

SOCIEDAD CHILENA DE INGENIERÍA HIDRÁULICA
XXVI CONGRESO CHILENO DE INGENIERÍA HIDRÁULICA

**UN PRODUCTO PROBABILÍSTICO DIARIO DE PRECIPITACIÓN Y
TEMPERATURA PARA CHILE**

DIEGO PINTO ^{1,2}
PABLO MENDOZA ^{3,4}
MIGUEL LAGOS ³

RESUMEN

En las últimas décadas, los productos meteorológicos grillados han ganado relevancia en ciencias del clima e hidrología, utilizándose diversas metodologías para su desarrollo a partir de observaciones, modelos climáticos globales y productos satelitales. En Chile, se han desarrollado productos como CR2Met y RF-MEP durante los últimos años, los cuales no consideran explícitamente una caracterización de la incertidumbre en sus estimaciones. En este trabajo, se presenta un producto grillado probabilístico diario de precipitación y temperatura para Chile. Para ello, se utiliza la herramienta *Gridded Meteorological Ensemble Tool* (GMET), que utiliza regresiones múltiples ponderadas localmente a partir de observaciones y predictores fisiográficos, obteniéndose un set de forzantes con 30 miembros del conjunto para el periodo 1988-2021. Se evalúa además la bondad de los modelos mediante el método de validación cruzada *Leave One Out*. Resultados preliminares en validación cruzada muestran un buen desempeño de la precipitación y temperatura a escala anual ($R^2 > 0,8$), con sesgos húmedos en las probabilidades de ocurrencia para eventos con $PoP > 0,5$. Los patrones climáticos de la zona son bien representados en general, encontrándose diferencias localizadas con respecto al CR2Met y ERA5. Existen sesgos secos en la zona Austral y cálidos en la zona cordillerana del norte. Se concluye que la metodología es aplicable en Chile, aunque existen oportunidades de mejora con respecto a la representatividad espacial.

¹ Ingeniero de proyectos en el Advanced Mining Technology Center (AMTC) de la Universidad de Chile.

² Estudiante de magíster de recursos hídricos del departamento de Ingeniería Civil de la Universidad de Chile.

³ Investigador Asociado, Advanced Mining Technology Center (AMTC) de la Universidad de Chile.

⁴ Profesor Asociado, Departamento de Ingeniería Civil, Universidad de Chile.

1. INTRODUCCIÓN

Los productos meteorológicos grillados han revolucionado el análisis del clima y el tiempo, logrando caracterizar la evolución espacio-temporal de variables meteorológicas (e.g., Rienecker et al., 2011; Beck et al., 2017; Tang et al., 2021; Báez-Villanueva et al., 2020). Además, son herramientas esenciales para diversas aplicaciones, incluyendo la predicción del tiempo (e.g., Saha et al., 2006; Johnson et al., 2019), modelación climática (e.g., Hersbach et al., 2020), y la predicción hidrológica (e.g., Araya et al., 2023), entre otras.

La creación de productos meteorológicos grillados puede involucrar una combinación de métodos de interpolación espacial, modelación numérica y técnicas de asimilación de datos (Sun et al., 2018). Además, requiere de datos observacionales de diversas fuentes, incluyendo estaciones meteorológicas, satélites e instrumentos de teledetección, los cuales son sometidos a un control de calidad para garantizar su precisión y consistencia (e.g., Beck et al., 2017). Las técnicas de interpolación espacial, como el *krigging* (e.g. Collinao, 2022) y ponderación por inverso de la distancia se utilizan comúnmente para estimar variables meteorológicas en ubicaciones no muestreadas. A menudo, se emplean métodos estadísticos (Clark y Slater, 2006; Newman et al., 2015, 2019) y algoritmos de aprendizaje automático (RFMEP; Báez-Villanueva et al., 2020) para mejorar la calidad y la representación espacial de los productos espacialmente distribuidos.

Si bien se han generado diversos productos meteorológicos grillados para Chile (e.g., Boisier et al., 2018; Báez-Villanueva et al., 2020; Castro et al., 2014), ninguno ha considerado la incertidumbre asociada a la metodología o a los datos utilizados. En este trabajo se generan estimaciones probabilísticas de precipitación diaria y temperaturas extremas diarias utilizando la herramienta *Gridded Meteorological Ensemble Tool* (GMET de ahora en adelante; Newman et al., 2019). Para el desarrollo del producto, se plantean el uso de dos criterios fundamentales: (1) una alta capacidad predictiva en puntos con información disponible, y (2) capacidad de reproducir los principales patrones climáticos dentro del área de estudio. Se buscan responder las siguientes preguntas científicas:

1. ¿Cómo es su desempeño con respecto a otros productos existentes?
2. ¿Cuál es el desempeño en alta montaña?
3. ¿Es suficiente la red actual de estaciones meteorológicas para el desarrollo del producto?

2. ZONA DE ESTUDIO Y DISPONIBILIDAD DE INFORMACIÓN

La zona de estudio corresponde al territorio nacional comprendido entre las latitudes 17°S y 56°S (i.e., Chile continental). La Figura 1 muestra la elevación media y valores promedio de temperatura media anual y precipitación anual. La elevación proviene del modelo de elevación digital (MDE) MERIT Hydro (Yamazaki et al., 2019), y las climatologías de precipitación y temperatura del producto grillado CR2Met (Boisier et al., 2018). Los gradientes están altamente influenciados por la presencia de la cordillera de los Andes, disminuyendo la temperatura y aumentando la precipitación hacia el este con elevaciones por sobre los 3000 msnm. Además, existe un marcado gradiente latitudinal que impacta en disminución de las temperaturas y aumento de la precipitación hacia el sur (Aceituno et al., 2021).

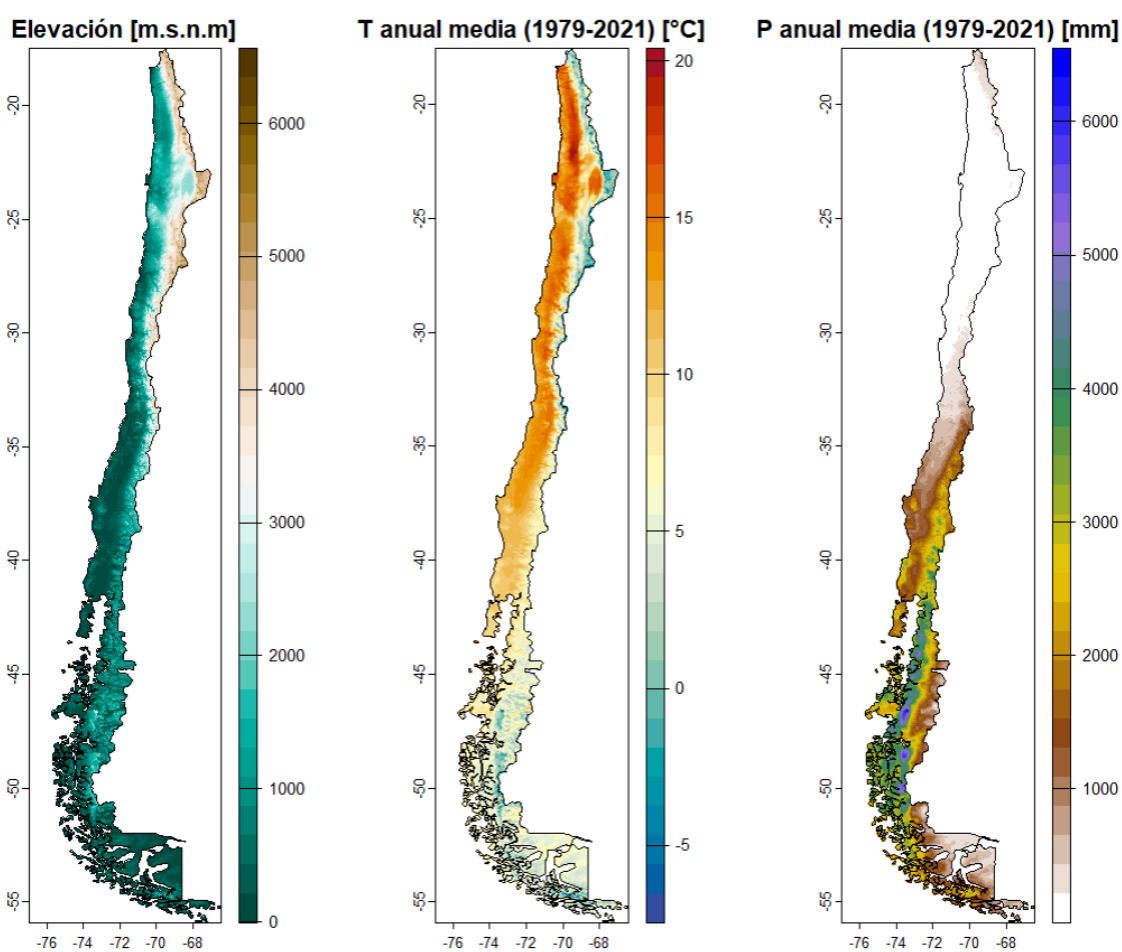


Figura 1. Elevación [m s.n.m.], temperatura media anual [°C] y precipitación anual promedio [mm] de la zona de estudio (periodo 1979-2021) a partir del DEM MERIT-Hydro y CR2Met.

Los datos utilizados provienen de estaciones meteorológicas de la Dirección General de Aguas (DGA) y de la Dirección Meteorológica de Chile (DMC). La disponibilidad temporal de observaciones de precipitación, temperatura mínima y máxima se muestra en la Figura 2,

donde se aprecia que la densidad espacial de la red observacional disminuye considerablemente para las variables de temperatura.

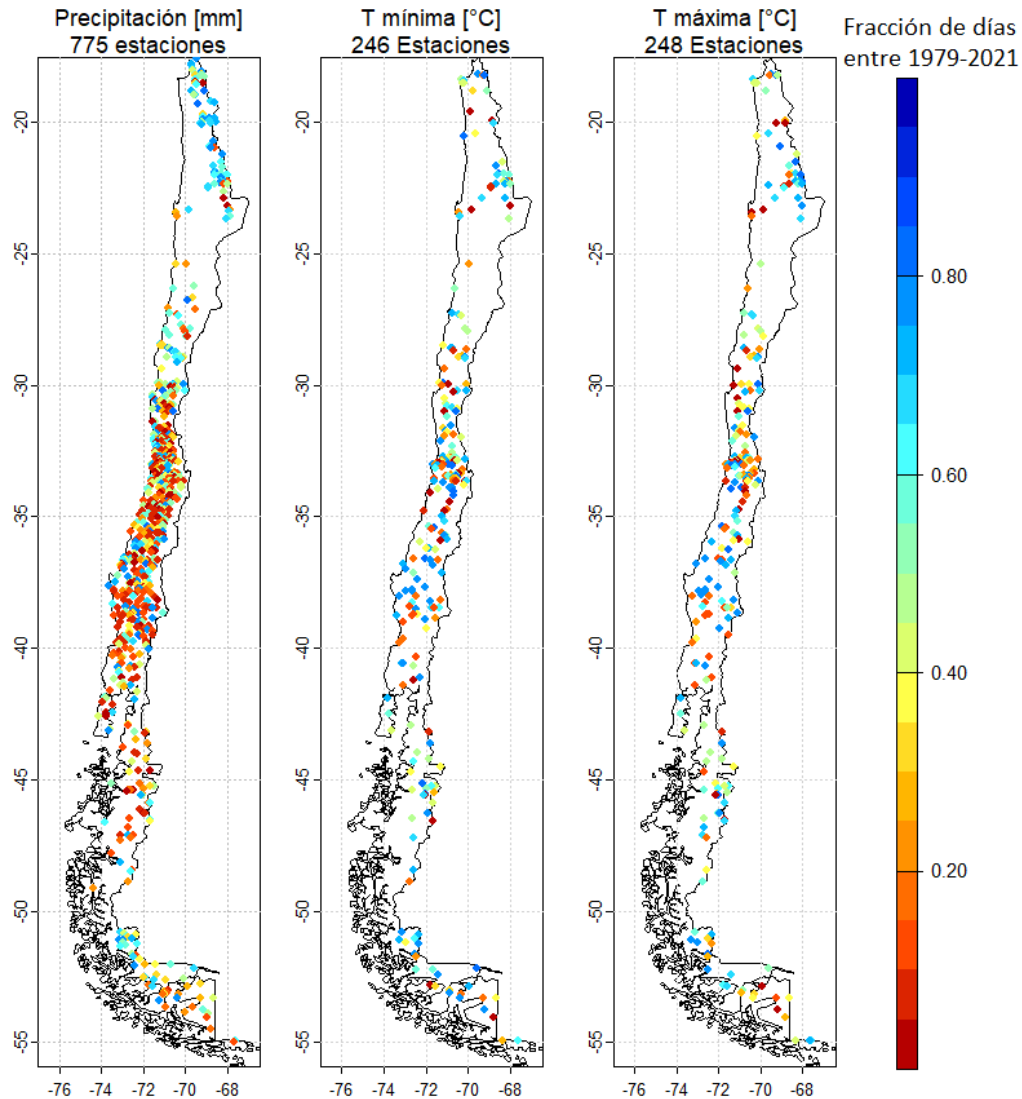


Figura 2: Fracción de días con observaciones en la red meteorológica de estaciones de la DGA y DMC durante el periodo 1979-2021.

3. METODOLOGÍA

La Figura 3 muestra un esquema general de la metodología adoptada en este trabajo, la cual contempla los siguientes pasos: (1) relleno de datos, (2) recopilación de información de los predictores, (3) estimaciones de precipitación y temperatura, y (4) generación del producto grillado probabilístico de conjuntos. El procedimiento detallado de cada fase se describe en las subsecciones siguientes.

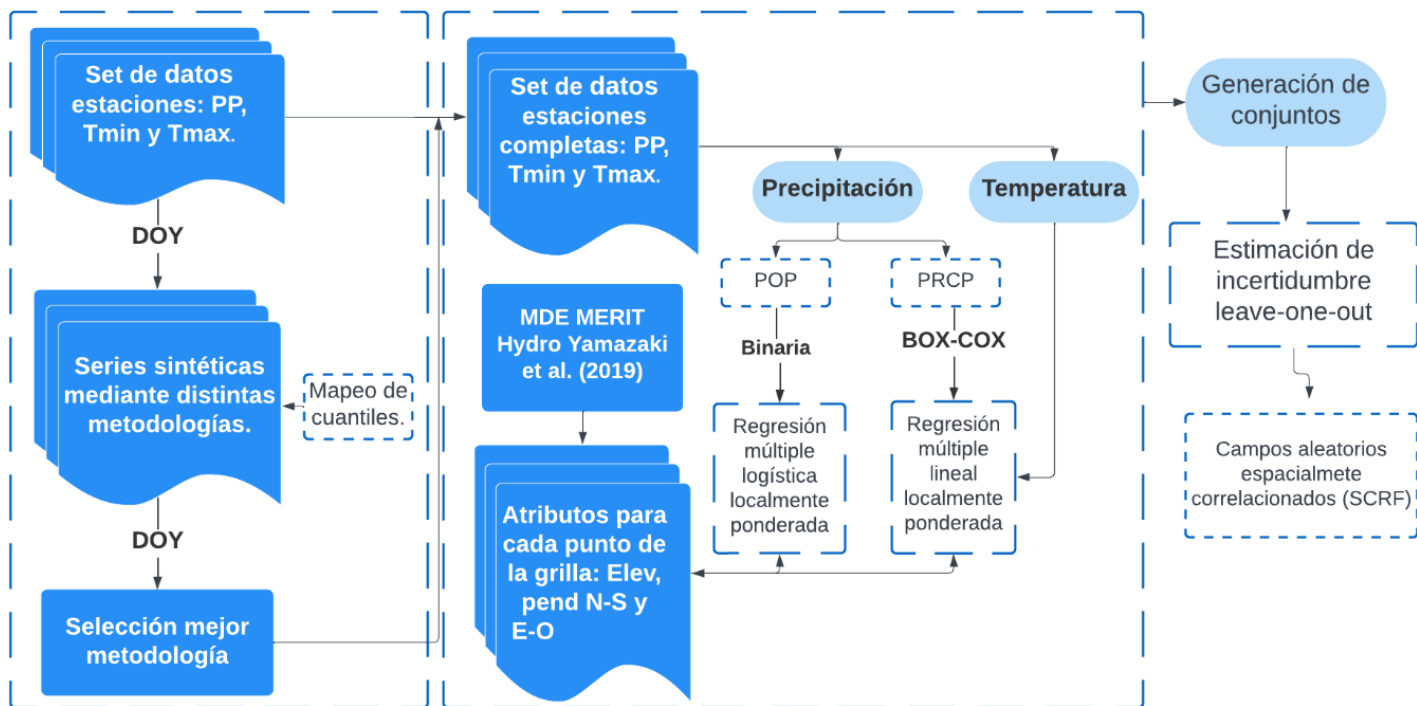


Figura 3. Esquema metodológico del relleno de datos y desarrollo del producto grillado.

3.1. Selección de estaciones y relleno de datos

Del set de datos inicial de estaciones, se seleccionan aquellas que tienen más de 8 años de información, posteriormente se utilizan técnicas de relleno automáticas de mapeo de cuantiles con estaciones vecinas testeadas por Tang et al. (2021). La primera de ellas corresponde al relleno con la estación vecina que tenga mejor coeficiente de correlación de Pearson (QMN1), la segunda y tercera corresponden a una ponderación de los rellenos de todas las estaciones vecinas ponderadas por coeficiente de correlación (QMN2) y por la distancia (QMN3), y finalmente, la cuarta corresponde a la mediana de los tres métodos mencionados anteriormente. Se considera un radio de 100 km. para buscar estaciones vecinas, y el relleno se realiza por día del año (DOY; por sus siglas en inglés) con una ventana temporal 31 días. Es decir, para rellenar el 16 de enero, se consideran las series entre el 1 y 31 de enero de todo el período de estudio (1979-2021). Finalmente, para cada día del año se selecciona la mejor metodología según el desempeño en cuanto al coeficiente de correlación. Lo anterior con el fin de no generar inconsistencias en el algoritmo GMET (Newman et al., 2015), que requiere estaciones completas para las regresiones.

3.2. Generación del producto grillado

La base metodológica de GMET fue originalmente propuesta por Clark & Slater (2006) para la estimación de precipitación en Colorado. Desde entonces se han generado diversos productos grillados a escala local (e.g., Newman et al., 2015) y global (e.g., Tang et al., 2021). Para la interpolación espacial de la precipitación, temperatura mínima y máxima, GMET

considera el uso de regresiones lineales múltiples ponderadas localmente (Ecuación 1). Para la precipitación diaria, se estima la probabilidad de ocurrencia (PoP) mediante regresiones logísticas ponderadas localmente (Ecuación 2); posteriormente, se modela el monto aplicando una transformación de Box-Cox (Box & Cox, 1964; Ecuación 3) debido a la poca normalidad de la variable (Newman et al. 2015; Tang et al. 2022). En todos los casos, se utilizan como predictores la latitud, longitud, elevación, pendiente norte-sur y este-oeste provenientes del MDE MERIT-Hydro (Yamazaki et al., 2019), considerando una grilla de 0,05° lat-lon (~5km). Para ajustar los coeficientes de las ecuaciones de regresión, se consideran las estaciones dentro de un radio de 100 kilómetros, buscando al menos 25 estaciones. Si en ese radio no existen suficientes estaciones cercanas, este se va expandiendo hasta alcanzar el número de estaciones mínimas

$$VAR_{i\ grilla} = Z_{i\ grilla} \gamma^a \quad (1)$$

$$PoP_{i\ grilla} = \frac{1}{1 + \exp(-Z_{i\ grilla} \gamma)} \quad (2)$$

$$P'_{estación} = \frac{(P^{\frac{1}{\lambda}} - 1)}{\frac{1}{\lambda}} \quad (3)$$

Donde VAR puede ser la precipitación mediante transformación de Box-Cox, la temperatura mínima o máxima. Z representa la matriz de predictores y γ los coeficientes de las regresiones múltiples.

La generación de conjuntos viene de la utilización de campos aleatorios espacialmente correlacionados (SCRF; por sus siglas en inglés) generados con la distribución estándar (Clark & Slater, 2006) para cada día y variable. Específicamente, se muestrean valores progresivamente, considerando aquellos generados por los miembros del conjunto anteriores, lo que permite mantener la correlación espacial entre puntos de la cuadrícula.

En este trabajo, se evalúa la calidad del relleno de datos mediante la eficiencia de (KGE; Gupta *et al.*, 2009; Ecuaciones 4 a 7), el coeficiente de determinación (R^2) y el porcentaje de sesgo, calculados con respecto a las observaciones. Además, se evalúa el desempeño del producto en validación cruzada en los puntos con observación a través de diagramas de confiabilidad (Hamill, 1997) y diagramas de discriminación (Clark & Slater, 2006). En el diagrama de confiabilidad, se comparan las probabilidades de ocurrencia de precipitación (i.e., $P > 0$ mm) estimadas con las frecuencias observadas para dichas predicciones, siendo el caso ideal un ajuste 1:1. Por ejemplo, si se estiman probabilidades de ocurrencia iguales a 0,5, se espera que la precipitación sea mayor que cero durante la mitad de los días para los que se realizó dicha predicción. Por otro lado, el diagrama de discriminación busca comparar las funciones de densidad de probabilidad para predicciones probabilísticas de eventos de ocurrencia de precipitación ($PP > 0$) vs. días con $PP = 0$. Es decir, para los eventos de lluvia ($PP > 0$), se espera que la frecuencia relativa observada sea baja cuando las probabilidades estimadas sean cercanas a 0, en cambio, se espera tener frecuencias relativas observadas altas cuando se estiman altas probabilidades de ocurrencia. El análisis para los días sin precipitación es inverso, se espera una frecuencia relativa alta (baja) observada para probabilidades de lluvia bajas (altas). Se considera que un sistema predictivo tiene buena discriminación si las dos curvas de densidad de probabilidad se encuentran bien distanciadas.

$$KGE = 1 - ED \quad (4)$$

$$ED = \sqrt{(r - 1)^2 + (\sigma - 1)^2 + (\beta - 1)^2} \quad (5)$$

$$\sigma = \frac{\sigma_s}{\sigma_o} \quad (6)$$

$$\beta = \frac{\mu_s}{\mu_o} \quad (7)$$

Donde r es el coeficiente de correlación lineal, σ representa la variabilidad mediante el cociente entre de la desviación estándar simulada y observada, y β representa el sesgo mediante el cociente entre las medias de las series observadas y simuladas.

Finalmente, se compara la climatología generada con GMET (utilizando el promedio de los 30 miembros) para el periodo 1988-2021 con la de los productos grillados CR2Met v2.5 y ERA5.

4. RESULTADOS

Las distintas metodologías de relleno de datos muestran un buen desempeño con respecto a las observaciones, alcanzándose valores de KGE y R^2 generalmente mayores a 0,8 y sesgos cercanos a 0 (Figura 4). Para la mayoría de las métricas, el método que utiliza la estación vecina con mayor correlación (QMN1) es la que muestra mejores resultados, sin embargo, disminuyen los sesgos cuando se utiliza una ponderación por el coeficiente de correlación (QMN2).

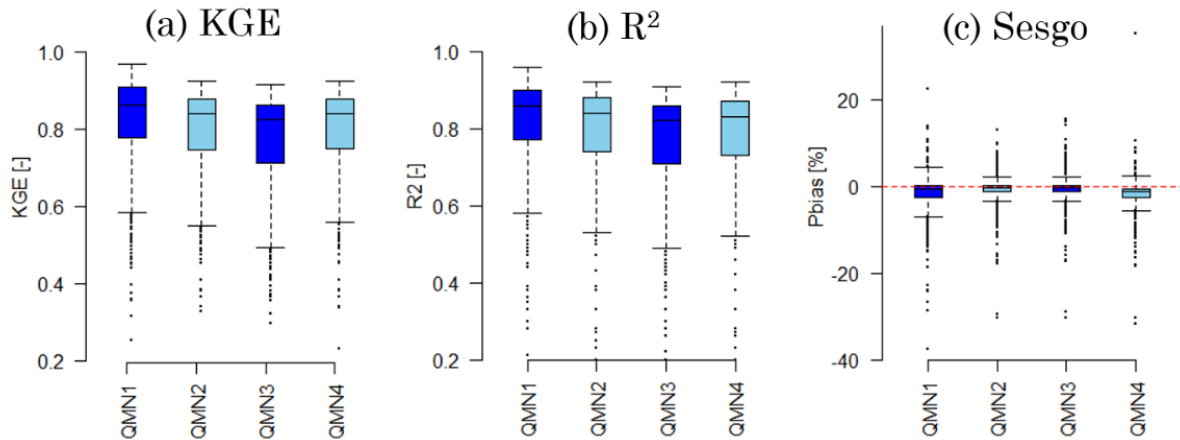


Figura 4. Métricas de evaluación de las series generadas a partir de mapeo de cuantiles para el periodo 1979-2021: (a) eficiencia de Kling-Gupta, (b) coeficiente de determinación, y (c) sesgo. Las cajas provienen de la métrica obtenida entre la serie sintética y observada de cada estación.

En validación cruzada, las precipitaciones tienen un adecuado desempeño ($R^2 > 0,7$), aumentando la dispersión para precipitaciones por sobre los 3000mm. En el caso de la temperatura, la mayor densidad de puntos se encuentra en la línea 1:1, teniendo mejor desempeño para la temperatura máxima ($R^2 = 0,8$) que para la temperatura mínima ($R^2 = 0,75$). Para el caso de la temperatura, existen puntos que se alejan de la línea 1:1, lo que puede ser

atribuido a la baja densidad de estaciones de medición de dicha variable (Figura 2), forzando ajustes de los modelos de regresión con estaciones que se encuentran alejadas.

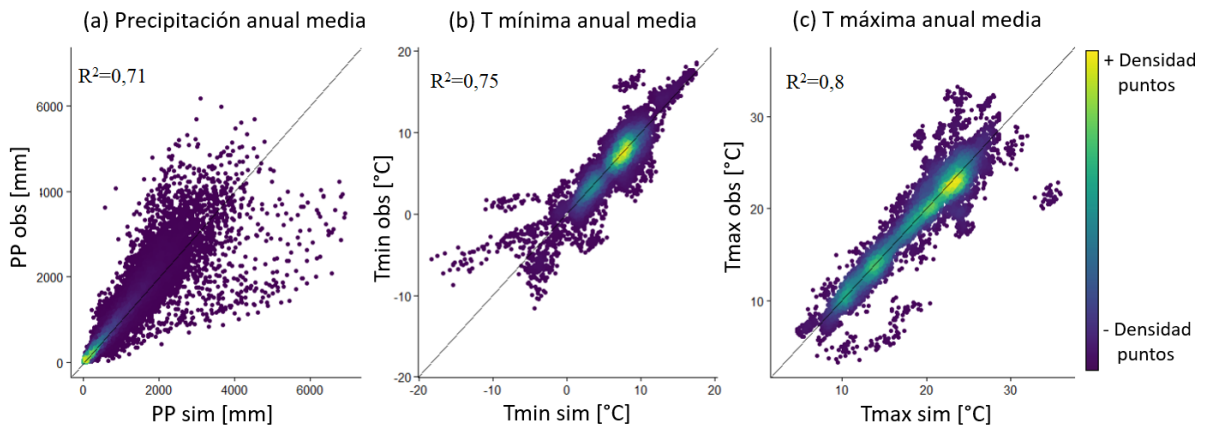


Figura 5. Observaciones versus estimaciones de la precipitación, temperatura mínima y máxima a escala anual con respecto a las observaciones. La escala de colores indica que los sectores con puntos más claros (amarillos) tienen mayor densidad de puntos solapados.

La Figura 6a muestra diagramas de confiabilidad para eventos con $P > 0$ mm. La línea azul se separa de la línea 1:1 a partir de los eventos con probabilidad mayor a 0,5, indicando que el producto tiene un sesgo húmedo. Por otro lado, la Figura 6b muestra que existe una buena discriminación de eventos; sin embargo, para los eventos de lluvia, la línea roja empieza a aumentar levemente la frecuencia relativa observada a partir de probabilidades de 0,5, lo que se condice con el sesgo húmedo ya indicado.

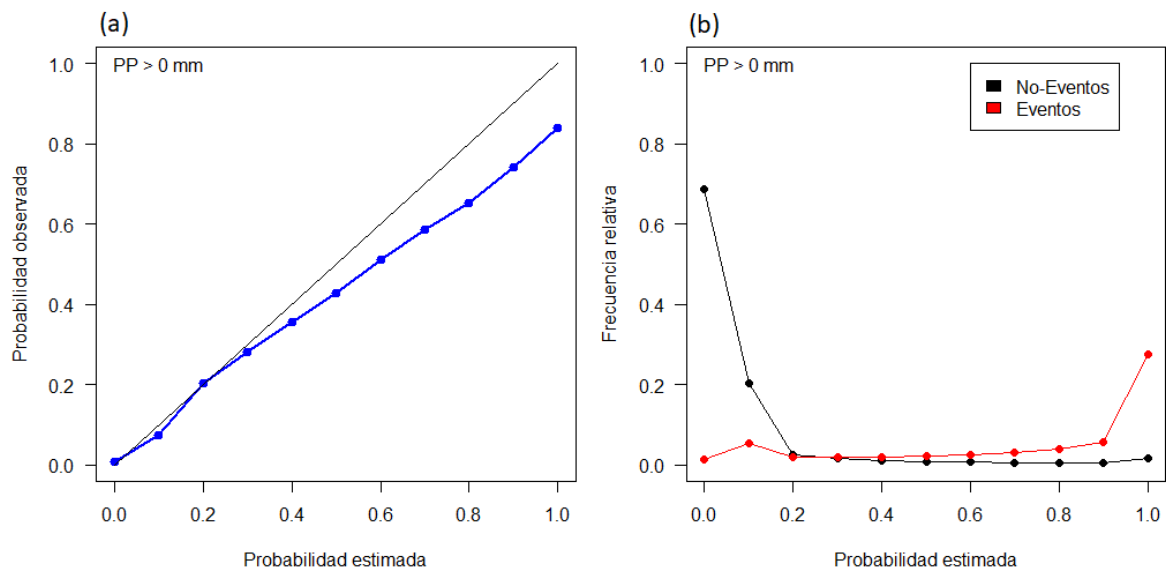


Figura 6. Diagrama de confiabilidad (a) y de discriminación (b) de la estimación de la probabilidad de ocurrencia en validación cruzada.

Las Figuras 7a y 8a muestran las distribuciones espaciales de los promedios climatológicos para la precipitación anual y la temperatura media anual (obtenida a partir del promedio

simple entre la máxima y la mínima), que concuerdan con los gradientes observados: la precipitación aumenta hacia la cordillera y el sur, mientras que para la temperatura se invierten los gradientes aumentando hacia el norte y la costa.

Los paneles (b), (c) y (d) de las Figuras 7 y 8 muestran una comparación del producto generado con respecto a CR2Met y ERA5. En cuanto a la precipitación (Figura 7), las mayores diferencias se dan en la zona Austral alrededor de los 45°S, las que alcanzan valores de hasta 5000 mm/año debido a que GMET no reconoce un aumento de la precipitación por sobre los 4000 mm que muestra el CR2Met, proveniente de otros predictores del reanálisis ERA5 (Figura 8d).

El producto meteorológico generado aquí sigue la misma distribución espacial de temperatura de CR2Met y ERA5 (Figura 8), con la excepción de una zona cálida en la cordillera alrededor de los 25°S, en la cual las diferencias alcanzan los 10°C. En dicha zona, no existen estaciones muy cercanas, razón por la que probablemente GMET estima mayores temperaturas. En el caso del CR2Met, el uso de predictores extraídos de ERA5 permite suplir la falta de observaciones en varias partes del dominio para obtener gradientes de temperatura y precipitación.

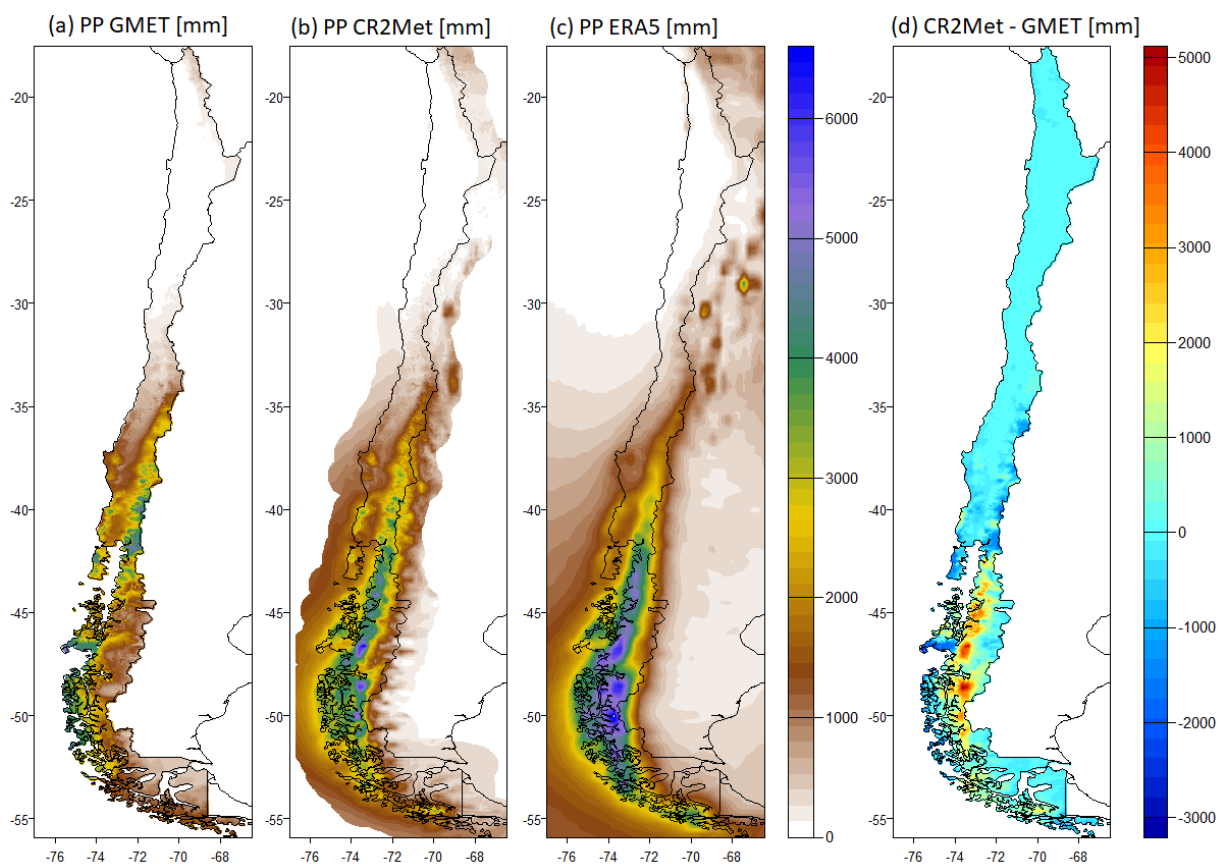


Figura 7. Precipitación anual media entre 1988-2021 de (a) GMET, (b) CR2Met, (c) ERA5, y (d) comparación entre GMET y CR2Met.

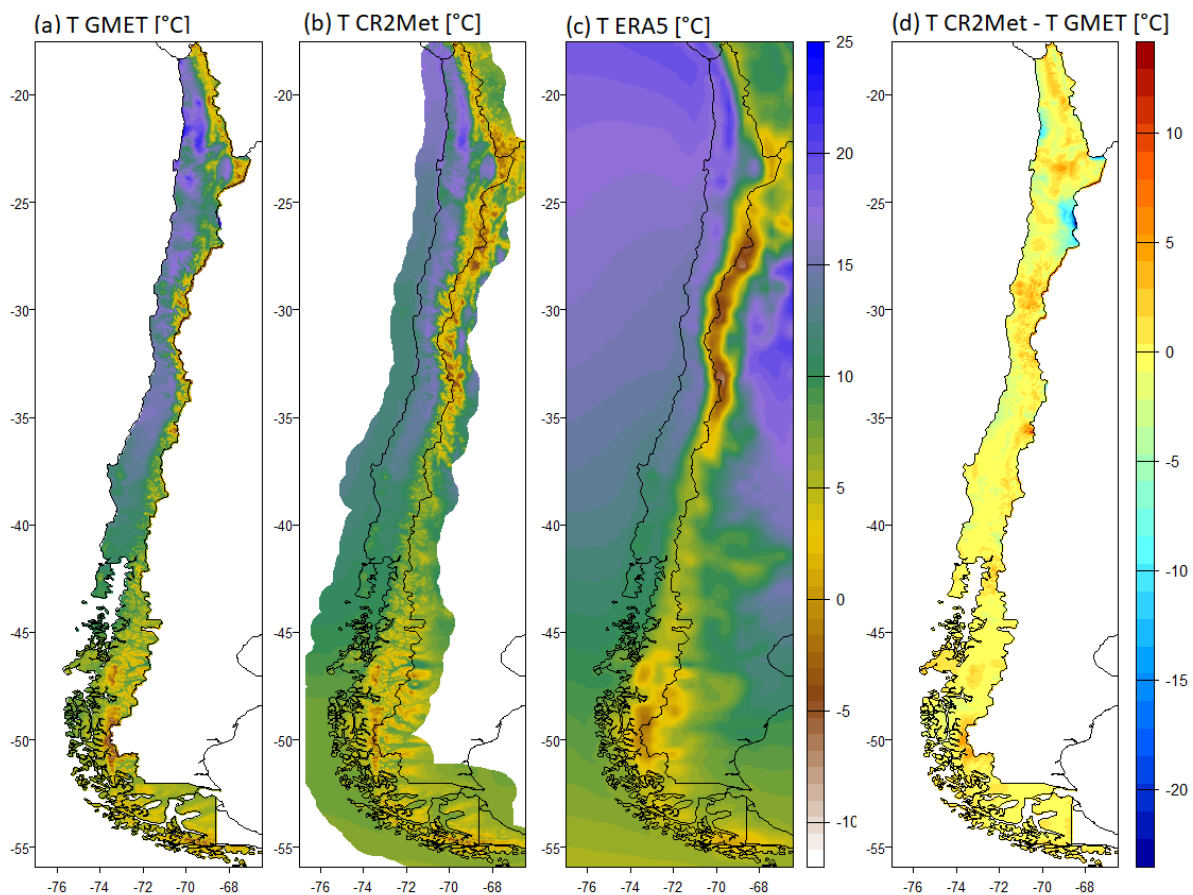


Figura 8. Temperatura media anual entre 1988-2021 de (a) GMET, (b) CR2Met, (c) ERA5, y (d) comparación entre GMET y CR2Met.

5. CONCLUSIONES

En primer lugar, el relleno de datos con mapeo de cuantiles tiene un buen desempeño con respecto a las observaciones, obteniéndose valores de KGE y R^2 mayores a 0,8 para muchas estaciones. La cordillera del norte (25°S) es la zona con menor densidad de estaciones, lo que es una limitante para la estimación de precipitación y temperatura. Existe espacio para implementar otras metodologías de relleno como métodos de interpolación y de aprendizaje automático.

La precipitación estimada tiene un buen desempeño en validación cruzada a escala anual ($R^2 > 0,7$) en puntos con información. Existe un sesgo húmedo en las probabilidades de precipitación para eventos con PoP mayores a 0,5, entregando valores altos a eventos que en realidad no ocurrieron. Con respecto a CR2Met y ERA5, las mayores diferencias se dan en la zona Austral, mostrando menores precipitaciones con la metodología GMET.

Para la temperatura mínima y máxima a escala anual en validación cruzada se obtuvieron R^2 mayores a 0,7, con algunas estaciones que se alejan de la línea 1:1. Lo que se debe principalmente a la menor densidad de estaciones de temperatura. Con respecto a la

distribución espacial, los gradientes norte-sur y este-oeste son bien representados, con sesgos cálidos locales de hasta 7°C en la zona cordillerana alrededor de la latitud 25°S.

Hay oportunidades de mejora con respecto al relleno de datos y otras decisiones, como por ejemplo la transformación de la precipitación. Además, se evalúa la posibilidad de utilizar predictores provenientes de otros productos globales para suplir la falta de observaciones en distintas zonas del país.

Los resultados de este estudio muestran que la herramienta GMET tiene mucho potencial para generar un producto grillado probabilístico para Chile. Esto resulta un gran avance en la representación de la meteorología del país, buscando cuantificar la incertidumbre en base a la calidad y cantidad de observaciones disponibles. El trabajo busca obtener un producto que por sí mismo logre evidenciar la variabilidad posible en las condiciones climáticas, siendo un insumo para que los usuarios puedan enfrentar distintas posibilidades según los alcances de sus estudios. Se pretende entregar recomendaciones por parte de los autores para su uso en función de la incertidumbre obtenida en las distintas macrozonas del país.

AGRADECIMIENTOS

Los autores agradecen el financiamiento del proyecto “Actualización de modelos para pronóstico de caudales de deshielo integrando la red nival D.G.A y su disposición en aplicación informática” por parte de la DGA a la división de hidrología del AMTC.

Esta investigación fue parcialmente apoyada por la infraestructura de supercómputo del NLHPC.

REFERENCIAS

Aceituno, Patricio, et al. "Climate and weather in Chile." *Water resources of Chile* (2021): 7-29.

Boisier, J. P., Álvarez-Garretón, C., Cepeda, J., Osses, A., Vásquez, N., y Rondanelli, R. (2018, April). CR2MET: A high-resolution precipitation and temperature dataset for hydroclimatic research in Chile. In *EGU General Assembly Conference Abstracts* (p. 19739).

Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Beck, H. E., McNamara, I., Ribbe, L., Nauditt, A., Birkel, C., Verbist, K., Giraldo-Osorio, J. y Thinh, N. X. (2020). RF-MEP: A novel Random Forest method for merging gridded precipitation products and ground-based measurements. *Remote Sensing of Environment*, 239, 111606.

Box, G. E. P. & Cox, D. (1964). An analysis of transformation, *Journal of the Royal Statistical Society, Series B* 26, 211-252.

Castro, L. M., Gironás, J., & Fernández, B. (2014). Spatial estimation of daily precipitation in regions with complex relief and scarce data using terrain orientation. *Journal of Hydrology*, 517, 481-492.

Clark, M. P., y Slater, A. G. (2006). Probabilistic quantitative precipitation estimation in complex terrain. *Journal of Hydrometeorology*, 7(1), 3-22.

Collinao Arenas, L. R. (2022). Uso de métodos geoestadísticos para la regionalización de variables meteorológicas.

Hamill, T. M. (1997). Reliability diagrams for multicategory probabilistic forecasts. *Weather and forecasting*, 12(4), 736-741.

Hersbach H, Bell B, Berrisford P, et al. The ERA5 global reanalysis. *QJR Meteorol Soc*. 2020;146:1999–2049. <https://doi.org/10.1002/qj.3803>.

Newman, A. J., Clark, M. P., Craig, J., Nijssen, B., Wood, A., Gutmann, E., Mizukami, N., Brekke, L. y & Arnold, J. R. (2015). Gridded ensemble precipitation and temperature estimates for the contiguous United States. *Journal of Hydrometeorology*, 16(6), 2481-2500.

Newman, A. J., Clark, M. P., Longman, R. J., Gilleland, E., Giambelluca, T. W., & Arnold, J. R. (2019). Use of daily station observations to produce high-resolution gridded probabilistic precipitation and temperature time series for the Hawaiian Islands. *Journal of Hydrometeorology*, 20(3), 509-529.

Saha S, Nadiga S, Thiaw C, Wang J, Wang W, Zhang Q, van den Dool HM, Pan H-L, Moorthi S, Behringer D, Stokes D, Peña M, Lord S, White G, Ebisuzaki W, Peng P, Xie P. 2006. The NCEP Climate Forecast System. *J. Climate* 19: 3483– 3517.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., y Hsu, K.-L. (2018). A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics*, 56, 79–107. <https://doi.org/10.1002/2017RG000574>.

Tang, G., Clark, M. P., & Papalexiou, S. M. (2021). SC-earth: a station-based serially complete earth dataset from 1950 to 2019. *Journal of Climate*, 34(16), 6493-6511.

Tang, G., Clark, M. P., Papalexiou, S. M., Newman, A. J., Wood, A. W., Brunet, D., & Whitfield, P. H. (2021). EMDNA: An ensemble meteorological dataset for North America. *Earth System Science Data*, 13(7), 3337-3362.

Tang, Guoqiang, Martyn P. Clark, and Simon Michael Papalexiou. "EM-Earth: The ensemble meteorological dataset for planet Earth." *Bulletin of the American Meteorological Society* 103.4 (2022): E996-E1018.

Yamazaki D., D. Ikeshima, J. Sosa, P.D. Bates, G.H. Allen, T.M. Pavelsky. MERIT Hydro: A high-resolution global hydrography map based on latest topography datasets *Water Resources Research*, vol.55, pp.5053-5073, 2019, doi: 10.1029/2019WR024873