

SOCIEDAD CHILENA DE INGENIERÍA HIDRÁULICA
XXV CONGRESO CHILENO DE INGENIERÍA HIDRÁULICA

**¿CÓMO SELECCIONAR MODELOS DE CIRCULACIÓN GENERAL PARA
ESTUDIOS REGIONALES?
PROPUESTA METODOLÓGICA BASADA EN EL DESEMPEÑO HISTÓRICO.**

FELIPE GATEÑO M.¹
XIMENA VARGAS M.²
PABLO MENDOZA Z.^{2,3}

RESUMEN

En la actualidad, los resultados de varios modelos climáticos están públicamente disponibles. Sin embargo, diferencias importantes en sus desempeños hacen necesaria una evaluación apropiada para la posterior selección. Con el fin de abordar este problema, se propone una metodología para generar un ranking de modelos generales de circulación (GCMs) que permita una selección para posteriores estudios regionales de cambio climático basada en tres métricas de diagnóstico que buscan: (i) evaluar si el modelo reproduce correctamente la tendencia de la serie anual observada; (ii) cuantificar el desempeño del modelo en reproducir la varianza a nivel interanual; y (iii) cuantificar la habilidad del modelo en simular el ciclo intra-anual. Esta metodología se aplica a 14 GCMs (13 modelos más un modelo artificial que resulta de promediar los resultados de modelos individuales, llamado MM13), obteniéndose que son 3 los modelos mejor evaluados para el territorio nacional (MPI-ESM1-2-HR, CanESM5 y NESM3). Por una parte, se concluye que la selección del modelo depende de la región analizada, ya que el mejor modelo identificado para una región puede llegar a ser relevado a las últimas posiciones para otra región. Adicionalmente, el modelo MM13 no supera la octava posición en el *ranking* de modelos -independiente de la región- principalmente debido a que suele subestimar la variabilidad interanual de las precipitaciones. Finalmente, al evaluar la reproducción de las tendencias anuales de las temperaturas extremas, se identifica un error sistemático en las zonas costeras para todos los GCMs, donde la tendencia observada es contraria a la modelada.

¹ Estudiante de M.Sc. en Recursos Hídricos, Departamento de Ingeniería Civil, Universidad de Chile - email: felipe.gateno@ug.uchile.cl

² Académico, Departamento de Ingeniería Civil, Universidad de Chile

³ Investigador Asociado, Centro Avanzado de Tecnología para la Minería (AMTC)

INTRODUCCIÓN

Uno de los grandes desafíos de la comunidad científica es el desarrollo de modelos de circulación general (GCMs por sus siglas en inglés) que permitan representar adecuadamente procesos atmosféricos para su posterior análisis en estudios de impacto del cambio climático (World Meteorological Organization, 1992). Gracias al esfuerzo coordinado de varios grupos de investigación, se ha logrado un avance sustancial en esta área y, junto con ello, se han realizado evaluaciones sistemáticas de dichos modelos en los distintos Proyectos de Intercomparación de Modelos Acoplados (CMIP), las cuales han sido resumidas en los reportes generados por el Panel Intergubernamental sobre el Cambio Climático (Flato et al., 2013; Houghton et al., 1995; Houghton et al., 2001; Randall et al., 2007). A la fecha, cerca de 60 modelos están públicamente disponibles en el Programa Mundial de Investigación del Clima (WRCP, <https://esgf-node.llnl.gov/search/cmip6/>).

Los modelos climáticos son las herramientas más comúnmente utilizadas en la investigación de la respuesta del sistema Tierra ante cambios en las forzantes externas (Flato et al., 2013), pero el alto costo computacional asociado a su ejecución hace necesario que los resultados estén discretizados a resoluciones horizontalmente gruesas, que generalmente difieren con la escala de los procesos regionales (Towler et al., 2010; Vano et al., 2015). Para abordar este problema, generalmente se realiza un proceso de escalamiento en que los resultados del modelo son llevados a una resolución más fina, permitiendo la representación de las variables climáticas a la escala local. Este proceso es comúnmente realizado mediante técnicas estadísticas (i.e., escalamiento estadístico: Cannon, 2016, 2018; Maurer et al., 2010; Wilby & Wigley, 1997) que típicamente corrigen el sesgo entre los campos atmosféricos modelados y observados.

Este proceso, sin embargo, no suele compensar los errores estructurales de los modelos climáticos, por lo que, aún posterior al escalamiento, se pueden identificar errores en las señales climáticas (Gutmann et al., 2014; Vaittinada Ayar et al., 2016). Maraun (2016) revisó los métodos de escalamiento estadístico más comúnmente utilizados, concluyendo que estos no reparan las deficiencias estructurales de los GCMs, como por ejemplo una incorrecta representación de la orografía o deficiencias en la representación de procesos a escala de sub-grilla. Debido a esto, es necesario realizar un examen cuidadoso de los modelos, enfocado en su desempeño en la región de interés.

Hasta la fecha, no hay consenso en la comunidad científica sobre cómo seleccionar y diagnosticar los modelos previo al proceso de escalado. Sin embargo, existen principalmente dos enfoques: (i) seleccionar el GCM basado en su desempeño durante un periodo histórico (Nashwan & Shahid, 2020; Pierce et al., 2009), el cual consiste en seleccionar un set de modelos que entregue la mejor representación de las observaciones; y (ii) un enfoque basado en la envolvente de modelos (Lutz et al., 2016), donde se seleccionan aquéllos que cubren el rango completo de los posibles escenarios futuros. Ambas metodologías tienen ventajas y desventajas asociadas. Sin embargo, debido a que el segundo enfoque requiere filtrar modelos -plausibles- que proyecten cambios moderados en las señales climáticas, en este trabajo se ha optado por diagnosticar los GCMs basados en su desempeño en el periodo histórico. Bajo este contexto, son dos las preguntas científicas que se quieren abordar:

- ¿La selección de GCMs depende de la zona de análisis?
- ¿Es el *modelo promedio* (MM13, calculado a partir del promedio de los 13 GCMs disponibles) el mejor de los modelos si es que no se incluyen métricas de diagnóstico asociadas al error en los montos?

1. ZONA DE ESTUDIO Y DATOS

El dominio de estudio consiste en Chile Continental, el cual se divide en 6 macrozonas siguiendo el criterio de Alvarez-Garreton et al. (2018): (1) Norte Grande, desde la región de Arica y Parinacota hasta Antofagasta; (2) Norte Chico, desde Atacama hasta Coquimbo; (3) Zona Central, desde Valparaíso hasta el Maule; (4) Zona Sur, desde Bío-Bío hasta Los Lagos; (5) Zona Austral (región de Aysén); y (6) Patagonia Sur (región de Magallanes).

Como set de datos observacional, se utilizó el producto grillado CR2METV2.0 (disponible en <https://www.cr2.cl/datos-productos-grillados>), el cual contiene registros diarios de precipitación y temperaturas extremas a una resolución horizontal de 0,05°x0,05° para todo Chile Continental, desde enero del 1979 hasta diciembre del 2018. Este producto se construye a partir de un escalamiento estadístico que utiliza las variables del reanálisis atmosférico ERA5 e índices topográficos como predictores, y observaciones in-situ como predictandos. Debido a su resolución horizontal, ERA5 suele subestimar los montos de precipitación a lo largo del territorio chileno (DGA, 2017), lo que repercute en una subestimación de los montos estimados en CR2METV2.0 (DGA, 2019). Debido a esto, se utilizó un set de datos corregido siguiendo la metodología de Yang et al. (1998) (también usado en DGA, 2019).

Si bien hay alrededor de 40 modelos disponibles del Proyecto de Intercomparación de Modelos Acoplados, fase 6 (CMIP6 por sus siglas en inglés) (O'Neill et al., 2016), se seleccionaron aquellos que cuenten con simulaciones tanto de precipitación como de temperaturas extremas, obteniéndose un set de 13 GCMs (disponibles en <https://esgf-node.llnl.gov/search/cmip6/>). Adicionalmente, se incluyó un modelo generado a partir del promedio diario de los 13 modelos en cada píxel de la zona de estudio, de ahora en adelante denominado MM13. Si bien el periodo histórico de los modelos abarca desde 1850 hasta 2014, el periodo de análisis para el presente estudio comienza en enero de 1979 y finaliza en diciembre del 2014.

Debido a diferencias en la resolución espacial, y con el objetivo de evaluar el efecto de esta en el diagnóstico de los GCMs, los campos atmosféricos modelados (observados), en el período base, son interpolados (agregados) a 1, 0,75 y 0,5 grados.

2. MÉTRICAS DE DIAGNÓSTICO Y METODOLOGIA DE SELECCIÓN DE GCMs

Las métricas seleccionadas para realizar el diagnóstico consideran: (i) que los GCMs no son desarrollados para que los campos modelados y observados estén temporalmente correlacionados, si no para que provean una estimación de los patrones atmosféricos desde la escala estacional hacia escalas temporales mayores (Flato et al., 2013; Maraun, 2016; Randall et al., 2007); (ii) que, usualmente, los GCMs sobrestiman la cantidad de días con precipitación -también llamado como efecto llovizna- (Dettinger et al., 2004; Gutmann et al.,

2014; Maurer et al., 2010), y por lo tanto, la elección de los modelos basado en métricas de diagnóstico a escala diaria repercutiría en errores sistemáticos; y (iii) que posterior a la selección de los GCMs se aplicará un proceso de escalado que corregirá su sesgo, por lo que no se incluirán métricas asociadas al error en los montos entre el modelo y las observaciones.

Primero, a partir de las series anuales observadas y modeladas en cada píxel, se generan series anuales suavizadas utilizando medias móviles de 10, 15 y 20 años para obtener el estadístico τ producto de aplicar la prueba de Mann-Kendall (Mann, 1945). Dicho estadístico puede tomar valores entre -1 y 1, donde valores positivos (negativos) cercanos a 1 (-1) indican que existe una tendencia monótonamente creciente (decreciente), mientras que valores de τ cercanos a 0 reflejan que no existe una tendencia en la serie. Al aplicar la prueba a la serie observada y modelada se obtienen entonces τ_{obs} y τ_{GCM} , respectivamente, con lo que se construye la primera métrica de diagnóstico:

$$\Delta\tau = sgn(\tau_{obs})sgn(\tau_{GCM})|\tau_{GCM} - \tau_{obs}| \quad (1)$$

Donde $sgn(\tau_{obs})$ y $sgn(\tau_{GCM})$ corresponden al signo del estadístico τ_{obs} y τ_{GCM} , respectivamente. El objetivo de esta métrica es evaluar la habilidad de los modelos en representar correctamente la tendencia histórica observada de cada variable: valores positivos (negativos) de $\Delta\tau$ indican que la tendencia del modelo se condice (contradice) con la observada, y mientras mayor sea la magnitud de $\Delta\tau$ mayor es la diferencia en la magnitud de las tendencias (i.e. el valor deseado es $\Delta\tau$, priorizando valores positivos sobre negativos).

Para evaluar la correcta reproducción de la variabilidad interanual, se calcula el coeficiente de variación de las series anuales observadas y modeladas en cada píxel. Luego, se calcula la tercera métrica de diagnóstico como $\Delta CV = CV_{GCM} - CV_{obs}$, donde CV_{obs} y CV_{GCM} son los coeficientes de variación para la serie observada y modelada, respectivamente. Valores de ΔCV positivos (negativos) indicarían una sobrestimación (subestimación) de la variabilidad interanual de la variable, mientras que valores cercanos a 0 indicarían una correcta representación de este aspecto.

Finalmente, para evaluar el desempeño del modelo en reproducir el ciclo intra-anual, se calcula el promedio mensual de cada variable por píxel en todo el periodo de estudio, tanto para el set de datos observacional como para los GCMs. Luego, se calcula el coeficiente de correlación de Pearson entre 12 valores simulados y 12 observados, obteniéndose la segunda métrica de diagnóstico, donde valores cercanos a 1 (-1) indicarían una correcta (incorrecta) reproducción del ciclo intra-anual.

Las tres métricas anteriores son calculadas para tres resoluciones horizontales ($0,5^\circ$, $0,75^\circ$ y 1°), generándose un diagnóstico independiente para cada resolución y macrozona en base a los siguientes pasos (Figura 1): (a) las tres métricas son calculadas para cada variable y píxel; (b) se construye la función de probabilidad acumulada (CDF) para cada métrica, y se calcula el área entre esta y su respectivo valor deseado. (c) Por construcción, la menor área en cada métrica estará asociada con el modelo que tenga el mejor desempeño a nivel de macrozona y, por lo tanto, dicho modelo obtendrá el mejor lugar en el ranking; (d) Se repite el paso anterior para el resto de las métricas, y luego se promedian los rankings respectivos asignando el mismo peso para cada par variable-métrica; y (e) el GCM con el menor promedio es seleccionado como el mejor modelo para la macrozona respectiva.

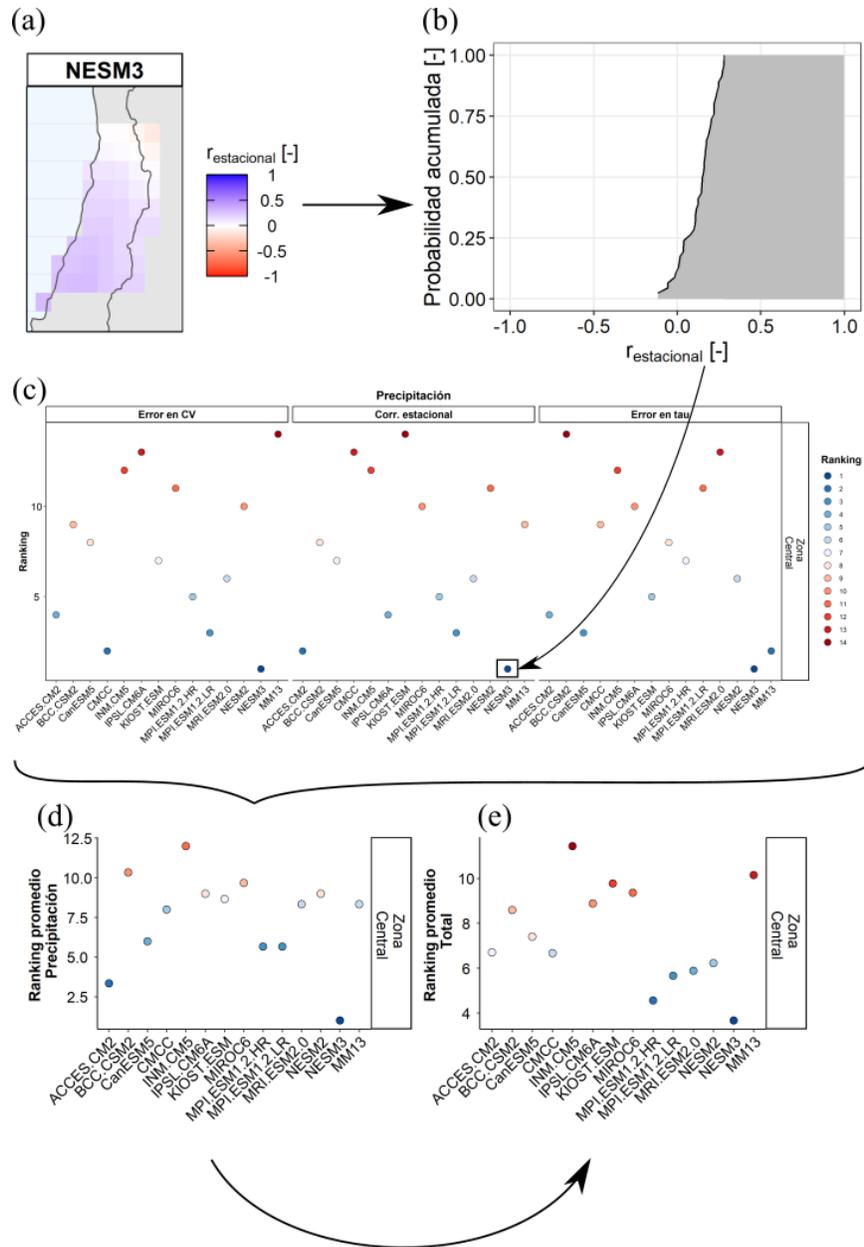


Figura 1: Esquema de la metodología de ranking. (a) Se calcula una métrica para un par GCM-variable en cada píxel asociado a una macrozona. En el panel (a), se presentan los resultados de la correlación estacional para la precipitación modelada por NESM3 en la macrozona Zona Central. (b) Se construye la CDF y se calcula el área entre esta y el valor deseado de la variable. En el panel (b) la línea negra representa la CDF de la correlación estacional, y el área gris representa el área entre esta y 1 (el valor deseado para dicha métrica). (c) El GCM recibe el lugar en el ranking según el área calculada en (2), y se repite el proceso para el resto de las métricas. En este caso el modelo NESM3 recibe el primer lugar para la correlación estacional de la precipitación (rectángulo negro en panel (c)) debido a que es el modelo con la menor área en dicho par métrica-variable. (d) Se calcula el ranking promedio del par GCM-variable, promediando el ranking de las cuatro métricas (panel (d)) y se repite

el proceso para las demás variables. (e) Finalmente, se calcula el ranking promedio total al promediar el ranking obtenido para las 3 variables (panel (e)).

3. RESULTADOS

3.1 EFECTO DE LA RESOLUCIÓN EN LA SELECCIÓN DEL MEJOR GCM

Los resultados indican que no hay diferencias considerables asociadas a la resolución horizontal utilizada para el diagnóstico, ya que los *rankings* obtenidos por los modelos en las tres resoluciones (Figura 2a) no varían considerablemente: en 3 de las 6 macrozonas el mejor modelo es el mismo independiente de la resolución. En las demás macrozonas, pequeñas diferencias en el ranking promedio total entre los primeros GCMs causan diferencias en la selección del mejor modelo dependiendo de la resolución. Debido a esto, los resultados presentados a continuación se restringen a la resolución horizontal de 0,50°. Adicionalmente, tampoco se identificaron diferencias considerables producto de utilizar ventanas móviles de distinto largo al calcular la métrica $\Delta\tau$, por lo que la métrica utilizada para calcular diagnósticos está construida a partir de ventanas móviles de 15 años.

En la Figura 2b se muestra el ranking de los modelos por macrozona para dicha resolución: en el Norte Grande, Zona Central y Patagonia Sur el mejor modelo de acuerdo con la metodología propuesta es NESM3; en Norte Chico y Zona Sur el mejor modelo es MPI-ESM1-2-HR; y el mejor modelo para la Zona Austral es CanESM5.

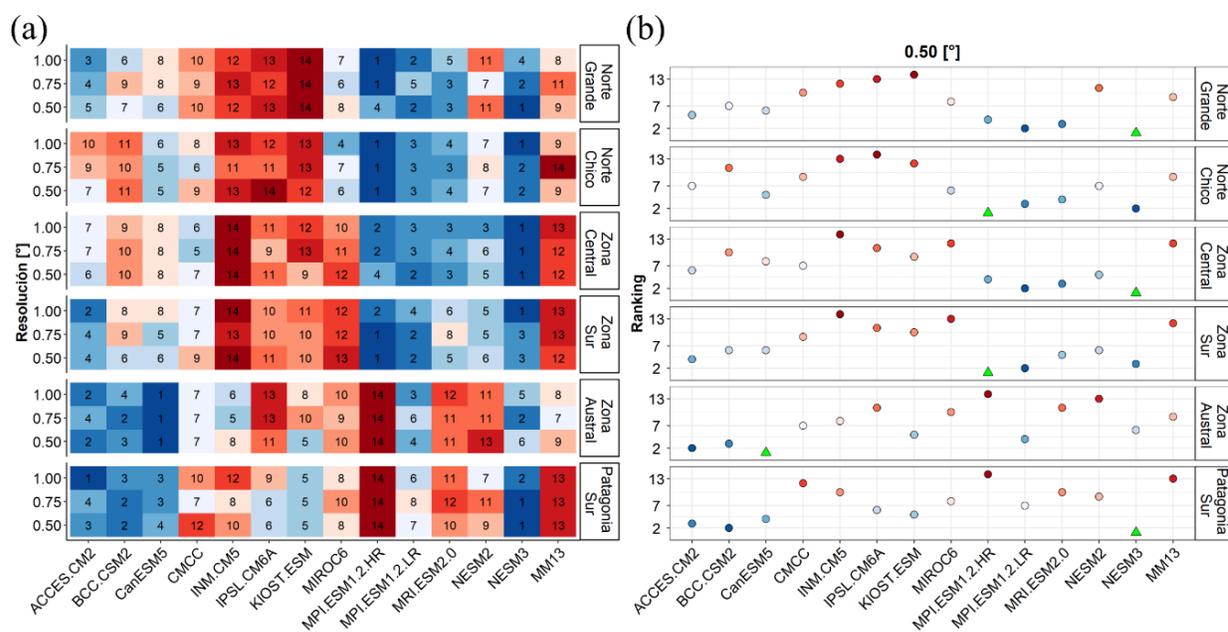


Figura 2: (a) Ranking de los GCMs por macrozona para las tres resoluciones de diagnóstico (colores azules y rojos representan los mejores y peores modelos, respectivamente). (b) Ranking de los GCMs a la resolución de 0,50° (mejores modelos representados con triángulos verdes).

3.2 COMPARACIÓN ENTRE MM13 Y LOS GCMs SELECCIONADOS

La Figura 3 muestra la distribución espacial de las métricas asociadas a la precipitación para el modelo promedio (MM13) y los tres mejores modelos obtenidos en la sección 3.1 (MPI-ESM1-2-HR, NESM3 y CanESM5). Los modelos MPI-ESM1-2-HR, NESM3 y MM13 presentan un mejor desempeño en la reproducción de las tendencias anuales (Figura 3a) entre las macrozonas del Norte Grande y Zona Sur en comparación con las dos macrozonas restantes, donde se obtienen una mayoría de píxeles con valores de $\Delta\tau$ negativos, indicando que los GCMs modelan una tendencia contraria a la observada en las regiones ubicadas al sur del territorio nacional. Por otro lado, si bien el modelo CanESM5 presenta tendencias contrarias a las observadas en la zona cordillerana de las macrozonas Norte Grande y Norte Chico, es el modelo que reporta mayor cantidad de píxeles con valores de $\Delta\tau$ cercanos a 0 en la macrozona de la Zona Austral, lo que justifica su elección como mejor modelo en dicha región.

En Figura 3b se observa un comportamiento similar para los 4 modelos en la reproducción del ciclo intra-anual: los mejores desempeños se obtienen para las macrozonas Zona Central y Zona Sur, mientras que en las demás regiones se obtienen, en general, coeficientes de correlación de Pearson negativos para la serie de valores medios mensuales. Estos resultados son similares para los demás 10 modelos estudiados (no presentado).

Finalmente, en la Figura 3c se presentan los resultados obtenidos para la métrica asociada al error en la variabilidad interanual. Si bien los cuatro modelos reportan, en general, una subestimación de la variabilidad (valores de ΔCV menores a 0), el modelo MM13 subestima la variabilidad en todo el territorio nacional, mientras que el área asociada a la subestimación en los modelos MPI-ESM1-2-HR, CanESM5 y NESM3 corresponde a ~71%, 74% y ~55%, de Chile continental, respectivamente. Para las temperaturas extremas se obtuvieron resultados similares (no presentado), siendo la variabilidad subestimada en ~81% del área para el modelo MM13, un 25%, 11% y 5% más que lo obtenido para los modelos MPI-ESM1-2-HR, CanESM5 y NESM3, respectivamente.

En la Figura 4 se presentan los resultados de $\Delta\tau$, ahora aplicados a las series anuales de temperaturas máximas. Si bien se identifican diferencias en el desempeño de los modelos para las macrozonas Zona Austral y Patagonia Sur, todos presentan tendencias contrarias a la observada en la zona costera de las regiones ubicadas entre las macrozonas Norte Grande y Zona Sur. Este resultado se obtiene para las tres resoluciones diagnosticadas, tanto para las series de temperaturas máximas como para las de temperaturas mínimas (no presentado). Este fenómeno podría ser explicado por una deficiente representación de la interfaz océano-continente en los GCMs, ya que la gruesa discretización horizontal de estos asociaría celdas oceánicas a píxeles que en realidad corresponden al continente cuando se analiza a una resolución más fina.

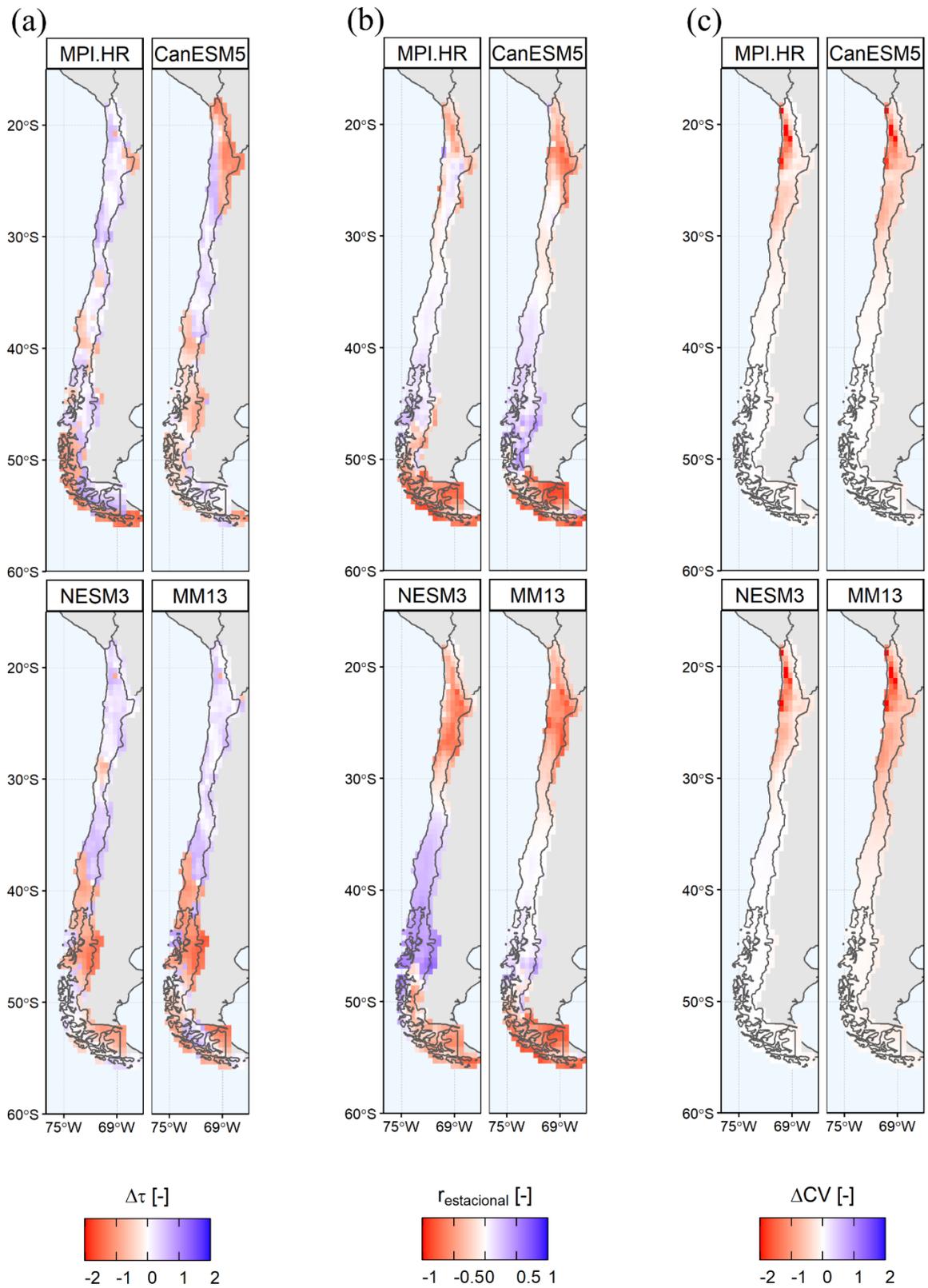


Figura 3: Diagnóstico de la precipitación para los modelos MPI-ESM1-2-HR (MPI.HR), CanESM5, NESM3 y MM13: (a) error en la tendencia anual ($\Delta\tau$); (b) reproducción del ciclo intra-anual ($r_{estacional}$); y (c) error en la variabilidad interanual (ΔCV).

Diagnóstico de temp. máximas a 0.5°:
error en las tendencias observadas

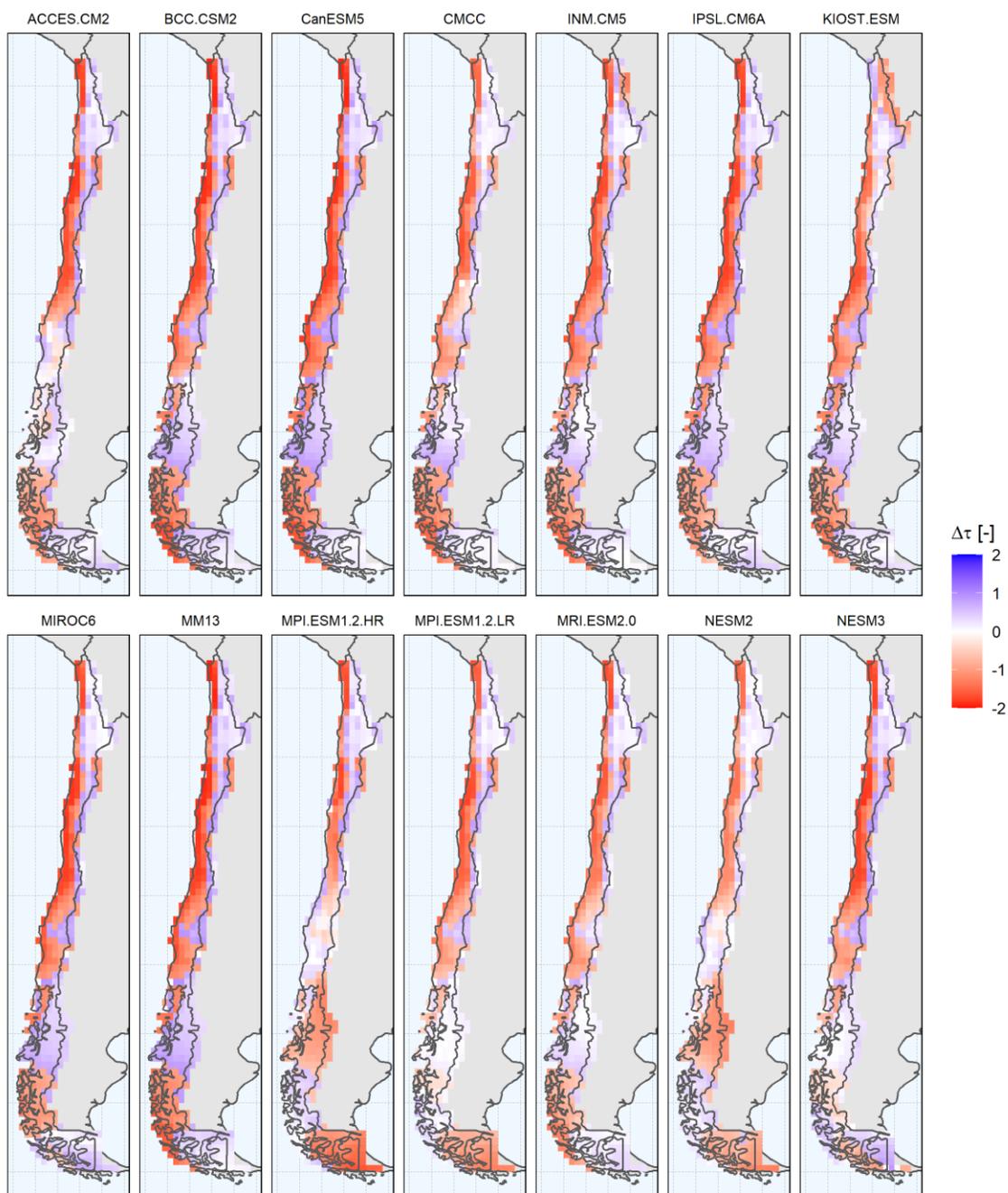


Figura 4: Diagnóstico de la reproducción de la tendencia anual de las temperaturas máximas.

En la Figura 5 se presenta la distribución de valores de ΔCV para las series de precipitación de los 14 GCMs. Se puede identificar que el modelo MM13 exhibe una distribución centrada en $\Delta CV = -0,1$ aproximadamente, mientras que los 13 modelos restantes presentan una distribución centrada en $\Delta CV = 0$. Estos resultados indican que promediar todos los modelos disponibles induce a una subestimación de la variabilidad interanual de la precipitación, una

desventaja que debe ser considerada en los estudios regionales que utilicen un GCM generado a partir del promedio, y que puede ser identificada al aplicar la metodología de selección de modelos propuesta. En efecto, los métodos de escalado estadístico más comúnmente usados no corrigen explícitamente variaciones interanuales, por lo que es esencial seleccionar un modelo que entregue un buen desempeño en esta métrica previo al proceso de escalado (Gutmann et al., 2014).

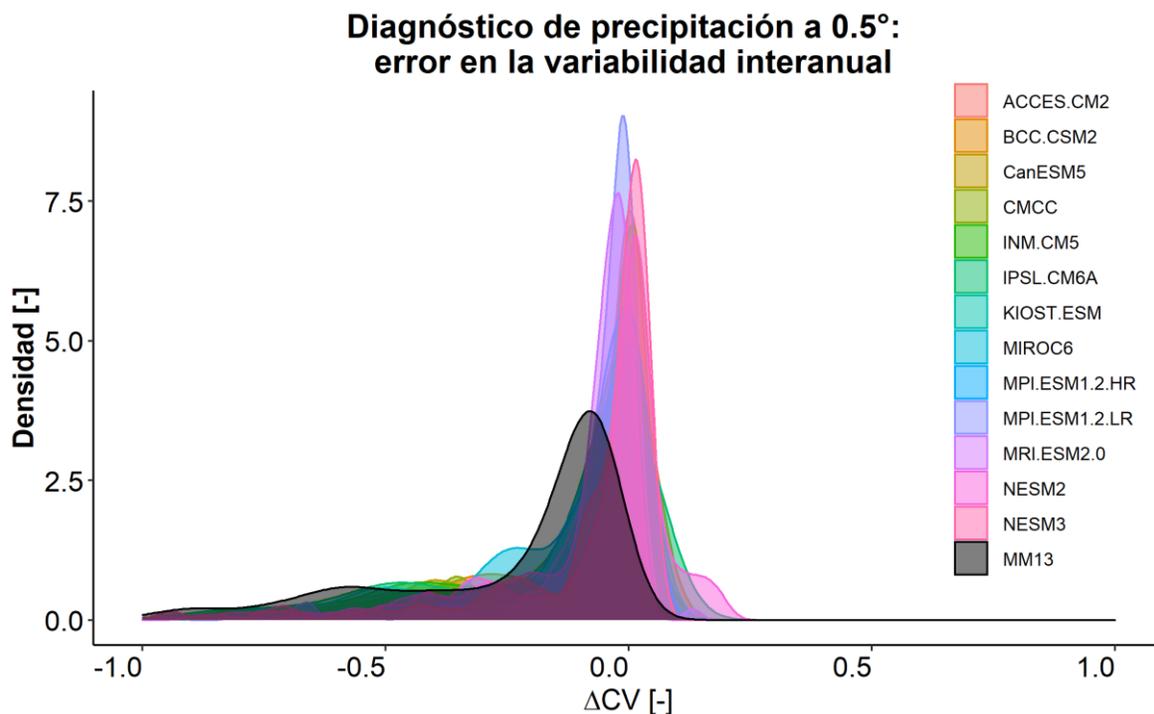


Figura 5: Gráfico de densidad del error en la variabilidad interanual de las precipitaciones.

4. CONCLUSIONES

Se propone una metodología para generar un ranking de GCMs para estudios regionales de cambio climático. Esta se basa en el cálculo de tres métricas que buscan evaluar los aspectos estructurales de los modelos generales de circulación: (i) $\Delta\tau$, construido para evaluar si el modelo reproduce correctamente la tendencia de la serie anual observada; (ii) el error en el coeficiente de variación (ΔCV), que cuantifica el desempeño del modelo en reproducir la varianza a nivel interanual; (iii) la correlación estacional entre el modelo y las observaciones ($r_{estacional}$), que cuantifica la habilidad del modelo en simular el ciclo intra-anual. Dichas métricas se calculan a partir de las series de precipitación y temperaturas extremas a escala de píxel y se realiza un proceso de *ranking* para cada una de las 6 macrozonas definidas para Chile Continental.

Al aplicar la metodología propuesta a 14 modelos se obtiene que son 3 los GCMs con mejor desempeño, el que a la vez depende de la macrozona que se estudie: (1) el modelo MPI-ESM1-2-HR posee el mejor desempeño entre las regiones de Arica y Parinacota y Coquimbo (macrozonas Norte Grande y Norte Chico) y entre las regiones de Bío-Bío y Los Lagos (Zona Sur); (2) entre las regiones de Valparaíso y del Maule (Zona Central) y en la región de

Magallanes (Patagonia Sur) el GCM mejor posicionado es NESM3; y (3) CanESM5 es el mejor evaluado en la región de Aysén (Zona Austral). En esa misma línea, el modelo MM13 (modelo generado a partir del promedio de los GCMs disponibles) es relevado a la octava e incluso a la décimo tercera posición del ranking dependiendo de la macrozona analizada.

Se identificó que la principal deficiencia que influye en el *ranking* de MM13 es que exhibe una clara subestimación en la variabilidad interanual de las precipitaciones, revelando la desventaja de usar el modelo “promedio” en estudios regionales. Para la precipitación, MM13 subestima la variabilidad en todo el territorio nacional, un 25% o incluso un 45% más que lo obtenido para los modelos seleccionados según la metodología propuesta.

Adicionalmente, se identificó que todos los modelos simulan tendencias de temperaturas extremas contrarias a las observadas en la zona costera de las macrozonas Norte Grande y Norte Chico, posiblemente asociado a una deficiencia en la representación de la interfaz océano-continente debido a la gruesa resolución horizontal de los GCMs.

La inclusión de las tres métricas propuestas en el diagnóstico de GCM permite seleccionar modelos que posean un correcto desempeño en los aspectos estructurales. Estos aspectos, a diferencia del sesgo en los montos, no son generalmente corregidos por los métodos de escalado, por lo que su consideración es crítica para elegir un modelo climático.

REFERENCIAS

- Alvarez-Garreton, C., Mendoza, P. A., Pablo Boisier, J., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Puelma, C., Cortes, G., Garreaud, R., McPhee, J., & Ayala, A. (2018). The CAMELS-CL dataset: Catchment attributes and meteorology for large sample studies-Chile dataset. *Hydrology and Earth System Sciences*, 22(11), 5817–5846. <https://doi.org/10.5194/hess-22-5817-2018>
- Cannon, A. J. (2016). Multivariate bias correction of climate model output: Matching marginal distributions and intervariable dependence structure. *Journal of Climate*, 29(19), 7045–7064. <https://doi.org/10.1175/JCLI-D-15-0679.1>
- Cannon, A. J. (2018). Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dynamics*, 50(1–2), 31–49. <https://doi.org/10.1007/s00382-017-3580-6>
- Dettinger, M. D., Cayan, D. R., Meyer, M. K., & Jeton, A. E. (2004). Simulated Hydrologic Responses to Climate Variations and Change in the Merced, Carson, and American River Basins, Sierra Nevada, California, 1900–2099. *Climatic Change*, 62(1–3), 283–317. <https://doi.org/10.1023/B:CLIM.0000013683.13346.4f>
- DGA. (2017). *Actualización del Balance Hídrico Nacional*.
- DGA. (2019). Aplicación de la metodología de actualización del balance hídrico nacional en la macrozona sur y parte norte de la macrozona austral. SIT N° 441. *Ministerio de Obras Públicas, Dirección General de Aguas, División de Estudios y Planificación. Elaborado Por: Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas. Santiago, Chile*. <https://snia.mop.gob.cl/sad/REH5889v1.pdf%0Ahttps://snia.mop.gob.cl/sad/REH5889v2.pdf%0Ahttps://snia.mop.gob.cl/sad/REH5889v3.pdf%0Ahttps://snia.mop.gob.cl/sad/REH5889v4.pdf>
- Flato, G., Marotzke, J., Abiodun, B., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., & Rummukainen, M. (2013). Evaluation of climate models. *Climate Change 2013 the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, 9781107057, 741–866. <https://doi.org/10.1017/CBO9781107415324.020>
- Gutmann, E., Pruitt, T., Clark, M. P., Brekke, L., Arnold, J. R., Raff, D. A., & Rasmussen, R. M. (2014). An intercomparison of statistical downscaling methods used for water resource assessments in the United States. *Water Resources Research*, 50(9), 7167–7186. <https://doi.org/10.1002/2014WR015559>

- Houghton, J. T., Meira Filho, L. G., Callander, B. A., Harris, N., Kattenber, A., & Maskell, K. (1995). Climate Change 1995: Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change. *Geoscience Canada*. https://doi.org/10.1007/978-3-662-03925-0_1
- Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Linden, P. j., Dai, X., Maskell, K., & Jhonson, C. A. (2001). CLIMATE CHANGE 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. *Journal of Shoulder and Elbow Surgery*, 881. [https://doi.org/10.1016/S1058-2746\(02\)86826-4](https://doi.org/10.1016/S1058-2746(02)86826-4)
- Lutz, A. F., ter Maat, H. W., Biemans, H., Shrestha, A. B., Wester, P., & Immerzeel, W. W. (2016). Selecting representative climate models for climate change impact studies: an advanced envelope-based selection approach. *International Journal of Climatology*, 36(12), 3988–4005. <https://doi.org/10.1002/joc.4608>
- Mann, H. B. (1945). Nonparametric Tests Against Trend. *Econometrica*, 13(3), 245. <https://doi.org/10.2307/1907187>
- Maraun, D. (2016). Bias Correcting Climate Change Simulations - a Critical Review. *Current Climate Change Reports*, 2(4), 211–220. <https://doi.org/10.1007/s40641-016-0050-x>
- Maurer, E. P., Hidalgo, H. G., Das, T., Dettinger, M. D., & Cayan, D. R. (2010). The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in California. *Hydrology and Earth System Sciences*, 14(6), 1125–1138. <https://doi.org/10.5194/hess-14-1125-2010>
- Nashwan, M. S., & Shahid, S. (2020). A novel framework for selecting general circulation models based on the spatial patterns of climate. *International Journal of Climatology*, 40(10), 4422–4443. <https://doi.org/10.1002/joc.6465>
- O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J. F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., & Sanderson, B. M. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, 9(9), 3461–3482. <https://doi.org/10.5194/gmd-9-3461-2016>
- Pierce, D. W., Barnett, T. P., Santer, B. D., & Gleckler, P. J. (2009). Selecting global climate models for regional climate change studies. *Proceedings of the National Academy of Sciences of the United States of America*, 106(21), 8441–8446. <https://doi.org/10.1073/pnas.0900094106>
- Randall, D. A., Wood, R. A., Bony, S., Colman, R., Fichetef, T., Fyfe, J., Katssov, V., Pitman, A., Shuckla, J., Srinivasan, J., Stouffer, R. J., Sumi, A., & Taylor, K. E. (2007). Climate Models and Their Application. *Climate Change 2007 the Physical Science Basis: Working Group I Contribution to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, 23–30. <https://doi.org/10.2110/scn.94.03.0023>
- Towler, E., Rajagopalan, B., Gilleland, E., Summers, R. S., Yates, D., & Katz, R. W. (2010). Modeling hydrologic and water quality extremes in a changing climate: A statistical approach based on extreme value theory. *Water Resources Research*, 46(11), 1–11. <https://doi.org/10.1029/2009WR008876>
- Vaittinada Ayar, P., Vrac, M., Bastin, S., Carreau, J., Déqué, M., & Gallardo, C. (2016). Intercomparison of statistical and dynamical downscaling models under the EURO- and MED-CORDEX initiative framework: present climate evaluations. *Climate Dynamics*, 46(3–4), 1301–1329. <https://doi.org/10.1007/s00382-015-2647-5>
- Vano, J. A., Nijssen, B., & Lettenmaier, D. P. (2015). Seasonal hydrologic responses to climate change in the Pacific Northwest. *Water Resources Research*, 51(4), 1959–1976. <https://doi.org/10.1002/2014WR015909>
- Wilby, R. L., & Wigley, T. M. L. (1997). Downscaling general circulation model output: A review of methods and limitations. *Progress in Physical Geography*, 21(4), 530–548. <https://doi.org/10.1177/030913339702100403>
- World Meteorological Organization. (1992). CLIMATE CHANGE: The IPCC 1990 and 1992 Assessments. In *Nature*. <https://doi.org/10.1038/360292e0>
- Yang, D., Goodison, B. E., Metcalfe, J. R., Golubev, V. S., Bates, R., Pangburn, T., & Hanson, C. L. (1998). Accuracy of NWS 8" standard nonrecording precipitation gauge: Results and application of WMO intercomparison. *Journal of Atmospheric and Oceanic Technology*, 15(1), 54–68. [https://doi.org/10.1175/1520-0426\(1998\)015<0054:AONSNP>2.0.CO;2](https://doi.org/10.1175/1520-0426(1998)015<0054:AONSNP>2.0.CO;2)