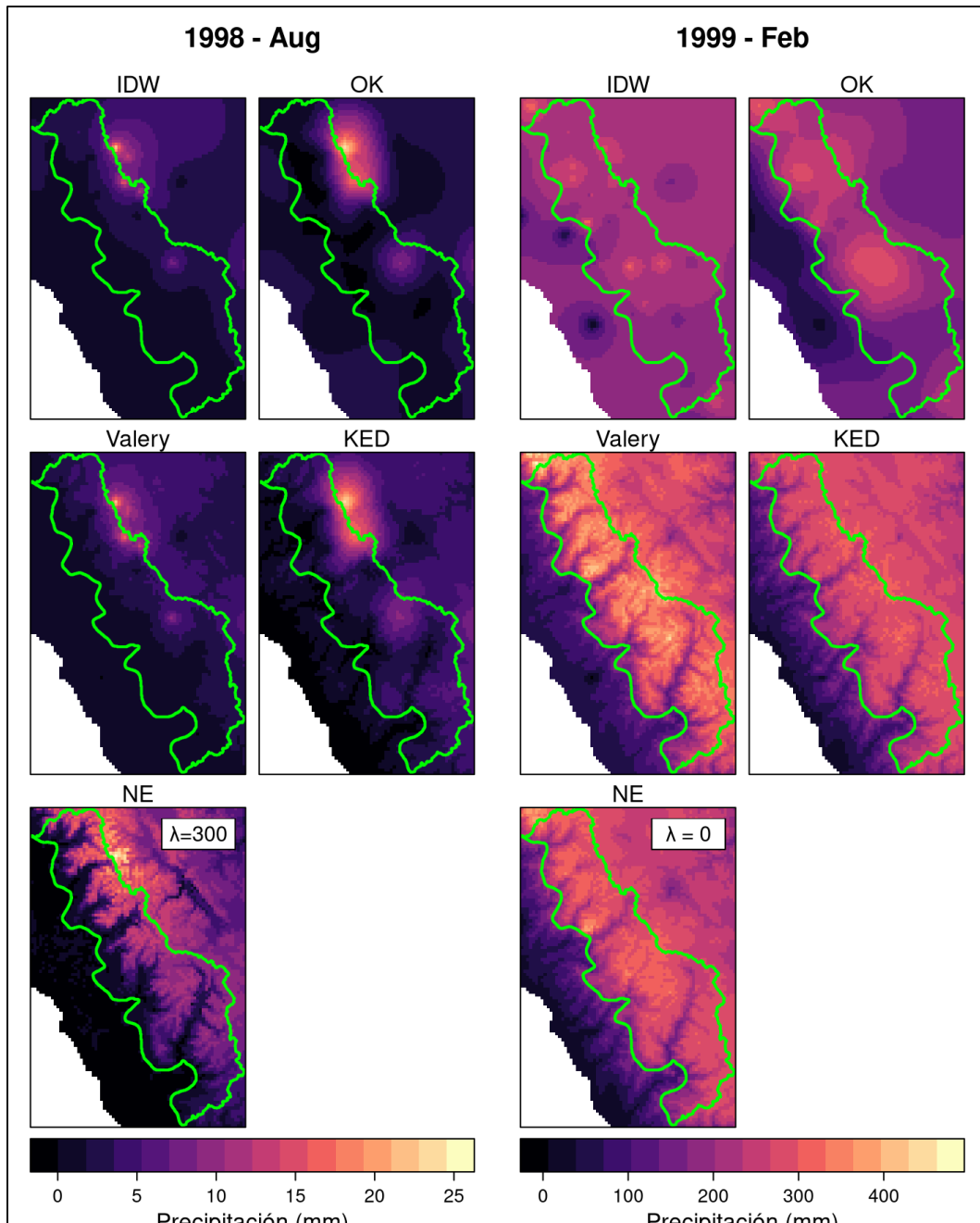


Figura N° 37. Mapas de precipitación obtenidos de métodos de interpolación en comparación en agosto de 1998 y febrero de 1999 – Zona 3

Fuente: Elaboración propia

En la Figura N° 37 y Figura N° 38 se presentan mapas de precipitación para meses de época húmeda y seca en la zona 3. Se observa una diferencia marcada entre los métodos que consideran la elevación y los que no; dado que en los primeros se marca la topografía, especialmente de valles, en la distribución espacial de precipitación.

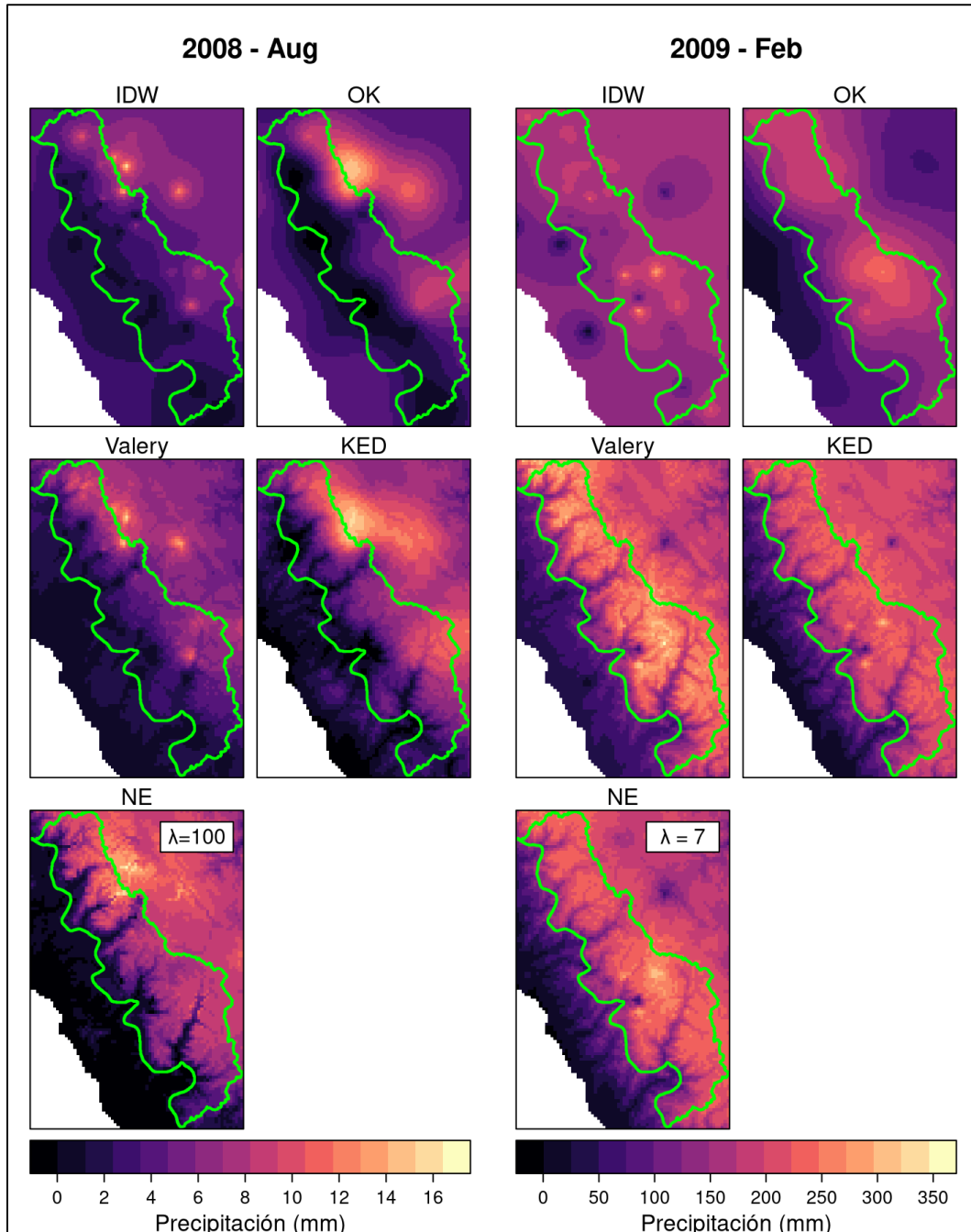


Figura N° 38. Mapas de precipitación obtenidos de métodos de interpolación en comparación en agosto de 2008 y febrero de 2009 – Zona 3

Fuente: Elaboración propia

Asimismo, en el método NE, se muestra una mayor estratificación a mayores λ , mientras que en caso de menores valores la distribución se vuelve más similar al método KED. Para los meses lluviosos se observa que el método de Valery tiende a mostrar mayores precipitaciones que los otros métodos, esto puede deberse a que el parámetro de gradiente altitudinal θ es calculado de manera global para la zona y por ello no se adapta necesariamente a las condiciones particulares de ciertos meses. Sin embargo, esta mayor estimación no se presenta en los meses de época seca mostrados. Además, se puede notar en los meses de época húmeda presentados, una ligera mayor estimación con el método NE que con el KED, esto especialmente en partes localizadas.

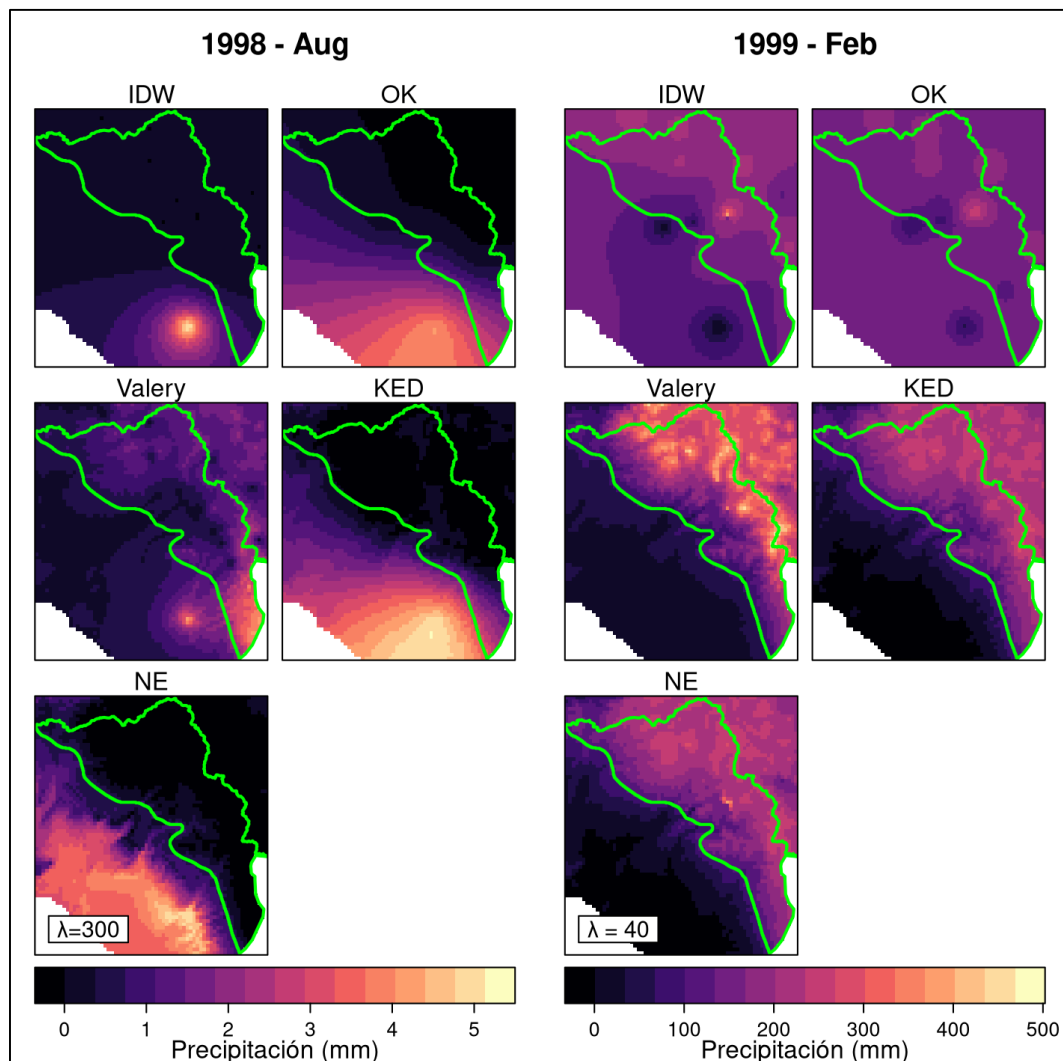


Figura N° 39. Mapas de precipitación obtenidos de métodos de interpolación en comparación en agosto de 1998 y febrero de 1999 – Zona 6

Fuente: Elaboración propia

Por otro lado, en la Figura N° 39 y Figura N° 40 se presentan similares mapas para la zona 6. Se observa que la mayor estimación presente en el método de Valery se da ahora en los meses húmedos y secos. Esta mayor estimación se presenta sobre todo en puntos localizados. Además, el método KED y NE muestran también gran similitud, y el relieve topográfico se muestra menos marcado que en la región 6.

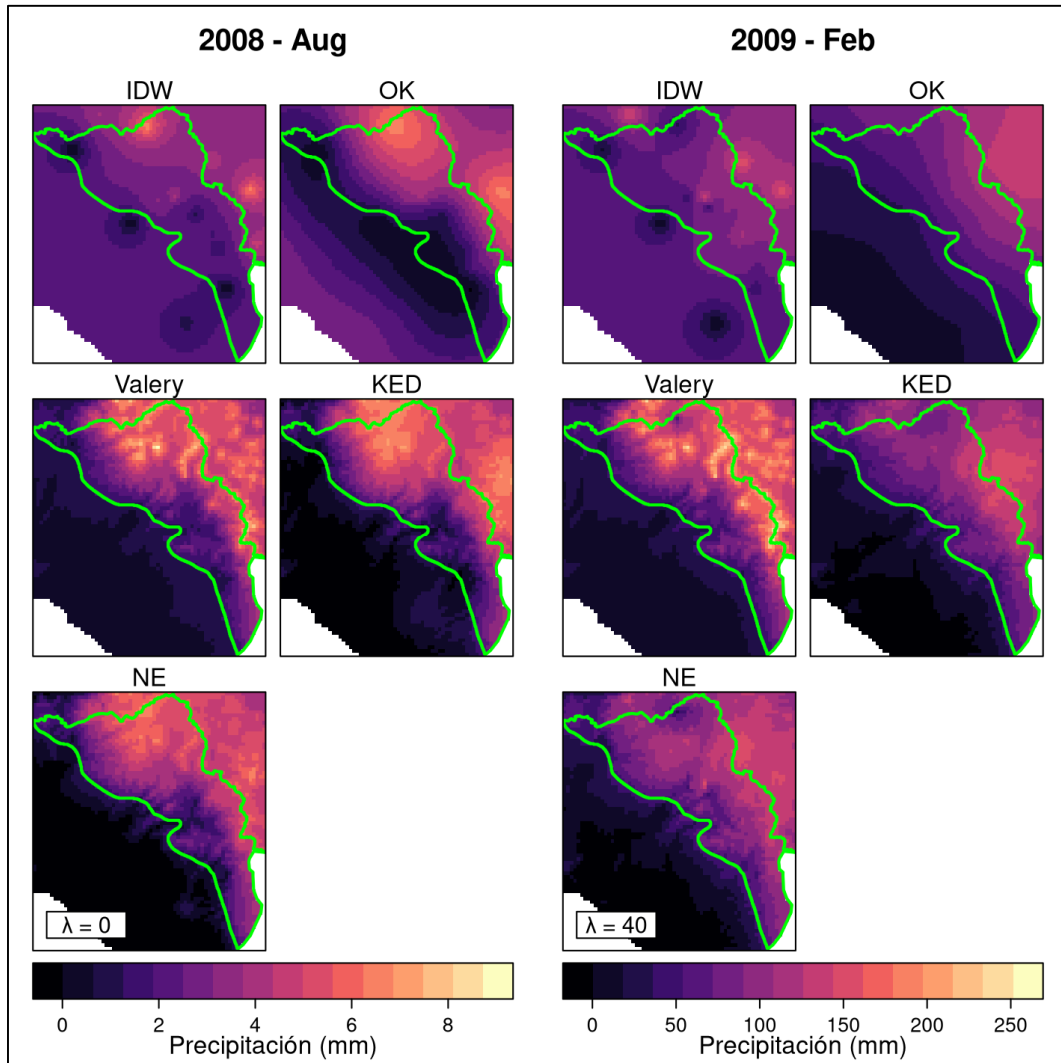


Figura N° 40. Mapas de precipitación obtenidos de métodos de interpolación en comparación en agosto de 2008 y febrero de 2009 – Zona 3

Fuente: Elaboración propia

5.4 DESEMPEÑO EN MODELAMIENTO HIDROLÓGICO

De forma similar a la sección 5.3, de entre las diferentes aproximaciones de distancias no euclídeas propuestas se ha considerado para esta etapa solo la NE2, dado que ha mostrado mejores resultados en la mayoría de las zonas de análisis. El método con esta distancia específica es nombrado como NE. Es así que los métodos de interpolación de precipitación que se han probado en esta etapa fueron cinco: IDW, OK, Valery, KED y NE.

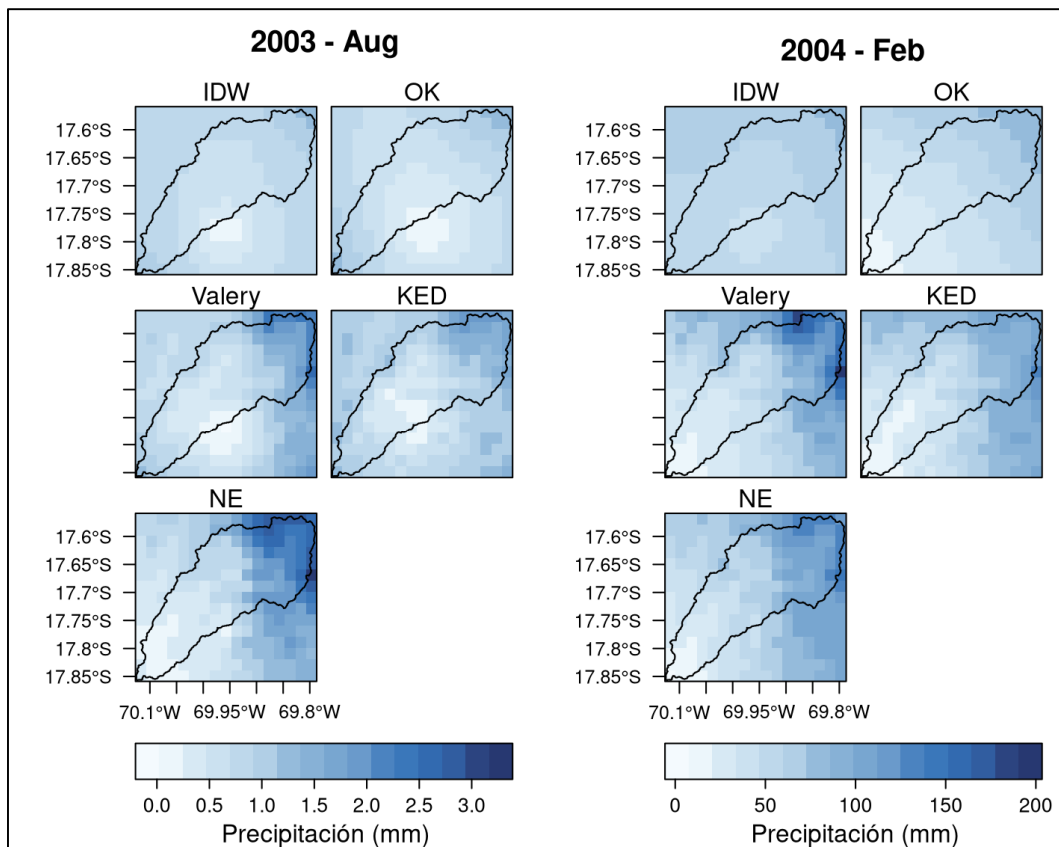


Figura N° 41. Interpolación de precipitación con diferentes métodos en la cuenca Caplina.

Fuente: Elaboración propia

En la cuenca Caplina las salidas de estos métodos de interpolación se muestran en la Figura N° 41 para un mes de característico de temporada húmeda (2004-Febrero) y otro para época seca (2003-Agosto). Se aprecia así que en esta área los métodos que consideran elevación muestran mayores estimaciones de precipitación en zonas altas, como era de esperarse. Además, una característica diferenciable es la mayor conectividad o similitud de magnitudes que existe en las partes altas con el método NE, especialmente en el mes seco.

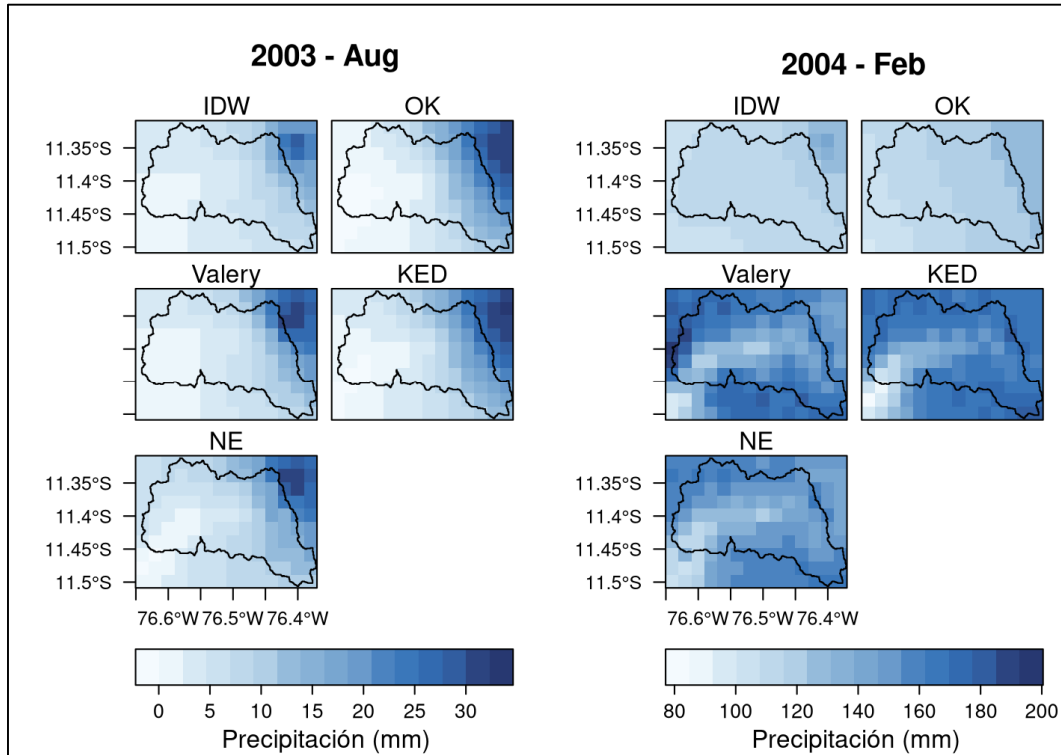


Figura N° 42. Interpolación de precipitación con diferentes métodos en la cuenca Obrajillo.

Fuente: Elaboración propia

De la misma forma, interpolaciones en meses específicos para la cuenca Obrajillo se muestran en la Figura N° 42. En este caso se toman los mismos meses que en la anterior figura. Es así que, en el caso del mes de temporada seca, los campos de precipitación estimados bajo los diferentes métodos tienden a ser bastante similares, incluso considerando los métodos que no consideran elevación. Por otro lado, en el mes de época seca la contribución de la elevación como predictor es mucho más resaltante. Si bien es cierto que en los métodos que consideran esta variable, los campos de precipitación son bastante similares, se observa una precipitación mayor localizada en el método de Valery. Esto posiblemente por tener un parámetro θ de gradiente altitudinal que se ha calibrado de forma global y que no se adapta necesariamente a cada mes, como se vio anteriormente.

Para todos los casos en ambas cuencas las series de precipitación media han sido calculadas a partir de estos campos de precipitación y se muestran en la Figura N° 43 y Figura N° 44, para las cuencas Caplina y Obrajillo respectivamente.

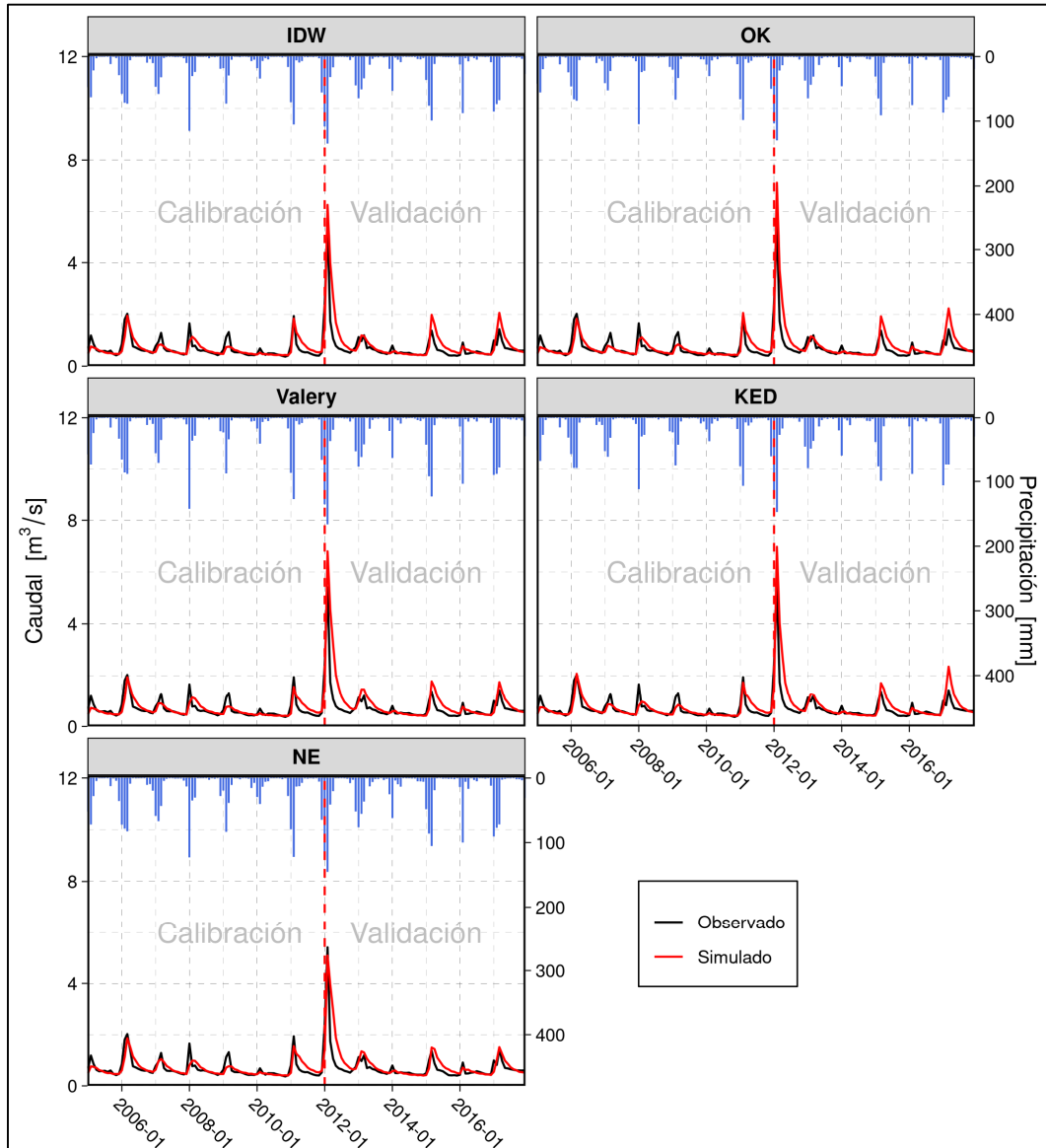


Figura N° 43. Simulación hidrológica mediante el modelo GR2M para diferentes series de precipitación - cuenca Caplina.

Fuente: Elaboración propia

El modelo hidrológico GR2M usado ha mostrado resultados satisfactorios. Las series de caudal simulado para la cuenca Caplina (Figura N° 43) muestran un comportamiento consistente con el observado y mantienen la estacionalidad para todos los métodos de interpolación. Sin embargo, se puede notar que en el método NE no se sobreestiman el caudal pico de inicios del 2012 y los siguientes dos mayores en magnitud de la etapa de validación, a diferencia de los otros métodos.

Los resultados del *Nash* para esta los métodos de interpolación evaluados considerando la etapa de calibración y validación se muestran en la Tabla N° 7. Se observa que los métodos que no consideran la elevación tienen una mayor performance en general que los métodos que si la consideran. Sin embargo, la simulación con las series del método NE, tiene índices equivalentes a los de los primeros. En particular muestra junto con el método OK, la mejor performance en la validación.

Además, si se considerasen las características particularmente áridas de la cuenca y añadiendo como un quinto parámetro del modelo un factor k que multiplica a la evapotranspiración potencial calculada, se obtienen *Nash* de 0.81 y 0.66 para calibración y validación respectivamente, con el método NE.

Tabla N° 7. Performance (Nash) de modelamiento hidrológico con diferentes series de precipitación – Cuenca Caplina
Fuente: Elaboración propia

Etapa	IDW	OK	KED	Valery	NE
Calibración	0.62	0.60	0.61	0.59	0.59
Validación	0.49	0.50	0.40	0.24	0.50

Adicionalmente, la Tabla N° 8 muestra el desempeño medido en forma de *KGE* y *Bias* (%). Se observa que un menor *Bias* es producido por el método NE en la etapa de validación, mientras que el *KGE* es mucho mayor al del resto de métodos. En la etapa de calibración, ambos índices se encuentran dentro del rango de los demás métodos.

Tabla N° 8. Performance (KGE / Bias [%]) de modelamiento hidrológico con diferentes series de precipitación – Cuenca Obrajillo
Fuente: Elaboración propia

Etapa	IDW	OK	KED	Valery	NE
Calibración	0.75 / 0.4	0.74 / 0.5	0.75 / -0.2	0.72 / 0.2	0.72 / 0.4
Validación	0.58 / 21.7	0.51 / 22.3	0.48 / 23.8	0.42 / 27.9	0.68 / 17.5

En la Figura N° 44 se muestra la comparación entre caudales simulados y observados para la cuenca Obrajillo considerando la aplicación de los diferentes

métodos de interpolación. Si bien es cierto que las simulaciones se muestran un comportamiento bastante similar a las series observadas, todas coinciden en estimar con bastante imprecisión los picos de los años 1999 y 2000. Asimismo, todas las simulaciones subestiman la mayoría de los picos en la etapa de validación.

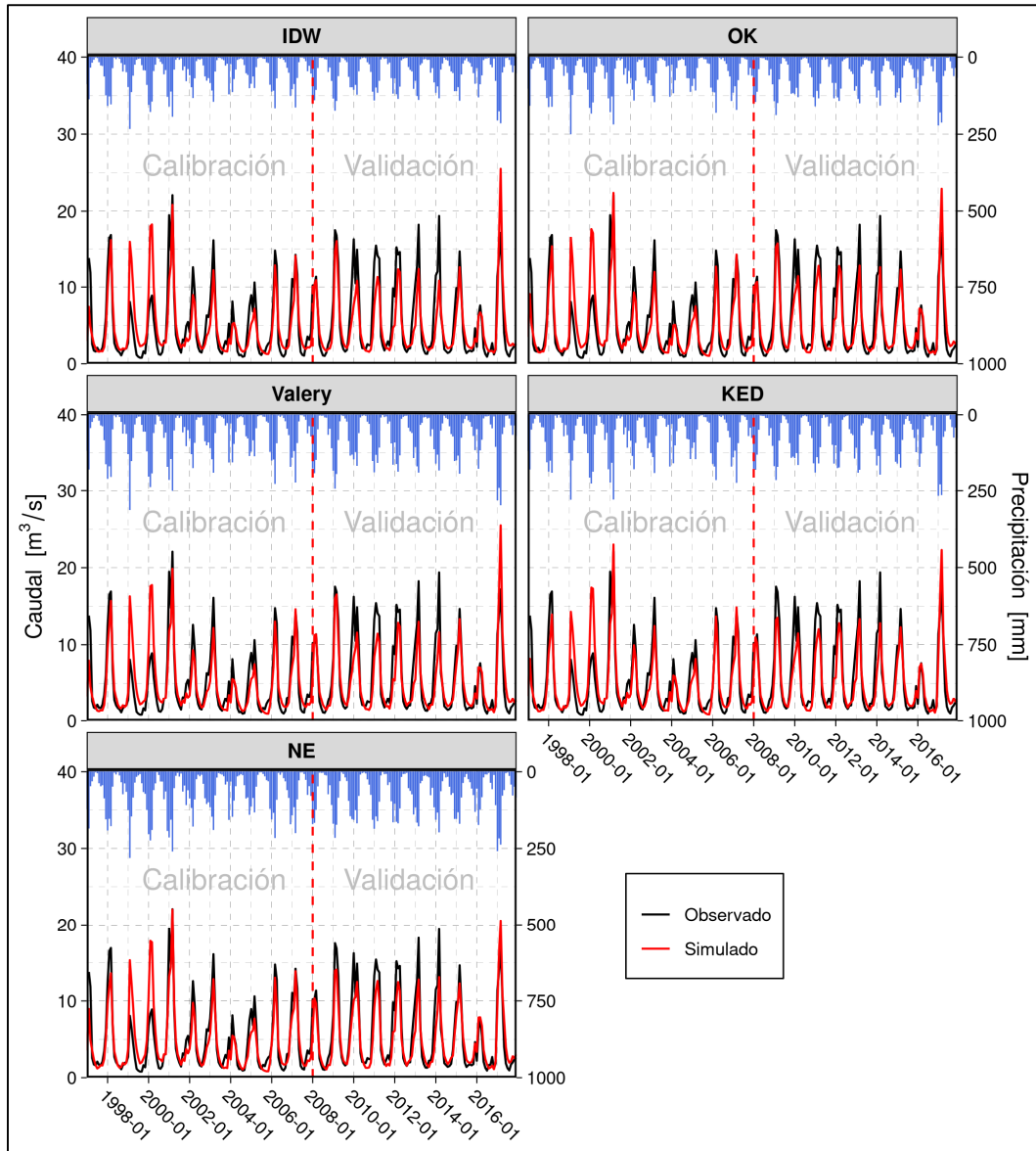


Figura N° 44. Simulación hidrológica mediante el modelo GR2M para diferentes series de precipitación - cuenca Obrajillo.

Fuente: Elaboración propia

Incluso considerando lo último, los *Nash* obtenidos muestran una buena performance (Tabla N° 9). En particular las series del método NE son las que mejor

desempeño muestran, tanto en calibración (junto al KED), como en validación. Es interesante agregar que, cambiando los periodos de calibración y validación, los índices de *Nash* pueden variar drásticamente. Por ejemplo, considerando las características descritas de los primeros años y tomando como calibración el periodo 2001-2009 y como validación el periodo 2010-2017, se obtienen índices de 0.93 y 0.89 respectivamente.

Asimismo, la performance con las series de precipitación NE medida en términos de *KGE* y *Bias* (%) muestra estar en el rango de la de los demás métodos (Tabla N° 10). En particular, en la etapa de validación el método Valery muestra tener mejores resultados, mientras que en la calibración el *Bias* se hace menor con el método IDW y el *KGE* es mayor con los métodos KED y NE.

Tabla N° 9. Performance (Nash) de modelamiento hidrológico con diferentes series de precipitación – Cuenca Obrajillo
Fuente: Elaboración propia

Etapa	IDW	OK	KED	Valery	NE
Calibración	0.67	0.71	0.74	0.70	0.74
Validación	0.72	0.81	0.81	0.78	0.83

Tabla N° 10. Performance (*KGE* / *Bias* [%]) de modelamiento hidrológico con diferentes series de precipitación – Cuenca Obrajillo
Fuente: Elaboración propia

Etapa	IDW	OK	KED	Valery	NE
Calibración	0.80 / -1.7	0.83 / -1.9	0.84 / -2.4	0.81 / -2.4	0.84 / -2.7
Validación	0.72 / -13.6	0.76 / -10.5	0.74 / -12.6	0.77 / -10.3	0.75 / -11.9

CONCLUSIONES

- Entre las variantes de distancias no euclídeas evaluadas, el uso de la distancia no euclídea 2 (NE 2), que consiste en la estimación para dos ubicaciones de la longitud del más corto recorrido espacial entre ellas considerando posibles trayectorias en la superficie topográfica y atmósfera y un factor de penalidad para tramos verticales, muestra una mejor performance en casi todas las doce zonas de análisis. Las otras dos variantes de distancias no euclídeas propuestas, muestran resultados y índices de desempeño muy cercanos a los de esta distancia. En ese sentido, la variante de distancia NE 1 es la que resultados más cercanos muestra (salvo en la zona 1 en términos de *RMSE* y en la zona 4 con respecto al *Nash*); es por ello que se considera a esta última aproximación como una alternativa más eficiente computacionalmente a la distancia NE 2.
- La metodología en base a distancias no euclídeas ha mostrado una mejor performance (considerando las métricas de *RMSE*, *Nash* y *r*) entre los métodos evaluados en la mayoría de las doce zonas analizadas; sin embargo, esto no ocurre en las zonas 10 y 11, donde los métodos IDW y Valery muestran el mejor desempeño respectivamente. Asimismo, en algunas zonas el método de Valery presenta mayores *Nash*.
- Los métodos de interpolación evaluados que consideran la elevación como variable auxiliar (KED, Valery, IED) han mostrado mejores índices de desempeño que los otros métodos que no hacen uso de esta información (IDW, OK) en la mayoría de las zonas de análisis. Un caso particular es el de la zona 12 correspondiente a la vertiente del Titicaca donde no hay una diferencia significativa entre el desempeño de ambos tipos de métodos; por otro lado, las mayores diferencias se aprecian en zonas con cambios topográficos fuertes en cortas distancias, como son la zona 11 y las pertenecientes a la vertiente del Pacífico. En tal sentido, los métodos de interpolación que consideran elevación, y en particular la metodología con distancias no euclídeas planteada, son más eficientes en regiones con cambios topográficos abruptos.

- Los resultados de la aplicación del modelo hidrológico GR2M en las cuencas Obrajillo y Caplina, muestran que la precipitación estimada en base a la metodología de distancias no euclídea permite obtener una mejora significativa en la performance del modelo hidrológico con respecto a otros métodos de interpolación que consideran la elevación.

RECOMENDACIONES

- En la presente investigación se evaluó el uso de la elevación y las distancias no euclídeas asociadas como información auxiliar. En base a ello se pudo estimar mejor la precipitación en los distintos esquemas de interpolación; sin embargo, otras variables auxiliares no fueron evaluadas. En ese sentido, se considera recomendable también evaluar otras variables que suelen ser usadas en la literatura, tales como la distancia con respecto al mar, porcentaje de cuerpos de agua, pendiente, latitud o longitud; las cuales, sin embargo, no suelen mostrar la misma significancia que la elevación. Asimismo, los productos de percepción remota proveen de información valiosa, que puede ayudar a mejorar la estimación. En cualquiera de los casos, se recomienda usar la elevación como variable auxiliar en compañía de alguna de las otras posibles variables mencionadas.
- La metodología planteada originalmente por Frei (2014) fue aplicada para la construcción de un producto grillado de temperaturas diarias y se consideraron perfiles no lineales. En ese sentido, se recomienda la evaluación de la presente metodología para la precipitación a paso diario, así como para la interpolación de temperaturas. Asimismo, el uso de perfiles diferentes al modelo lineal para representar la tendencia subyacente es aconsejable previo estudio del comportamiento de la precipitación en la zona, y especialmente para zonas extensas y con muchos datos.
- Existen otras metodologías de interpolación las cuales no se han evaluado en esta investigación. En tal sentido métodos como el de cópulas (Bárdossy y Pegram, 2013), u métodos que plantean una interpolación espacio-temporal pueden ser evaluados. Asimismo, técnicas basadas en métodos de machine learning han mostrado resultados positivos, tal es el caso evaluado por Navarro (2019) en Perú, que pueden profundizarse.
- Se recomienda evaluar la aplicación en modelamiento de escorrentía de los resultados de precipitación, de los métodos descritos, en otras cuencas y con otros modelos hidrológicos, así como en otras aplicaciones donde la precipitación sea variable de entrada.

REFERENCIAS BIBLIOGRÁFICAS

- Aalto, J., Pirinen, P., y Heikkinen, J. (2013). Spatial interpolation of monthly climate data for Finland: comparing the performance of kriging and generalized additive models. *Theoretical and Applied Climatology*, 112, 99–111. <https://doi.org/10.1007/s00704-012-0716-9>
- Ahrens, C. D., y Henson, R. (2018). *Essentials of meteorology : an invitation to the atmosphere* (8ª ed.). Cengage Learning.
- Aybar, C., Fernández, C., Huerta, A., Lavado, W., Vega, F., y Felipe-Obando, O. (2020). Construction of a high-resolution gridded rainfall dataset for Peru from 1981 to the present day. *Hydrological Sciences Journal*. <https://doi.org/10.1080/02626667.2019.1649411>
- Bachmaier, M., y Backes, M. (2011). Variogram or Semivariogram ? Variance or Semivariance ? Allan Variance or Introducing a New Term ? Variogram or Semivariogram ? Variance or Semivariance ? Allan Variance or Introducing a New Term ? *Mathematical geosciences*, 43(6), 735–740. <https://doi.org/10.1007/s11004-011-9348-3>
- Bárdossy, A., y Pegram, G. (2013). Interpolation of precipitation under topographic influence at different time scales. *Water Resources Research*, 49(8), 4545–4565. <https://doi.org/10.1002/wrcr.20307>
- Beniston, M., Diaz, H. F., y Bradley, R. S. (1997). Climatic change at high elevation sites: An overview. *Climatic Change*, 36(3–4), 233–251. <https://doi.org/10.1023/a:1005380714349>
- Bickel, P., Diggle, P., Fienberg, S., Gather, U., Olkin, I., y Zeger, S. (2007). *Model-based Geostatistics*. Springer. <https://doi.org/10.1007/978-0-387-48536-2>
- Bivand, R. S., Pebesma, E., y Gómez-Rubio, V. (2013). *Applied Spatial Data Analysis with R* (2ª ed.). Springer. <https://doi.org/10.1007/978-1-4614-7618-4>
- Bourgin, P.-Y. (2010). *Spatial interpolation of precipitation in the Norte Chico region, Chile*. [Tesis de maestría, UPMC/MINES ParisTech/AgroParisTech]. <https://webgr.inrae.fr/wp-content/uploads/2012/07/2010-BOURGIN-MASTER.pdf>

- Branisavljević, N., Prodanović, D., Arsić, M., Simić, Z., y Borota, J. (2009). Hydro-meteorological data quality assurance and improvement. *Journal of Serbian Society for Computational Mechanics*, 3(1), 228–249.
- Brovkin, V. (2002). Climate-vegetation interaction. *Journal De Physique. IV: JP*, 12(10), 4–8. <https://doi.org/10.1051/jp4:20020452>
- Brunet-Moret, Y. (1979). Homogénéisation des précipitations. *Orstom*, XVI, 147–170.
- Cabrera, J. (2009). *Calibración de Modelos Hidrológicos*. Instituto para la Mitigación de los Efectos del Fenómeno el Niño, Universidad Nacional de Ingeniería.
- Carter, C. (2002). Great Circle Distances, Computing the distance between two points on the surface of the Earth. *SiRF*, 1–6.
- Chamorro, A. (2015). *Stochastic and hydrological modelling for climate change prediction in the Lima region, Peru* [Tesis de doctorado, Universität Stuttgart]. <https://elib.uni-stuttgart.de/handle/11682/655>
- Chen, F. W., y Liu, C. W. (2012). Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. *Paddy and Water Environment*, 10(3), 209–222. <https://doi.org/10.1007/s10333-012-0319-1>
- Condom, T., Rau, P., y Espinoza, J. C. (2011). Correction of TRMM 3B43 monthly precipitation data over the mountainous areas of Peru during the period 1998-2007. *Hydrological Processes*, 25(12), 1924–1933. <https://doi.org/10.1002/hyp.7949>
- Coron, L., Thirel, G., Delaigue, O., Perrin, C., y Andréassian, V. (2017). The suite of lumped GR hydrological models in an R package. *Environmental Modelling & Software*, 94, 166–171. <https://doi.org/10.1016/J.ENVSOFT.2017.05.002>
- Curriero, F. C. (2006). On the use of non-euclidean distance measures in geostatistics. *Mathematical Geology*, 38(8), 907–926. <https://doi.org/10.1007/s11004-006-9055-7>
- Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1), 269–271. <https://doi.org/10.1007/BF01386390>

- Dingman, S. L. (2015). *Physical hydrology* (3ª ed.). Waveland Press.
- Espinoza, J. C., Ronchail, J., Guyot, J. L., Cochonneau, G., Naziano, F., Lavado, W., de Oliveira, E., Pombosa, R., y Vauchel, P. (2009). Spatio-temporal rainfall variability in the Amazon basin countries (Brazil, Peru, Bolivia, Colombia, and Ecuador). *International Journal of Climatology*, 29(11), 1574–1594. <https://doi.org/10.1002/joc.1791>
- Fekete, B. M., Vörösmarty, C. J., Roads, J. O., y Willmott, C. J. (2004). Uncertainties in precipitation and their impacts on runoff estimates. *Journal of Climate*, 17(2), 294–304. [https://doi.org/10.1175/1520-0442\(2004\)017<0294:UIPATI>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<0294:UIPATI>2.0.CO;2)
- Frei, C. (2014). Interpolation of temperature in a mountainous region using nonlinear profiles and non-Euclidean distances. *International Journal of Climatology*, 34(5), 1585–1605. <https://doi.org/10.1002/joc.3786>
- Giraldo, R. (2002). *Introducción a la geoestadística: Teoría y aplicación*. Universidad Nacional de Colombia.
- Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, 228(1), 113–129. [https://doi.org/https://doi.org/10.1016/S0022-1694\(00\)00144-X](https://doi.org/https://doi.org/10.1016/S0022-1694(00)00144-X)
- Gottardi, F., Obled, C., Gailhard, J., y Paquet, E. (2012). Statistical reanalysis of precipitation fields based on ground network data and weather patterns: Application over French mountains. En *Journal of Hydrology* (Vols. 432–433, pp. 154–167). <https://doi.org/10.1016/j.jhydrol.2012.02.014>
- Gupta, H. V., Kling, H., Yilmaz, K. K., y Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1), 80–91. <https://doi.org/https://doi.org/10.1016/j.jhydrol.2009.08.003>
- Gupta, H. V., Sorooshian, S., y Yapo, P. O. (1999). Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration. *Journal of Hydrologic Engineering*, 4(2), 135–143. [https://doi.org/10.1061/\(asce\)1084-0699\(1999\)4:2\(135\)](https://doi.org/10.1061/(asce)1084-0699(1999)4:2(135))
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A.,

- Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., y Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160), 850–853. <https://doi.org/10.1126/science.1244693>
- Hengl, T., Heuvelink, G., y Stein, A. (2003). *Comparison of kriging with external drift and regression-kriging. Technical note*. International Institute for Geo-information Science and Earth (ITC).
- Heuvelink, G. B. M., y Pebesma, E. J. (2002). Is the ordinary kriging variance a proper measure of interpolation error? *Proceedings of the fifth international symposium on spatial accuracy assessment in natural resources and environmental sciences, July*, 179–186.
- Hillel, D., Hatfield, J. H., Powlson, D. S., Rosenzweig, C., Scow, K. M., Singer, M. J., y Sparks, D. L. (Eds.). (2005). *Encyclopedia of Soils in the Environment*. Academic Press.
- Hingray, B., Picouet, C., y Musy, A. (2014). *Hydrology: A Science for Engineers*. CRS Press.
- Insel, N., Poulsen, C. J., y Ehlers, T. A. (2010). Influence of the Andes Mountains on South American moisture transport, convection, and precipitation. *Climate Dynamics*, 35(7), 1477–1492. <https://doi.org/10.1007/s00382-009-0637-1>
- Isaaks, E. H. (1989). *Applied Geostatistics*. Oxford University Press.
- Keller, V. D. J., Tanguy, M., Prosdocimi, I., Terry, J. A., Hitt, O., Cole, S. J., Fry, M., Morris, D. G., y Dixon, H. (2015). CEH-GEAR: 1 km resolution daily and monthly areal rainfall estimates for the UK for hydrological use. *Earth System Science Data Discussions*, 8(1), 83–112. <https://doi.org/10.5194/essdd-8-83-2015>
- Kling, H., Fuchs, M., y Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, 424–425, 264–277. <https://doi.org/https://doi.org/10.1016/j.jhydrol.2012.01.011>
- Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Southern African Institute of Mining and*

Metallurgy, 52(6), 119–139.

L'Hostis, y Alain. (2020). All geographical distances are optimal. <http://journals.openedition.org/cybergeog>, 2020. <https://doi.org/10.4000/CYBERGEO.35048>

Lavado, W., y Fernández, C. (2014). *Análisis comparativo de métodos de interpolación espacial para la elaboración de mapas climáticos de precipitación*. Servicio Nacional de Meteorología e Hidrología.

Legates, D. R., y McCabe Jr., G. J. (1999). Evaluating the use of “goodness-of-fit” Measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35(1), 233–241. <https://doi.org/10.1029/1998WR900018>

Lloyd, C. D. (2005). Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. *Journal of Hydrology*, 308, 128–150. <https://doi.org/10.1016/j.jhydrol.2004.10.026>

Mair, A., y Fares, A. (2011). Comparison of rainfall interpolation methods in a mountainous region of a tropical island. *Journal of Hydrologic Engineering*, 16(4), 371–383. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000330](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000330)

Majani, B. S. (2007). *Analysis of External Drift Kriging Algorithm with application to precipitation estimation in complex orography*. [Tesis de maestría, International Institute for Geo-information Science and Earth Observation]. <https://www.iirs.gov.in/node/310>

Masson, D., y Frei, C. (2014). Spatial analysis of precipitation in a high-mountain region: Exploring methods with multi-scale topographic predictors and circulation types. *Hydrology and Earth System Sciences*, 18(11), 4543–4563. <https://doi.org/10.5194/hess-18-4543-2014>

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58(8), 1246–1266. <https://doi.org/https://doi.org/10.2113/gsecongeo.58.8.1246>

McMillan, H., Jackson, B., Clark, M., Kavetski, D., y Woods, R. (2011). Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. *Journal of Hydrology*, 400(1–2), 83–94. <https://doi.org/10.1016/j.jhydrol.2011.01.026>

- MIDAGRI. (s. f.). *Clima de la sierra y Selva*.
<https://www.midagri.gob.pe/portal/datero/53sector-agrario/el-clima>
- Mouelhi, S., Michel, C., Perrin, C., y Andréassian, V. (2006). Stepwise development of a two-parameter monthly water balance model. *Journal of Hydrology*, 318(1–4), 200–214. <https://doi.org/10.1016/j.jhydrol.2005.06.014>
- Mourre, L., Condom, T., Junquas, C., Lebel, T., E. Sicart, J., Figueroa, R., y Cochachin, A. (2016). Spatio-temporal assessment of WRF, TRMM and in situ precipitation data in a tropical mountain environment (Cordillera Blanca, Peru). *Hydrology and Earth System Sciences*, 20(1), 125–141. <https://doi.org/10.5194/hess-20-125-2016>
- Musy, A., y Higy, C. (2010). Hydrology: A science of nature. En *Hydrology: A Science of Nature*. <https://doi.org/10.1201/b10426>
- Nash, J. E., y Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3), 282–290. [https://doi.org/https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/https://doi.org/10.1016/0022-1694(70)90255-6)
- Navarro, E. (2019). *Estudio de la distribución espacial de la precipitación mediante productos de percepción remota en la cuenca alta del río Piura* [Tesis de grado, Universidad Nacional de Ingeniería]. <http://cybertesis.uni.edu.pe/handle/uni/17836>
- Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., y Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 - Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. *Journal of Hydrology*, 303(1–4), 290–306. <https://doi.org/10.1016/j.jhydrol.2004.08.026>
- Pardo-Igúzquiza, E. (1998). Comparison of geostatistical methods for estimating the areal average climatological rainfall mean using data on precipitation and topography. *International Journal of Climatology*, 18, 1031–1047. [https://doi.org/https://doi.org/10.1002/\(SICI\)1097-0088\(199807\)18:9%3C1031::AID-JOC303%3E3.0.CO;2-U](https://doi.org/https://doi.org/10.1002/(SICI)1097-0088(199807)18:9%3C1031::AID-JOC303%3E3.0.CO;2-U)
- Perrin, C., Michel, C., y Andréassian, V. (2007). *Modèles hydrologiques du Génie Rural (GR)*.

- R Core Team. (2020). R: A language and environment for statistical computing. En *R Foundation for Statistical Computing*, Vienna, Austria. <https://www.r-project.org/>
- Rau, P. (2017). *Precipitation, runoff and water balance regimes variability along the Peruvian Pacific slope and coast: ENSO influence and sensitivity to hydroclimatic change*. [Tesis de doctorado, Université Toulouse 3 Paul Sabatier]. <https://tel.archives-ouvertes.fr/tel-01925488>
- Rau, P., Bourrel, L., Labat, D., Melo, P., Dewitte, B., Frappart, F., Lavado, W., y Felipe, O. (2017). Regionalization of rainfall over the Peruvian Pacific slope and coast. *International Journal of Climatology*, 37(1), 143–158. <https://doi.org/10.1002/joc.4693>
- Rau, P., Bourrel, L., Labat, D., Ruelland, D., Frappart, F., Lavado, W., Dewitte, B., y Felipe, O. (2019). Assessing multidecadal runoff (1970–2010) using regional hydrological modelling under data and water scarcity conditions in Peruvian Pacific catchments. *Hydrological Processes*, 33(1), 20–35. <https://doi.org/10.1002/hyp.13318>
- Ritter, A., y Muñoz-Carpena, R. (2013). Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. *Journal of Hydrology*, 480, 33–45. <https://doi.org/https://doi.org/10.1016/j.jhydrol.2012.12.004>
- Ruelland, D. (2020). Should altitudinal gradients of temperature and precipitation inputs be inferred from key parameters in snow-hydrological models? *Hydrology and Earth System Sciences*, 24(5), 2609–2632. <https://doi.org/10.5194/hess-24-2609-2020>
- Singh, P., y Verma, P. (2019). A Comparative Study of Spatial Interpolation Technique (IDW and Kriging) for Determining Groundwater Quality. En *GIS and Geostatistical Techniques for Groundwater Science* (pp. 43–56). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-815413-7.00005-5>
- Thiessen, A. H. (1911). Precipitation averages for large areas. *Monthly Weather Review*, 39(7), 1082–1089. [https://doi.org/10.1175/1520-0493\(1911\)39<1082b:pafla>2.0.co;2](https://doi.org/10.1175/1520-0493(1911)39<1082b:pafla>2.0.co;2)

- Valéry, A., Andréassian, V., y Perrin, C. (2010). Regionalization of precipitation and air temperature over high-altitude catchments – learning from outliers. *Hydrological Sciences Journal*, 55(6), 928–940. <https://doi.org/10.1080/02626667.2010.504676>
- Wang, Q. J., Pagano, T. C., Zhou, S. L., Hapuarachchi, H. A. P., Zhang, L., y Robertson, D. E. (2011). Monthly versus daily water balance models in simulating monthly runoff. *Journal of Hydrology*, 404(3–4), 166–175. <https://doi.org/10.1016/j.jhydrol.2011.04.027>
- Weather, H., Sorooshian, S., y Sharma, K. D. (2008). *Hydrological Modeling in Arid and Semi-Arid Areas*. Cambridge University Press.
- Webster, R., y Oliver, M. A. (2007). *Geostatistics for Environmental Scientists*. John Wiley & Sons.
- World Meteorological Organization. (2008). *Guide to Hydrological Practices: Vol. I*. WMO-No. 168.

ANEXOS

Anexo 1.

Listado de estaciones seleccionadas del proceso de control de calidad y que pertenecen a la región de estudio seleccionada.

N°	CÓDIGO	ESTACIÓN	LONGITUD	LATITUD	ALTITUD	CUENCA
1	000605	ABANCAY	-72.8706	-13.6083	2776	INTERCUENCA ALTO APURÍMAC
2	000687	ACOMAYO	-71.6839	-13.9216	3212	INTERCUENCA ALTO APURÍMAC
3	000739	PAMPA GALERAS	-74.4036	-14.6692	3998	CUENCA GRANDE
4	155107	PAMPA LIBRE	-76.9700	-10.8658	1756	CUENCA HUAURA
5	158309	PAMPA UMALZO (TITIJONES)	-70.4236	-16.8750	4609	CUENCA TAMBO
6	000751	PAMPACOLCA	-72.5668	-15.7167	2950	CUENCA CAMANÁ
7	157205	PAMPAHUASI	-74.2503	-14.4836	3722	CUENCA ACARÍ
8	000762	PAMPAHUTA	-70.6758	-15.4835	4400	CUENCA COATA
9	157207	PAMPAMARCA	-74.2003	-14.2336	4222	CUENCA PAMPAS
10	156217	PAMPAS	-73.8247	-13.4367	2031	CUENCA PAMPAS
11	000508	PAMPAS	-74.8841	-12.4020	3245	CUENCA MANTARO
12	156203	PARAS	-74.6336	-13.5503	3330	CUENCA PAMPAS
13	157408	PARATIA	-70.6003	-15.4503	4364	CUENCA COATA
14	155112	PARIACANCHA	-76.5027	-11.3936	3842	CUENCA CHILLÓN
15	000362	BAMBAMARCA	-78.5183	-6.6766	2495	INTERCUENCA ALTO MARAÑÓN IV
16	156131	PARIONA	-75.0669	-13.5336	4425	CUENCA ICA
17	155212	PARQUIN	-76.7198	-10.9667	3571	CUENCA HUAURA
18	000686	PARURO	-71.8447	-13.7670	3050	INTERCUENCA ALTO APURÍMAC
19	155211	PATON	-76.7003	-10.6669	4423	CUENCA HUAURA
20	157210	PAUCACORRAL	-74.1003	-14.6669	3398	CUENCA ACARÍ
21	158328	PAUCARANI	-69.7804	-17.5329	4556	CUENCA USHUSUMA
22	000792	PAUCARAY	-73.6388	-14.0438	3238	CUENCA PAMPAS
23	000658	PAUCARBAMBA	-74.5359	-12.5557	3366	CUENCA MANTARO
24	000689	PAUCARTAMBO	-71.5906	-13.3244	3042	CUENCA URUBAMBA
25	000745	PAUZA	-73.3411	-15.2753	2477	CUENCA OCOÑA
26	000542	PICOY	-76.7370	-10.9216	2920	CUENCA HUAURA
27	000648	PILCHACA	-75.0841	-12.3994	3578	CUENCA MANTARO
28	158208	PILLONES	-71.2167	-15.9813	4455	CUENCA QUILCA - VITOR - CHILI
29	154110	PIRA	-77.7072	-9.5853	3625	CUENCA CASMA
30	155214	PIRCA	-76.6551	-11.2328	3342	CUENCA CHANCAY - HUARAL
31	000844	PISAC	-71.8509	-13.4195	3002	CUENCA URUBAMBA
32	000644	ACORA	-75.3669	-13.7836	2114	CUENCA ICA
33	000881	PIZACOMA	-69.3686	-16.9070	3930	CUENCA MAURI CHICO
34	000443	POMABAMBA	-77.4573	-8.8217	2985	INTERCUENCA ALTO MARAÑÓN V

35	000812	POMACANCHI	-71.5726	-14.0278	3681	CUENCA URUBAMBA
36	157329	PORPERA	-71.3167	-15.3500	4195	CUENCA CAMANÁ
37	000778	PROGRESO	-70.3678	-14.6717	3913	CUENCA AZÁNGARO
38	000815	PUCARA	-70.3666	-15.0457	3866	CUENCA PUCARÁ
39	000625	ACOSTAMBO	-75.0573	-12.3608	3603	CUENCA MANTARO
40	157307	PUICA	-72.6926	-15.0606	3661	CUENCA OCOÑA
41	000708	PUNO	-70.0121	-15.8263	3812	INTERCUENCA 0173
42	000848	PUQUINA	-71.1848	-16.6298	3085	CUENCA TAMBO
43	000736	PUQUIO	-74.1316	-14.6993	3168	CUENCA ACARÍ
44	157322	PUSA PUSA	-71.6503	-15.2169	4226	INTERCUENCA ALTO APURÍMAC
45	157414	PUTINA	-69.8757	-14.9210	3858	CUENCA HUANCANÉ
46	153211	QUEBRADA HONDA	-78.6503	-6.9003	3691	CUENCA JEQUETEPEQUE
47	153108	QUEBRADA SHUGAR	-78.4569	-6.6878	3292.7	INTERCUENCA ALTO MARAÑON IV
48	158331	BOCATOMA	-69.6267	-17.5802	4260	CUENCA CAPLINA
49	000309	QUILCATE	-78.7440	-6.8227	3082	CUENCA CHANCAY - LAMBAYEQUE
50	157401	QUILLISANI	-70.7503	-15.3836	4758	CUENCA COATA
51	154102	QUIRUVILCA	-78.3079	-8.0042	4047	CUENCA MOCHE
52	157733	RAMIS	-69.8664	-15.2371	3850	INTERCUENCA RAMIS
53	156207	RAYUSCA	-74.4169	-13.8836	3924	CUENCA PAMPAS
54	000441	RECUAY	-77.4537	-9.7292	3431	CUENCA SANTA
55	155229	RICRAN	-75.5271	-11.5404	3667	CUENCA PERENE
56	000821	RINCON DE LA CRUZ	-69.8096	-15.9902	3880	INTERCUENCA 0173
57	151210	RIO BLANCO	-76.2597	-11.7345	3503	CUENCA RÍMAC
58	005232	RUNATULLO	-75.0123	-11.6181	3473	CUENCA PERENE
59	155232	RUNATULLO	-75.0503	-11.5836	3690	CUENCA PERENE
60	000672	SAISA	-74.4003	-14.9336	3097	CUENCA ACARÍ
61	000801	SALAMANCA	-72.8320	-15.5029	3203	CUENCA OCOÑA
62	006200	SALCABAMBA	-74.7858	-12.1994	3280	CUENCA MANTARO
63	000398	SALPO	-78.6071	-8.0053	3418	CUENCA MOCHE
64	154100	SALPO	-78.6169	-8.0003	3250	CUENCA MOCHE
65	155225	SAN JOSE DE PARAC	-76.2581	-11.8005	3829	CUENCA RÍMAC
66	000369	SAN JUAN	-78.4911	-7.2976	2228	CUENCA JEQUETEPEQUE
67	156114	SAN JUAN DE CASTROVIRREYNA	-75.6276	-13.2120	1856	CUENCA SAN JUAN
68	151506	SAN JUAN DE HUIRPACANCHA	-75.2336	-13.9503	3650	CUENCA ICA
69	156116	SAN JUAN DE JARPA	-75.4169	-12.0336	3920	CUENCA MANTARO
70	000594	SAN JUAN DE JARPA	-75.4318	-12.1245	3670	CUENCA MANTARO
71	153213	SAN JUAN DE LACAMACA	-78.5336	-6.7003	2671	INTERCUENCA ALTO MARAÑON IV
72	156113	SAN JUAN DE YANAC	-75.7870	-13.2106	2513	CUENCA SAN JUAN
73	156102	SAN LAZARO DE ESCOMARCA	-76.3521	-12.1809	3758	CUENCA LURÍN
74	000655	SAN LORENZO	-75.3836	-11.8503	3320	CUENCA MANTARO
75	004431	CABANA	-78.0046	-8.3906	3364	CUENCA SANTA

76	000370	SAN MARCOS	-78.1727	-7.3225	2287	CUENCA CRISNEJAS
77	111175	SAN MATEO DE HUANCHOR	-76.3010	-11.7602	3015	CUENCA RÍMAC
78	000308	SAN MIGUEL	-78.8531	-6.9968	2658	CUENCA JEQUETEPEQUE
79	000795	CABANACONDE	-71.9741	-15.6226	3333	CUENCA CAMANÁ
80	006205	SAN PEDRO DE CACHI	-74.3907	-13.0874	2990	CUENCA MANTARO
81	156205	SAN PEDRO DE CACHI	-74.3586	-13.0653	2990	CUENCA MANTARO
82	155228	SAN PEDRO DE CHUCLU	-75.5003	-11.7503	3380	CUENCA MANTARO
83	000857	SAN PEDRO DE HUACARPANA	-75.6463	-13.0480	3772	CUENCA SAN JUAN
84	156115	SAN PEDRO DE HUACARPANA	-75.6503	-13.0503	3763	CUENCA SAN JUAN
85	151212	SAN PEDRO DE PILAS	-76.2256	-12.4548	2707	CUENCA OMAS
86	000780	CABANILLAS	-70.3466	-15.6391	3885	CUENCA COATA
87	000552	SAN RAFAEL	-76.1766	-10.3293	2687	INTERCUENCA ALTO HUALLAGA
88	000740	SANCOS	-73.9503	-15.0669	2721	CUENCA YAUCA
89	000477	SANTA ANA	-75.2216	-12.0096	3289	CUENCA MANTARO
90	155202	SANTA CRUZ	-76.6349	-11.1983	3583	CUENCA CHANCAY - HUARAL
91	157405	SANTA LUCIA	-70.6003	-15.7003	4080	CUENCA COATA
92	153223	CACHACHI	-78.2686	-7.4511	3203	CUENCA CRISNEJAS
93	000865	SANTA LUCIA	-70.5992	-15.6940	4034	CUENCA COATA
94	000823	SANTA ROSA	-70.7928	-14.6178	3957	CUENCA PUCARÁ
95	154103	CACHICADAN	-78.1494	-8.0918	2900	CUENCA SANTA
96	000426	SANTIAGO ANTUNEZ DE MAYOLO	-77.5249	-9.5165	3079	CUENCA SANTA
97	156123	SANTIAGO DE CHOCORVOS	-75.2508	-13.8331	2700	CUENCA ICA
98	000436	SANTIAGO DE CHUCO	-78.1669	-8.1336	2900	CUENCA SANTA
99	000752	SANTO TOMAS	-72.0886	-14.3994	3253	INTERCUENCA ALTO APURÍMAC
100	150700	SAYAPULLO	-78.4503	-7.4169	3432	CUENCA CHICAMA
101	151213	SHEQUE	-76.5023	-11.6610	3188	CUENCA RÍMAC
102	155480	SHULLCAS	-75.1667	-12.0000	3750	CUENCA MANTARO
103	000759	SICUANI	-71.2367	-14.2374	3534	CUENCA URUBAMBA
104	154111	SIHUAS	-77.6500	-8.5667	2716	INTERCUENCA ALTO MARAÑÓN V
105	000645	SINTO	-75.3336	-13.3336	4039	CUENCA PISCO
106	158318	SITAJARA	-70.1356	-17.3725	3132	CUENCA SAMA
107	158313	CAIRANI	-70.3389	-17.2881	3920	CUENCA LOCUMBA
108	000323	SONDOR-MATARA	-78.2126	-7.2369	2908	CUENCA CRISNEJAS
109	000373	CAJABAMBA	-78.0513	-7.6217	2625	CUENCA CRISNEJAS
110	157328	SUMBAY	-71.3669	-15.9836	4294	CUENCA QUILCA - VITOR - CHILI
111	000502	SURASACA	-76.7836	-10.5169	4546	CUENCA HUAURA
112	158317	SUSAPAYA	-70.1347	-17.3477	3468	CUENCA SAMA
113	000882	TAHUACO - YUNGUYO	-69.0747	-16.3078	3882	INTERCUENCA 0157
114	158323	TALABAYA	-69.9997	-17.5592	3420	CUENCA SAMA
115	156122	TAMBO	-75.2742	-13.6916	3138	CUENCA ICA
116	154108	CAJAMARQUILLA	-77.7414	-9.6320	3286	CUENCA CASMA

117	156106	TANTA	-76.0167	-12.1167	4323	CUENCA CAÑETE
118	151507	TANTARA	-75.6336	-13.0669	3190	CUENCA SAN JUAN
119	155101	TAPACOCHA	-77.5669	-10.0003	3440	CUENCA HUARMEY
120	000816	TARACO	-69.9725	-15.3117	3849	INTERCUENCA RAMIS
121	000540	CAJATAMBO	-76.9897	-10.4779	3432	CUENCA PATIVILCA
122	000877	TARATA	-70.0366	-17.4794	3050	CUENCA SAMA
123	000554	TARMA	-75.6903	-11.3969	3000	CUENCA PERENE
124	000997	TAZA OROYA	-75.9169	-11.5169	3717	CUENCA MANTARO
125	156117	TICRAPO	-75.4336	-13.3836	2174	CUENCA PISCO
126	155111	TINGO	-76.4833	-11.6167	3980	CUENCA RÍMAC
127	157325	TISCO	-71.3064	-15.3142	4175	CUENCA CAMANÁ
128	155169	TOMAS	-75.7503	-12.2336	3948	CUENCA CAÑETE
129	158325	TOQUELA	-69.9489	-17.6608	3566	CUENCA LOCUMBA
130	000647	TUNEL CERO	-75.0860	-13.2593	4498	CUENCA PAMPAS
131	157412	UMAYO	-70.1503	-15.7336	3850	CUENCA ILPA
132	000688	URCOS	-71.6336	-13.7003	3666	CUENCA URUBAMBA
133	156204	URUBAMBA	-74.4167	-14.1167	4028	CUENCA PAMPAS
134	000683	URUBAMBA	-72.1246	-13.3052	2850	CUENCA URUBAMBA
135	000574	USIBAMBA	-75.4669	-12.0503	3588	CUENCA MANTARO
136	158320	VILACOTA	-70.0503	-17.1169	4438	CUENCA MAURI
137	000861	VILACOTA	-70.0430	-17.0791	4468	CUENCA MAURI
138	156111	VILCA	-75.8262	-12.1146	3832	CUENCA CAÑETE
139	156211	VILCASHUAMAN	-73.9524	-13.6688	3522	CUENCA PAMPAS
140	000608	VIQUES	-75.2283	-12.1560	3186	CUENCA MANTARO
141	156210	VISCHONGO	-74.0003	-13.5836	3278	CUENCA PAMPAS
142	153101	CALLANCAS	-78.4773	-7.7675	1501	CUENCA CHICAMA
143	000664	WAYLLAPAMPA	-74.2165	-13.0765	2470	CUENCA MANTARO
144	000475	YANAHUANCA	-76.5129	-10.4896	3129	INTERCUENCA ALTO HUALLAGA
145	000864	YANAQUIHUA	-72.8758	-15.7765	3130	CUENCA OCOÑA
146	157321	YANQUE	-71.5836	-15.6503	4104	CUENCA CAMANÁ
147	155121	YANTAC	-76.4045	-11.3458	4617	CUENCA MANTARO
148	155235	YAULI	-76.0833	-11.6667	4141	CUENCA MANTARO
149	000757	YAURI	-71.4308	-14.8029	3916	INTERCUENCA ALTO APURÍMAC
150	000636	YAUYOS	-75.9104	-12.4917	2285	CUENCA CAÑETE
151	000674	YUCAY	-72.0836	-13.3003	3481	CUENCA URUBAMBA
152	000682	ZURITE	-72.2669	-13.4669	3506	CUENCA URUBAMBA
153	151205	CANCHACALLA	-76.5314	-11.8448	2400	CUENCA RÍMAC
154	000457	CANCHAN	-76.3096	-9.9210	1986	INTERCUENCA ALTO HUALLAGA
155	000876	CANDARAVE	-70.2554	-17.2678	3476	CUENCA LOCUMBA
156	000547	CANTA	-76.6257	-11.4712	2818	CUENCA CHILLÓN
157	000440	AIIA	-77.6067	-9.7816	3508	CUENCA HUARMEY
158	000788	CAPACHICA	-69.8422	-15.6134	3822	INTERCUENCA RAMIS
159	153221	CAPACHIQUE	-78.3169	-7.8503	2750	CUENCA CHICAMA

160	158326	CAPAZO	-69.7355	-17.1877	4530	CUENCA MAURI
161	155223	CARAMPOMA	-76.5151	-11.6550	3424	CUENCA RÍMAC
162	156109	CARANIA	-75.8722	-12.3444	3820	CUENCA CAÑETE
163	155115	CARHUACAYAN	-76.2850	-11.2027	4127	CUENCA MANTARO
164	156214	CARHUANCA	-73.7836	-13.7336	3278	CUENCA PAMPAS
165	157220	CARHUANILLAS	-73.7336	-15.1336	3479	CUENCA YAUCA
166	000873	CARUMAS	-70.6958	-16.8118	3054	CUENCA TAMBO
167	155446	CASAPALCA	-76.2333	-11.6380	4233	CUENCA RÍMAC
168	000998	CASARACRA	-75.9503	-11.4503	3850	CUENCA MANTARO
169	000809	CAY CAY	-71.7003	-13.6003	3150	CUENCA URUBAMBA
170	000753	CAYLLOMA	-71.7702	-15.1888	4318	INTERCUENCA ALTO APURÍMAC
171	000690	CCATCCA	-71.5601	-13.6099	3681	CUENCA URUBAMBA
172	157215	CCECCAÑA	-74.0003	-14.6003	4406	CUENCA ACARÍ
173	000738	CECCHAPAMPA	-74.0003	-14.8336	3900	CUENCA ACARÍ
174	157209	CERRO CONDORILLO	-74.2669	-14.6003	3256	CUENCA ACARÍ
175	000593	CERRO DE PASCO	-76.2642	-10.6934	4359	CUENCA MANTARO
176	155518	CHACLLABAMBA	-71.7200	-13.1085	2701	CUENCA URUBAMBA
177	000455	CHAGLLA	-75.9071	-9.8501	3032	INTERCUENCA ALTO HUALLAGA
178	000747	CHALHUANCA	-73.1792	-14.3928	3358	INTERCUENCA ALTO APURÍMAC
179	007747	CHALHUANCA II	-73.1667	-14.3333	3548	INTERCUENCA ALTO APURÍMAC
180	155119	CHALILLA	-76.3336	-11.9336	3975	CUENCA LURÍN
181	006670	CHALLABAMBA	-71.6472	-13.2175	2740	CUENCA URUBAMBA
182	006640	CHALLACA CO	-75.3821	-13.7905	1800	CUENCA ICA
183	000009	CHALLACA PLU	-75.3906	-13.7919	1880	CUENCA ICA
184	158327	CHALLAPALCA	-69.8131	-17.2372	4280	CUENCA MAURI
185	000662	ALLPACHACA	-74.2669	-13.3836	3600	CUENCA MANTARO
186	151206	CHAUTE	-76.5044	-11.9350	2215	CUENCA RÍMAC
187	000445	CHAVIN	-77.1753	-9.5860	3140	INTERCUENCA ALTO MARAÑÓN V
188	000742	CHAVIÑA	-73.8336	-14.9836	3207	CUENCA YAUCA
189	155116	ALPAMARCA	-76.4503	-11.2169	4715	CUENCA MANTARO
190	000847	CHIGUATA	-71.4091	-16.4064	2902	CUENCA QUILCA - VITOR - CHILI
191	156212	CHILCAYOC	-73.7266	-13.8826	3400	CUENCA PAMPAS
192	158329	CHILLIGUA	-69.6778	-16.5422	4164	CUENCA ILAVE
193	157305	CHINCHAYLLAPA	-72.7336	-14.9169	4497	CUENCA OCOÑA
194	157214	CHIPAO	-73.8842	-14.3669	3342	CUENCA PAMPAS
195	000538	CHIQUIAN	-77.1595	-10.1476	3386	CUENCA PATIVILCA
196	156307	CHITAPAMPA	-71.9747	-13.4150	4306	CUENCA URUBAMBA
197	000758	CHIVAY	-71.6017	-15.6416	3644	CUENCA CAMANÁ
198	000303	CHOTA	-78.6486	-6.5471	2468	CUENCA CHAMAYA
199	153235	CHOTANO LAJAS	-78.7483	-6.5597	2163.4	CUENCA CHAMAYA
200	000860	CHUAPALCA	-69.6436	-17.3050	4177	CUENCA MAURI
201	153208	CHUGUR	-78.7372	-6.6709	2757	CUENCA CHANCAY - LAMBAYEQUE
202	156220	CHUNGUI	-73.6169	-13.2169	3645	CUENCA PAMPAS

203	000750	CHUQUIBAMBA	-72.6506	-15.8460	2850	CUENCA CAMANÁ
204	000764	CHUQUIBAMBILLA	-70.7157	-14.7848	3971	CUENCA PUCARÁ
205	156206	CHUSCHI	-74.3503	-13.5836	3160	CUENCA PAMPAS
206	151211	AMBAR	-77.2833	-10.7500	2100	CUENCA SUPE
207	000643	COCAS	-75.3669	-13.2669	3461	CUENCA PISCO
208	000545	COCHAQUILLO	-76.6669	-10.8003	4575	CUENCA HUAURA
209	000827	COJATA	-69.3556	-15.0167	4344	CUENCA SUCHES
210	006620	COLCABAMBA	-74.6668	-12.4147	3055	CUENCA MANTARO
211	156306	COLQUEPATA	-71.6731	-13.3631	3729	CUENCA URUBAMBA
212	000560	COMAS	-75.1301	-11.7479	3650	CUENCA PERENE
213	000826	ANANEA	-69.5345	-14.6787	4660	CUENCA AZÁNGARO
214	000756	COMBAPATA	-71.4333	-14.1000	3464	CUENCA URUBAMBA
215	000358	CONCHAN	-78.6503	-6.4336	2474	INTERCUENCA ALTO MARAÑÓN IV
216	000437	CONCHUCOS	-77.8503	-8.2669	3200	CUENCA SANTA
217	155104	CONGAS	-77.4336	-10.3336	3018	CUENCA FORTALEZA
218	157420	CONIMA	-69.4336	-15.4503	3869	INTERCUENCA 0171
219	000743	CORACORA	-73.7794	-15.0119	3156	CUENCA YAUCA
220	000669	ANDAHUAYLAS	-73.3668	-13.6486	2990	CUENCA PAMPAS
221	157101	CORDOVA	-75.1870	-14.0406	3225	CUENCA GRANDE
222	000438	CORONGO	-77.9003	-8.5669	3150	CUENCA SANTA
223	155201	ANDAJES	-76.9000	-10.7833	3950	CUENCA HUAURA
224	000489	COSMOS	-75.5669	-12.1503	4547	CUENCA MANTARO
225	153216	COSPAN	-78.5336	-7.4336	2227	CUENCA CHICAMA
226	003216	COSPAN	-78.5411	-7.4286	2423	CUENCA CHICAMA
227	000749	COTAHUASI	-72.8933	-15.2113	2683	CUENCA OCOÑA
228	154113	COTAPARACO	-77.5836	-9.9836	3170	CUENCA HUARMEY
229	007415	CRUCERO	-70.0259	-14.3642	4128	CUENCA AZÁNGARO
230	000803	CRUCERO ALTO	-70.9124	-15.7645	4521	CUENCA CAMANÁ
231	155301	ANDAMARCA	-74.8336	-11.7169	3535	CUENCA MANTARO
232	000677	CURAHUASI	-72.7349	-13.5526	2737	INTERCUENCA ALTO APURÍMAC
233	000737	ANDAMARCA	-73.9669	-14.3836	3509	CUENCA PAMPAS
234	158315	CURIBAYA	-70.3064	-17.4039	3546	CUENCA LOCUMBA
235	114108	CURPAHUASI	-72.6667	-14.0614	3579	INTERCUENCA ALTO APURÍMAC
236	156121	CUSICANCHA	-75.2963	-13.5040	3253	CUENCA PISCO
237	157418	CUYO CUYO	-69.5490	-14.4889	3414	CUENCA INAMBARI
238	156222	ANDARAPA	-73.3669	-13.5169	2819	CUENCA PAMPAS
239	004450	DOS DE MAYO	-76.7736	-9.7169	3442	INTERCUENCA ALTO MARAÑÓN V
240	000849	EL FRAYLE	-71.1891	-16.1548	4131	CUENCA QUILCA - VITOR - CHILI
241	000684	ANTA ANCACHURO	-72.2158	-13.4683	3340	CUENCA URUBAMBA
242	151208	GORGOR	-77.0403	-10.6214	3025	CUENCA PATIVILCA
243	000607	GRANJA KCAYRA	-71.8753	-13.5569	3219	CUENCA URUBAMBA
244	000359	GRANJA PORCON	-78.6334	-7.0375	3149	CUENCA JEQUETEPEQUE
245	157332	HACIENDA COLINI	-70.8836	-15.6503	4380	CUENCA COATA

246	000372	HACIENDA JOCOS	-78.0003	-7.5169	2720	CUENCA CRISNEJAS
247	153212	HACIENDA LLAGADEN	-78.5836	-7.2669	2150	CUENCA JEQUETEPEQUE
248	153214	HACIENDA LLAUCAN	-78.5336	-6.7503	2566	INTERCUENCA ALTO MARAÑÓN IV
249	000656	HACIENDA TOCAZ	-74.6503	-12.4503	3096	CUENCA MANTARO
250	151509	HORNADA	-75.1003	-13.4336	4530	CUENCA PISCO
251	154112	HUACAMARCANGA	-78.2931	-8.1214	3883	CUENCA SANTA
252	157213	HUACAÑA	-73.8836	-14.1669	3120	CUENCA PAMPAS
253	156303	HUACHIBAMBA	-72.1003	-13.0503	4225	CUENCA URUBAMBA
254	151503	HUACHOS	-75.5335	-13.2206	2744	CUENCA SAN JUAN
255	151602	HUAC-HUAS	-74.9406	-14.1319	3180	CUENCA GRANDE
256	156100	ANTIOQUIA	-76.5000	-12.0833	1839	CUENCA LURÍN
257	000363	HUALGAYOC	-78.6169	-6.7669	3815	INTERCUENCA ALTO MARAÑÓN IV
258	000460	HUALLANCA	-76.9503	-9.8836	3796	INTERCUENCA ALTO MARAÑÓN V
259	000374	HUAMACHUCO	-78.0401	-7.8192	3186	CUENCA CRISNEJAS
260	155209	HUAMANTANGA	-76.7500	-11.5000	3392	CUENCA CHILLÓN
261	156223	HUANCABAMBA	-73.0364	-13.7336	3925	CUENCA PAMPAS
262	000786	HUANCANE	-69.7583	-15.2069	3829	CUENCA HUANCANÉ
263	000665	HUANCAPI	-74.0681	-13.7490	3102	CUENCA PAMPAS
264	156218	HUANCASANCOS	-74.3336	-13.9169	3440	CUENCA PAMPAS
265	151207	HUANCATA	-76.2407	-12.1862	2684	CUENCA MALA
266	000649	HUANCAVELICA	-75.0361	-12.7803	3860	CUENCA MANTARO
267	153327	HUANGACOCKA	-78.0681	-7.9376	3763	CUENCA CRISNEJAS
268	156110	HUANGASCAR	-75.8337	-12.8986	2500	CUENCA CAÑETE
269	000660	HUANTA	-74.2683	-12.9274	2542	CUENCA MANTARO
270	112190	HUANTA GORE	-74.2333	-12.9475	2735	CUENCA MANTARO
271	000404	HUANUCO	-76.2486	-9.9520	1947	INTERCUENCA ALTO HUALLAGA
272	156103	HUAÑEC	-76.1390	-12.2936	3200	CUENCA MALA
273	000646	ACNOCOCHA	-75.0836	-13.2169	4520	CUENCA PAMPAS
274	000787	HUARAYA MOHO	-69.4843	-15.3883	3890	INTERCUENCA 0171
275	155407	HUARAZ	-77.5316	-9.5342	3052	CUENCA SANTA
276	000400	HUARAZ	-77.5249	-9.5165	3052	CUENCA SANTA
277	000553	HUARIACA	-76.1836	-10.4336	3355	INTERCUENCA ALTO HUALLAGA
278	151204	ARAHUAY	-76.7000	-11.6167	2800	CUENCA CHILLÓN
279	000633	HUAROCHIRI	-76.2343	-12.1389	3120	CUENCA MALA
280	155218	HUARIOS	-76.5758	-11.4074	3569	CUENCA CHILLÓN
281	000555	HUASAHUASI	-75.6391	-11.2677	2731	CUENCA PERENE
282	000635	HUAYAO	-75.3205	-12.0402	3316	CUENCA MANTARO
283	000678	HUYRO	-72.4503	-13.0669	2326	CUENCA URUBAMBA
284	158310	ICHUÑA	-70.5427	-16.1411	3815	CUENCA TAMBO
285	000783	ARAPA	-70.1182	-15.1363	3830	INTERCUENCA RAMIS
286	000879	ILAVE	-69.6404	-16.0599	3861	CUENCA ILAVE
287	000765	IMATA	-71.0906	-15.8427	4475	CUENCA QUILCA - VITOR - CHILI
288	000744	INCUYO	-73.5669	-15.2503	3296	CUENCA OCOÑA

289	155231	INGENIO	-75.2966	-11.8752	3373	CUENCA MANTARO
290	000817	ISLA SOTO	-69.4890	-15.5622	3825	LAGO TITICACA
291	000822	ISLA SUANA	-68.8503	-16.3299	3840	LAGO TITICACA
292	000818	ISLA TAQUILE	-69.6911	-15.7702	3863	LAGO TITICACA
293	000456	JACAS CHICO	-76.5003	-9.8836	3673	INTERCUENCA ALTO HUALLAGA
294	157403	JARPANA	-70.7836	-15.5169	4361	CUENCA COATA
295	000503	JAUIJA	-75.4869	-11.7866	3363	CUENCA MANTARO
296	000391	JESUS	-78.3884	-7.2457	2564	CUENCA CRISNEJAS
297	154101	JULCAN	-78.4861	-8.0423	3385	CUENCA MOCHE
298	000880	JULI	-69.4599	-16.2038	3830	INTERCUENCA 0157
299	007454	JULIACA	-70.2079	-15.4442	3830	CUENCA COATA
300	000754	LA ANGOSTURA	-71.6512	-15.1825	4258	INTERCUENCA ALTO APURÍMAC
301	155106	ACO	-77.2336	-10.5003	2490	CUENCA PATIVILCA
302	117054	ARICOTA	-70.2373	-17.3291	2837	CUENCA LOCUMBA
303	157323	LA CALERA	-72.0169	-15.2836	4540	CUENCA CAMANÁ
304	155120	LA CRUZADA	-76.4003	-11.0336	4600	CUENCA MANTARO
305	153331	LA ENCAÑADA	-78.3331	-7.1233	2980	CUENCA CRISNEJAS
306	153218	LA LLICA	-78.5169	-6.7836	2919	INTERCUENCA ALTO MARAÑÓN IV
307	000604	LA OROYA	-75.9576	-11.5718	3861	CUENCA MANTARO
308	000839	LA PAMPILLA	-71.5345	-16.4138	2326	CUENCA QUILCA - VITOR - CHILI
309	000663	LA QUINUA	-74.1416	-13.0555	3209	CUENCA MANTARO
310	000318	LA VICTORIA	-78.4595	-7.1912	2630	CUENCA CRISNEJAS
311	155217	LACHAQUI	-76.6284	-11.5533	3670	CUENCA CHILLÓN
312	155206	LAGUNA SURASACA	-76.7867	-10.5238	4385	CUENCA HUAURA
313	000763	LAGUNILLAS	-70.6613	-15.7680	4250	CUENCA COATA
314	000393	ASUNCION	-78.5158	-7.3260	2270	CUENCA JEQUETEPEQUE
315	000642	LAIVE	-75.3605	-12.2523	3829	CUENCA MANTARO
316	000356	LAJAS	-78.7453	-6.5589	2168	CUENCA CHAMAYA
317	157223	LAMPA	-73.3479	-15.1847	2790	CUENCA OCOÑA
318	000779	LAMPA	-70.3742	-15.3611	3866	CUENCA COATA
319	151214	LANGA	-76.4211	-12.1265	2863	CUENCA LURÍN
320	156128	LARAMARCA	-75.0336	-13.9503	3350	CUENCA GRANDE
321	000889	LARAQUERI	-70.0459	-16.1360	3933	CUENCA ILAVE
322	158209	LAS SALINAS	-71.1481	-16.3178	4378	CUENCA QUILCA - VITOR - CHILI
323	000659	ACOBAMBA	-74.5616	-12.8436	3371	CUENCA MANTARO
324	000657	LIRCAY	-74.7181	-12.9819	3360	CUENCA MANTARO
325	157320	LIVITACA	-71.6836	-14.3169	3741	INTERCUENCA ALTO APURÍMAC
326	000761	LLALLY	-70.8805	-14.9529	3985	CUENCA PUCARÁ
327	157212	AUCARA	-73.9669	-14.2836	3157	CUENCA PAMPAS
328	000388	LLAPA	-78.8112	-6.9783	2932	CUENCA JEQUETEPEQUE
329	000461	LLATA	-76.7836	-9.5503	3239	INTERCUENCA ALTO MARAÑÓN V
330	157200	LLAUTA	-74.9256	-14.2464	2445	CUENCA GRANDE
331	000706	LOS LIBERTADORES	-74.9169	-13.4003	4200	CUENCA PAMPAS

332	000782	LOS UROS	-69.9741	-15.8204	3819	LAGO TITICACA
333	156209	LURICOCHA	-74.2336	-12.8169	3790	CUENCA MANTARO
334	000304	AUGUSTO WEBERBAUER	-78.4931	-7.1675	2673	CUENCA CRISNEJAS
335	000679	MACHU PICCHU	-72.5459	-13.1666	2399	CUENCA URUBAMBA
336	000777	MACUSANI	-70.4221	-14.0668	4358	CUENCA INAMBARI
337	157317	MADRIGAL	-71.8065	-15.6100	3276	CUENCA CAMANÁ
338	000392	MAGDALENA	-78.6526	-7.2535	1307	CUENCA JEQUETEPEQUE
339	150901	MALVAS	-77.6552	-9.9272	2979	CUENCA HUARMEY
340	155122	AUTISHA	-76.6110	-11.7381	2181	CUENCA RÍMAC
341	000820	MAÑAZO	-70.3395	-15.7994	3924	CUENCA ILPA
342	000549	MARCAPOMACOCHA	-76.3249	-11.4044	4447	CUENCA MANTARO
343	153222	MATARA	-78.2336	-7.2169	2922	CUENCA CRISNEJAS
344	000548	MATUCANA	-76.3780	-11.8391	2417	CUENCA RÍMAC
345	150903	MAYORARCA	-77.3489	-10.1578	3335	CUENCA FORTALEZA
346	000878	MAZO CRUZ	-69.7154	-16.7389	3975	CUENCA ILAVE
347	155514	MILLOC	-76.3499	-11.5713	4384	CUENCA RÍMAC
348	150902	MILPO	-77.2333	-9.8833	4400	CUENCA SANTA
349	155113	MINA COLQUI	-76.4836	-11.5836	4600	CUENCA RÍMAC
350	152170	MINAS CAYLLOMA	-71.8503	-15.1503	4889	INTERCUENCA ALTO APURÍMAC
351	154106	MOLLEPATA	-77.9540	-8.1914	2708	CUENCA SANTA
352	000680	MOLLEPATA	-72.5503	-13.5169	2601	INTERCUENCA ALTO APURÍMAC
353	156104	AYAVIRI	-76.1367	-12.3821	3224	CUENCA MALA
354	000776	AYAVIRI	-70.5931	-14.8719	3928	CUENCA PUCARÁ
355	000785	MUÑANI	-69.9518	-14.7669	3948	CUENCA HUANCANÉ
356	000321	NAMORA	-78.3278	-7.2006	2744	CUENCA CRISNEJAS
357	157404	NUÑO A	-70.6336	-14.4836	4013	CUENCA AZÁNGARO
358	000781	AZANGARO	-70.1905	-14.9172	3851	CUENCA AZÁNGARO
359	156133	OBRAJILLO	-76.6221	-11.4526	2468	CUENCA CHILLÓN
360	157202	OCAÑA	-74.8169	-14.4003	2760	CUENCA GRANDE
361	155105	OCROS	-77.4003	-10.4046	3249	CUENCA PATIVILCA
362	000695	OLLACHEA	-70.4971	-13.8039	2850	CUENCA INAMBARI
363	154200	ONGON	-76.9669	-8.2669	2440	INTERCUENCA ALTO HUALLAGA
364	157311	ORCOPAMPA	-72.3386	-15.2606	3779	CUENCA CAMANÁ
365	157407	ORURILLO	-70.5169	-14.7336	4048	CUENCA AZÁNGARO
366	157201	OTOCA	-74.6836	-14.4836	2055	CUENCA GRANDE
367	158302	OTORA	-70.8536	-17.0072	2580	CUENCA ILO - MOQUEGUA
368	000361	OTUZCO	-78.5669	-7.9003	2660	CUENCA MOCHE
369	000541	OYON	-76.7687	-10.6677	3667	CUENCA HUAURA
370	155200	PACCHO	-76.9333	-10.9500	3250	CUENCA HUAURA
371	000634	PACHACHACA	-76.0169	-11.6169	4000	CUENCA MANTARO
372	155207	PACHAMACHAY	-76.9141	-11.0652	3175	CUENCA HUAURA
373	155204	PACHANGARA	-76.8169	-10.7836	3583	CUENCA HUAURA

374	158305	PACHAS	-70.7336	-16.4669	3857	CUENCA TAMBO
375	157217	PAICO	-73.6669	-14.0336	3589	CUENCA PAMPAS
376	158321	PALCA	-69.9586	-17.7779	3004	CUENCA CAPLINA
377	000735	PALCACHACRA	-74.2836	-14.5336	3300	CUENCA ACARÍ
378	155205	PALLAC	-76.8080	-11.3487	2367	CUENCA CHANCAY - HUARAL
379	A01	YURACMAYO	-76.1578	-11.8395	4311	CUENCA RÍMAC
380	13404001	CONDOROMA	-71.2900	-15.4000	4287	CUENCA CAMANÁ
381	A03	TACALAYA	-70.4167	-17.0000	4452	CUENCA LOCUMBA
382	158306	CAMILACA	-70.4333	-17.2667	3300	CUENCA LOCUMBA

Anexo 2.

Registro de series de precipitación mensual en las estaciones seleccionadas.

Los datos se encuentran en: <https://cutt.ly/SAiOCuA>