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

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

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN RECURSOS Y MEDIO AMBIENTE HÍDRICO

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL

FELIPE IGNACIO GATEÑO MENESES

PROFESORA GUÍA:

XIMENA VARGAS MESA

PROFESOR CO-GUÍA:

PABLO MENDOZA ZÚÑIGA

MIEMBROS DE LA COMISIÓN:

MIGUEL LAGOS ZÚÑIGA

NICOLÁS VÁSQUEZ PLACENCIA

SANTIAGO DE CHILE

2022

RESUMEN DE LA TESIS PARA OPTAR AL TÍTULO DE:
Grado de Magíster en ciencias de la ingeniería, mención recursos y medio ambiente hídrico
MEMORIA PARA OPTAR AL TÍTULO DE: Ingeniería Civil
POR: Felipe Ignacio Gateño Meneses
FECHA: 2022
PROFESOR GUÍA: Ximena Vargas Mesa

PROPUESTA METODOLÓGICA PARA LA SELECCIÓN DE MODELOS DE CIRCULACIÓN GENERAL PARA ESTUDIOS REGIONALES.

Estudios realizados durante las últimas décadas indican que el cambio climático afectará la disponibilidad de los recursos hídricos debido a cambios en temperaturas y precipitaciones. Para mitigar dichos efectos, se suelen realizar estudios hidrológicos que, a través de proyecciones climáticas generadas por modelos de circulación general (GCMs), permiten estimar la disponibilidad hídrica futura, cuantificar cambios proyectados, y generar planes que faciliten un manejo más eficiente del recurso hídrico. En la actual fase del Proyecto de Intercomparación de Modelos Acoplados (CMIP6) existen alrededor de 60 GCMs, los cuales simulan a escala diaria la climatología histórica (1850-2014) y futura (2015-2100) para todo el mundo bajo distintos escenarios. Utilizar la totalidad de ellos resulta en una tarea ardua debido a los altos costos computacionales de procesar tal magnitud de información, y, por lo tanto, es imprescindible someterlos a un proceso de diagnóstico que permita la selección de modelos que representen adecuadamente la climatología de la zona de estudio. Debido a esto, en el presente trabajo se propone una metodología de *ranking* de GCMs que facilite la selección para posteriores estudios regionales de cambio climático en Chile continental.

El estudio se lleva a cabo considerando 27 modelos (y el promedio de estos) asociados al CMIP6, los cuales son diagnosticados en el periodo 1979-2014, evaluando cuatro aspectos de las series de precipitaciones y temperaturas extremas: (i) variabilidad interanual; (ii) correlación espacial de la media climatológica; (iii) reproducción temporal del ciclo medio anual; y (iv) la tendencia anual. Se construyeron cuatro métricas, las cuales fueron calculadas para cada GCM y comparadas con el set de datos observacional CR2METv2.0, generándose un *ranking* de modelos a nivel general y para cada métrica y variable. Se analizó también el efecto de la resolución horizontal a la que se diagnostican los modelos, repitiendo el proceso de diagnóstico a cuatro resoluciones horizontales. Finalmente, se estudió la influencia que tiene el escalamiento estadístico en el rendimiento de los modelos al aplicarse, a la resolución de 0,25°, dos métodos a los GCMs para luego recalcular el *ranking* de modelos y compararlo con el obtenido previo al proceso de escalamiento. Los resultados indican que no existe un modelo que demuestre superioridad en todas las variables y métricas de manera simultánea, y a la vez, todos varían en su desempeño según la zona que se esté analizando. Los dos métodos de escalamiento estadístico aplicados producen una mejoría del ciclo anual climatológico y la correlación espacial, aunque no logran corregir errores en las tendencias históricas, donde todos los GCMs presentan tendencias contrarias a las observadas en precipitaciones y temperaturas extremas en ciertas zonas del país. Considerando la variabilidad espacial del desempeño de los GCMs, se recomienda realizar, a escala regional, una selección basada en el diagnóstico de múltiples aspectos y variables de la meteorología, donde se considere, al menos, la reproducción del ciclo anual, la variabilidad interanual y las tendencias observadas en el periodo histórico.

AGRADECIMIENTOS

Antes que todo me gustaría darle las gracias a mis papás, Alejandra y Alexander, por ser un apoyo y unos guías en todo el transcurso de mi vida. Desde pequeño estuvieron siempre para mí, preocupados de que no me faltara nada y preocupados de inculcarme los valores de la responsabilidad, el respeto y la empatía. Este trabajo está dedicado a ustedes, y espero que sea un primer paso para empezar a retribuirles todo lo que han hecho por mí. Agradecer a mis abuelos, por cuidarme y regalonearme durante esos veranos que pasaba en Peñaflor, a mis primos, Ignacia, Berni y Juanma, por alegrarme con sus bromas y risas de zarigüeyas, a Carola, por su constante preocupación y apoyo, y a Marcela, por acarrearme todas esas mañanas camino a Santiago. A pesar de que por distintas razones no seguimos siendo tan unidos como antes, quiero agradecer a Silvia y Hugo, mis tíos Pablo, Felipe, Paty y Claudia y a mis primos Javiera, Seba, Feña y Fran, por esas juntas familiares que hacíamos hace años.

Gracias a esas amistades que he mantenido desde el colegio: mis amigos de toda la vida Pablo e Ignacio, por todas esas historias que hemos creado juntos y por ser siempre un apoyo en todos mis logros y fracasos, y la Ale por esa década de amistad y confianza que, a pesar de la distancia que ha habido en ciertos momentos, hemos podido mantener como el primer día.

Este trabajo cierra también un ciclo universitario que ha sido profundamente alegre gracias a las hermosas personas que pude conocer. Barby y Luciano, gracias a ambos por estar siempre ahí para a escuchar, por apoyarme y por alegrarse de mis logros. Gracias Enzo, Paty, Max, Miguel y Conchi, por poder mantener nuestra amistad desde el primer año de universidad, por todas las risas, salidas y carretes que nos hemos mandado. Gracias al Pao, Benja, Mati, Alan, Tapia y Ace por los momentos que vivimos juntos, y que ha pesar de que nos distanciamos en los últimos años, siempre los tendré en la memoria.

Agradezco especialmente haber conocido a Javiera, mi polola, confidente y amiga, por ayudarme en los momentos de frustración, guiarme en los momentos de dudas, y por darme tantos momentos de alegría. Estoy completamente agradecido de que el destino haya cruzado nuestros caminos hace años, y que ahora me permita ser tu compañero y tú la mía. Contigo la rutina del día a día se vuelve mágica.

Por último, agradecer a aquellas personas que me han guiado en mi camino profesional. Gracias a mis profesores Ximena, Pablo y Miguel, porque gracias a la vocación que tienen como docentes y académicos he podido orientar mi carrera a lo que verdaderamente me gusta. Gracias a todo el equipo de Eridanus, Lalo, Mauricio, Rodrigo, Rossana, Ignacio y Javiera, por el buen ambiente con el que me han recibido, desde el día uno en mi práctica profesional hasta ahora, ya comenzando oficialmente mi carrera laboral.

Gracias infinitas, todos me han dejado una enseñanza que llevo conmigo en el día a día.

TABLA DE CONTENIDO

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

I. Introducción

En un contexto de cambio climático, donde la dinámica y disponibilidad de los recursos hídricos está (y seguirá) siendo afectada, se vuelve necesario tomar decisiones estratégicas que permitan prever y disminuir los impactos en la sociedad y en los ecosistemas en general (IPCC, 2021). Para ello, es esencial realizar estudios robustos de disponibilidad hídrica que permitan a quienes diseñan políticas públicas tomar decisiones informadas y bien documentadas.

La realización de dichos estudios requiere de decisiones metodológicas asociadas a la selección de: (i) el escenario climático (Lehner et al., 2020; Liang et al., 2020; Shen et al., 2018); (ii) los modelos de circulación general (Karmalkar et al., 2019; Lutz et al., 2016; Rupp et al., 2013); (iii) el método de escalamiento (Cannon, 2018; Gutmann et al., 2014; Vaittinada Ayar et al., 2016); (iv) el modelo hidrológico (Clark et al., 2011; Knoben et al., 2020); y (v) su respectiva estrategia de calibración (Liu et al., 2017; Mendoza et al., 2015). Todos estos elementos son parte de la cadena de modelación, donde las decisiones metodológicas llevadas a cabo en cada etapa tienen asociada una incertidumbre que debe ser cuantificada y reducida, para así generar resultados robustos que faciliten la toma de decisiones (Chegwidden et al., 2019; Clark et al., 2016; Wang et al., 2020).

Mientras varios estudios se han enfocado en caracterizar la incertidumbre de un elemento específico de la cadena de modelación (Blázquez & Nuñez, 2013; Gutmann et al., 2014; Hakala et al., 2018; Lutz et al., 2016), otros han estudiado el efecto conjunto en la dispersión de los resultados de la modelación hidrológica, identificando decisiones metodológicas más importantes que otras dependiendo de la aplicación (Shen et al., 2018). En Chegwidden et al. (2019), por ejemplo, identificaron que tanto la selección del escenario climático como la del GCM son las decisiones que más influyen a la incertidumbre asociada a la estimación de volúmenes de escorrentía. Similares resultados se encuentran en Wang et al. (2020), donde la selección del GCM se identificó como la principal fuente de incertidumbre para la proyección de volúmenes y eventos extremos de escorrentía. A pesar de la importancia de dicha decisión metodológica, aún no hay un consenso en la comunidad científica sobre cómo diagnosticar y/o seleccionar modelos climáticos.

Gracias al esfuerzo conjunto de varios centros de investigación, actualmente están disponibles los resultados de cerca de 60 GCMs en el sitio web del Programa Mundial de Investigación Climática (WRCP, www.wcrp-climate.org). Si bien pueden identificarse similitudes entre modelos debido a que comparten parametrizaciones y estructuras para representar los procesos climáticos (Brunner et al., 2020), se debe tener en cuenta que hay modelos que representan mejor ciertos procesos climáticos en comparación con otros (Clark et al., 2016; Knutti, 2010). Adicionalmente, trabajar con los resultados de todos los modelos disponibles puede convertirse en una tarea poco factible debido al costo y memoria computacional requerido para procesar las salidas de dichos GCMs, dados los tiempos de ejecución limitados en estudios. Estas razones, junto con la necesidad de incluir modelos que representen adecuadamente la climatología de la zona de estudio, justifican el diseño de un proceso robusto de selección de GCMs, el cual debiera ser incluido en todo estudio de impacto del cambio climático a modo de facilitar el análisis de la incertidumbre asociada a las proyecciones.

Si bien son varios los estudios que se han enfocado en esta etapa de la cadena de modelación (Salazar Morey, 2017; Sota Christie et al., 2022), no existe un consenso sobre cómo realizar la selección de GCMs, identificándose principalmente dos filosofías para ello; (i) seleccionar GCMs basado en su desempeño en un periodo histórico (e.g., Brunner et al., 2020; Rupp et al., 2013), filosofía que consiste en seleccionar un set de modelos que entregue la mejor representación de las observaciones basado en el diagnóstico de diversas métricas; y (ii) un enfoque basado en captar la envolvente de modelos (e.g., Lutz et al., 2016; Sa'adi et al., 2020), donde se seleccionan aquellos GCMs que cubren el rango completo de los posibles escenarios futuros. Ambas filosofías tienen ventajas y desventajas asociadas. Por ejemplo, seleccionar GCMs que cubran el rango de escenarios futuros implica filtrar modelos con señales de cambio moderadas, resultando en la eliminación de GCMs que representan una opción plausible. Por otro lado, debido a que existe una relación no estacionaria entre el modelo climático y las observaciones (i.e., los sesgos que el modelo tiene con respecto a las observaciones varían en el tiempo, e.g., Chen et al., 2015; Gutmann et al., 2014; Maraun, 2016), la elección de un GCM con un correcto desempeño en un periodo histórico no asegura que lo mismo sucederá en un periodo futuro (Wang et al., 2019). Sin embargo, debido a que el objetivo principal de seleccionar modelos climáticos es reducir la incertidumbre en la cadena de modelación, y que, por construcción, utilizar un enfoque basado en la envolvente de modelos contribuye a la dispersión de los campos climáticos, en el presente estudio se propone una metodología de selección basada en el desempeño histórico de los modelos.

Por otro lado, la incertidumbre de los modelos climáticos puede asociarse a principalmente dos fuentes, (i) la respuesta del modelo y (ii) la variabilidad interna del sistema (Brunner et al., 2020; Rupp et al., 2013).Para identificar la incertidumbre asociada a la primera fuente se suele utilizar un set de métricas que cuantifique el error del modelo con respecto a las observaciones, mientras que para cuantificar la incertidumbre asociada a la variabilidad interna se analiza la dispersión de las distintas simulaciones de un mismo modelo. Por ejemplo, Rupp et al. (2013) cuantificaron la incertidumbre de ambas fuentes, proponiendo un set de métricas clasificándolas según su robustez. Las métricas más robustas (i.e. aquellas que resultan en el mismo valor independiente de la corrida del modelo) identificadas en dicho estudio son el promedio, la desviación estándar y la correlación espacial de valores climatológicos.

A pesar de que las métricas utilizadas en los estudios de selección de modelos son variadas, se pueden identificar similitudes. Por ejemplo, en Chiew et al. (2009) propusieron, en la misma línea de Rupp et al. (2003), métricas que, además de evaluar la climatología promedio, diagnostiquen las tendencias de las series anuales. Además de la reproducción de la climatología, Karmalkar et al. (2019) , propusieron el diagnóstico de métricas basadas en procesos para identificar la correcta simulación de patrones de larga escala (e.g., ENSO, PDO, etc). Las métricas de diagnóstico propuestas en este estudio buscan aunar los criterios utilizados en investigaciones pasadas. Debido a que los métodos de escalamiento suelen corregir adecuadamente el sesgo del modelo climático (Gutmann et al., 2014; Maraun, 2013), las métricas no estarán enfocadas en cuantificar la subestimación o sobrestimación de los montos, si no que más bien se enfocarán en cuantificar la reproducción de la variabilidad interanual, las tendencias anuales, la reproducción del ciclo intra-anual y patrones espaciales.

En esta línea, la pregunta científica que será abordada en este estudio es:

¿Es posible elaborar un ranking de modelos climáticos en base a características climáticas históricas, que se mantenga invariable en caso de aplicar escalamiento estadístico?

1.1. Objetivos

Los objetivos generales y específicos del presente trabajo de tesis son los siguientes:

Objetivos Generales

Generar un *ranking* de modelos generales de circulación que facilite la selección para posteriores estudios regionales de cambio climático aplicados a Chile continental.

Objetivos Específicos

- Realizar un diagnóstico de los GCMs basado en el desempeño de éstos en un periodo histórico (1979-2014) mediante métricas que resuman aspectos estructurales de los GCMs sin corrección estadística.
- Generar el diagnóstico para distintas resoluciones espaciales con el objetivo de cuantificar los efectos de seleccionar una resolución.
- Comparar el desempeño histórico del *modelo promedio* con el de los mejores modelos diagnosticados.
- Aplicar técnicas de escalado estadístico a los GCMs para el período histórico y generar un diagnóstico post-escalamiento a la mayor resolución con el objetivo de identificar cuáles aspectos mejoran o empeoran luego de dicho proceso.

II. Artículo para publicación

A continuación, se presenta el artículo titulado "*Systematic selection of global climate models for regional studies based on past-perfomance approach*"*:*

An approach to rank global climate models for regional studies based on past performance

Felipe Gateño[1], Ximena Vargas[1] and Pablo A. Mendoza[1,2]
[1]Department of Civil Engineering, Faculty of Physical and Mathematical Sciences, Universidad de Chile, Chile.
[2]Advanced Mining Technology Center (AMTC), Faculty of Physical and Mathematical Sciences, Universidad de Chile, Chile.

Abstract

Estimating projected hydrologic changes requires several methodological choices, being the selection of General Circulation Models (GCMs) one of the most challenging and crucial, since it may contribute a large spread in hydrological projections. We propose a framework to rank GCMs and facilitate their selection for regional climate change studies. The study considers 28 CMIP6 models (27 models plus one 28th ensemble mean, called MMA), that are diagnosed across continental Chile during the period 1979-2014, using four metrics that evaluate (i) interannual variability, (ii) spatial correlation of the climatological mean, (iii) mean seasonal cycles and (iv) annual trend on the historical series of precipitation and extreme temperatures. These metrics are computed for each GCM and compared with an observational dataset to generate a ranking for each metric and variable, and an overall ranking of models. Additionally, we examine the effects of the horizontal resolution at which the models are diagnosed, re-gridding the GCMs at four different resolutions (0.25°, 0.50°, 0.75° and 1.00°) and conducting the diagnostic at each resolution. Finally, we explore the influence of statistical downscaling (SD) on GCM selection using two SD methods. The results show that no model demonstrates superiority in all variables and metrics simultaneously, either before or after the SD process. The results show that the GCM ranking has little sensitivities to the diagnostic resolutions. Finally, the two statistical downscaling methods yield an improvement in the climatological annual cycle and the spatial correlation with the observational dataset; however, they fail to correct errors in the historical trends. Considering the spatial variability of the performance of the GCMs, it is recommended to carry out, on a regional scale, a performance-based selection which considers multiple aspects and variables of the meteorology, diagnosing, at least, the reproduction of the annual cycle, the interannual variability and the trends in the historical period.

1. Introduction

Understanding climate change and its impacts on water resources is crucial to design mitigation and adaptation strategies (IPCC, 2021). . Further, quantifying climate change impacts on hydrological variables requires several methodological choices (Clark et al., 2016), including: (i) selection of climate change scenario (Lehner et al., 2020; Liang et al., 2020; Shen et al., 2018); (ii) selection of general circulation models (GCMs, Karmalkar et al., 2019; Lutz et al., 2016; Rupp et al., 2013); (iii) choice of initial conditions of the GCMs

(Ahmed et al., 2019; Hakala et al., 2018); (iv) choice of downscaling technique (Gutmann et al., 2014; Vaittinada Ayar et al., 2016); (v) selection of hydrological model structures (Clark et al., 2011; Knoben et al., 2020); and (vi) selection of hydrologic model parameters (Liu et al., 2017; Mendoza et al., 2015).

Previous studies have aimed to shed light on which methodological decisions should be further investigated when projecting changes in specific variables and/or metrics. For example, Shen et al. (2018) identified that GCM selection contributes more than the choice of climate scenario to the spread in mean annual streamflow projections in the Hanjiang River watershed, China. Chegwidden et al. (2019) analyzed four methodological decisions along the modeling chain – Representative Concentration Pathway (RCP), choice of GCM, choice of downscaling method, and hydrologic model set-up – and its effects on the spread of hydrologic change projections in the Columbia River basin, USA. They found that the selection of RCP and GCM are the main contributors to the spread in annual streamflow volume and timing. Similar results were obtained in other investigations (Wang et al., 2020; Zhang et al., 2021), concluding that GCM selection is one of the key contributors to the uncertainty in hydrologic change projections.

Hence, there is increasing interest in the community to develop and test techniques to select GCMs, which can be classified in two main groups: (i) past performance approaches, and (ii) envelope-based approaches. Past performance techniques (Brunner et al., 2020; Rupp et al., 2013), consist of evaluating GCMs with different metrics over a historical period. Because there is a non-stationary relationship between climate models and observations (Chen et al., 2015; Gutmann et al., 2014; Maraun, 2016), selecting GCMs with good historical performance does not warrantee that the same will happen in a future period (Wang et al., 2019). On the other hand, envelope-based approaches (Lutz et al., 2016; Sa'adi et al., 2020), aim to select those GCMs that cover the full range of possible future scenarios, which implies that models with moderate change signals, which can still represent provide a plausible scenario, may be dismissed (Chen et al., 2015; Gutmann et al., 2014; Maraun, 2016; Wang et al., 2019).

Because the primary goal of selecting climate models is to reduce uncertainty in the modeling chain, and envelope-based approaches contributes – by definition – to increased spread in climate projections, most previous efforts encourage the use of a historical performance-based approach (e.g.; Brunner et al., 2020; Chiew et al., 2009; Karmalkar et al., 2019; Pierce et al., 2009; Rupp et al., 2013; Shen et al., 2018; Wang et al., 2019). Because the ranking of models depends on the variable and metric analyzed (Gleckler et al. 2008), a robust assessment of GCMs requires defining a set of metrics that quantify different aspects defining the quality of simulated climate fields (Gleckler et al., 2008). For example, Rupp et al. (2013) used a suite of metrics to assess the capability to reproduce seasonality, long-term variance, regional teleconnections, and other features for precipitation and temperature. Further, they analyzed the intra-model variability by calculating the metrics for different runs of the same model, identifying the climatological mean and the spatial correlation between the mean observed values and the mean simulated values as the most robust metrics (i.e., metrics that do not vary significantly for different runs of the same model), being the metrics associated with regional teleconnections the least robust. In the same line, Chiew et al. (2009) and Pierce et al. (2009) quantified the model error in simulating the climatological mean through MSE, and the interannual variability using the coefficient of variation of the annual timeseries

or the standard deviation of the seasonal mean. Other studies have proposed the evaluation of historical trends, by fitting a linear regression to the annual timeseries (e.g.; Brunner et al., 2020; Tokarska et al., 2020). However, Rupp et al. (2013) found that modeled trends are highly sensitive to the run within the same model. Consequently, the intramodel variability is too high to comfortably rank models. Similar conclusions were proposed by Brunner et al. (2020).

On the other hand, (Towler et al., 2010; Vano et al., 2015) a post-processing step must be conducted on climate model outputs to address the mismatch between the horizontal resolution of GCMs and the scale of interest for water resources applications (Towler et al., 2010; Vano et al., 2015). Although, downscaling techniques are applied to reduce the absolute bias in mean fields (Gutmann et al., 2014; Wood et al., 2004), some authors have shown that such procedures cannot fix GCM weaknesses such as orographic misrepresentations, the lack of feedbacks with the land surface and/or sub-grid processes (Maraun, 2016; Vaittinada Ayar et al., 2016). Additionally, the downscaling process may affect the performance of climate models and, therefore, the results from the selection process. Recently, Wang et al. (2019) analyzed the effects of downscaling on the relative performance of GCMs. They used raw GCM outputs to simulate streamflows using different weightings schemes based on historical performance and repeated the process using bias-corrected GCM outputs. The weights obtained using raw models were unequal identifying models with better performance than others prior to downscaling (i.e., the GCMs with better historical performance obtained more weight than the GCMs with poorer performance). On the other hand, the weights calculated using the bias-corrected outputs are similar to the weights before downscaling, concluding that the downscaling process affects the ability to discriminate models.

Despite the existing body of work on GCM selection, there is no consensus on how to select GCMs (i.e., what diagnostic metric to consider) and the effect of the downscaling process on GCMs selection. Consequently, is it possible to elaborate a ranking of climate models based on their historical performance, which remains invariable in case of applying statistical downscaling? To seek for answers, we evaluate and rank the performance of 27 raw CMIP6 GCM outputs in reproducing historical precipitation and extreme temperatures considering four aspects: (i) interannual variability; (ii) intra-annual cycle; (iii) spatial correlation; and (iv) historical trends. We also bias-correct the models using two statistical downscaling methods and analyze the effects of including this process on the ranking.

2. Study domain

Our study domain is continental Chile (17°S-56°S, Figure *1*a) whose weather and climate are affected by the semi-permanent subtropical anticyclone over the Southeast Pacific, and a westerly wind regime at mid-latitudes (Aceituno et al., 2021). These features combine to generate a marked north-south precipitation gradient, with mean annual values ranging from less than 10 mm in the north to more than 3000 mm in the southern part of continental Chile (Figure 1d). Additionally, the Andes Cordillera disrupt the atmospheric circulation, resulting in a variety of mesoscale phenomena, from the Chilean Altiplano affected by the Bolivian High and water advection from the Amazonas, to frontogenetical activity in mi latitudes (Aceituno et al., 2021; Garreaud, 2009). Because

the Andes is present in almost all the Chilean territory, a pronounced west-east orographic gradient that decreases towards the south (Figure 1b and 1c) affects the spatial distribution of precipitation (Uribe et al., 2012; Viale et al., 2019).

Additionally, strong latitudinal and longitudinal temperature gradients are observed across the study domain, being the latter attributed to the marine front (Araya-Osses et al., 2020). Temperatures are colder in southern Chile (where averaged maximum daily values below 0°C have been measured) compared with northern Chile (where averaged maximum temperatures >30°C are frequently observed, see Figure 1e). Averaged minimum daily temperatures show a similar behavior (Figure 1f) ranging from approximately 10°C in the north to -10°C in southern Chile.

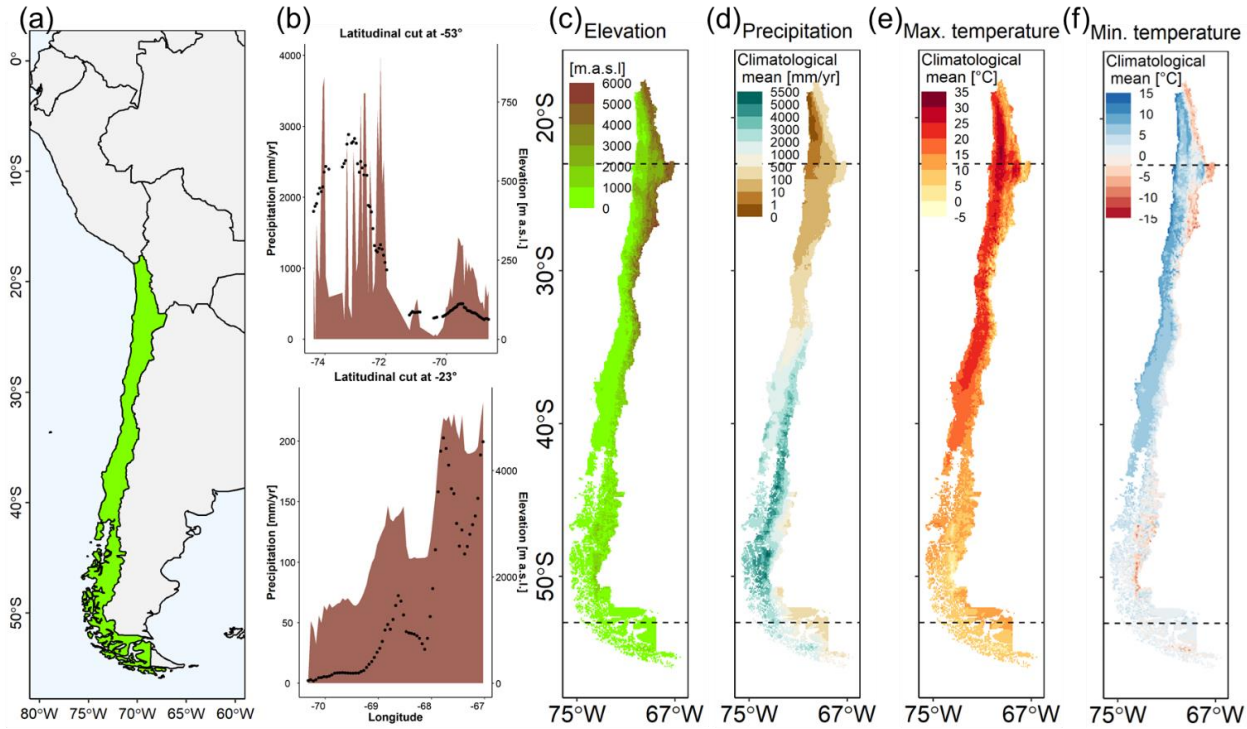

Figure 1: General description of study area: (a) location of Chile in South America; (b) longitudinal variation of mean annual precipitation (black dots) and elevation (brown area) for latitudes displayed as black lines in the following panels; (c) elevation map; (d) spatial distribution of CR2MET (v2.0) (d) precipitation, (e) maximum and (f) minimum temperature for period 1979-2018.

3. Data

The observational, reference dataset is a corrected version of the gridded product CR2MET v2.0 (Alvarez-Garreton et al., 2018; Boisier et al., 2018; available at https://www.cr2.cl/datos-productos-grillados), which contains time series of daily precipitation (Pr), and maximum (Tmax) and minimum (Tmin) daily temperatures at a horizontal resolution of 0.05°x0.05° for continental Chile from 1979 to 2018. CR2MET v2.0 builds upon a statistical downscaling technique that uses ERA5 atmospheric reanalysis variables and topographic indices as predictors, and in-situ observations as predictands. For temperature estimates, surface temperature from MODIS LST is also included. ERA5 usually underestimates precipitation amounts along the Chilean territory

due to its coarse spatial resolution (DGA, 2017) and, therefore, CR2MET v2.0 precipitation estimates inherit this problem (DGA, 2019) due to the lack of measurements in the Andes Cordillera to correct ERA5 biases. Hence, we use a corrected dataset, which incorporates the undercatch adjustment approach by Yang et al. (1998).

Although CMIP6 includes approximately 60 models, only those containing extreme temperatures and precipitation at daily time steps are used (i.e., 27 GCMs at the time of this writing, listed in Table 1). It should be noted that the horizontal resolutions of GCMs selected are between 0.7° and 2.8°. We include in all analyses MMA results obtained from daily averages of 27 GCM outputs. Table *1*Because historical GCM simulations span from 1850 to 2014, we define our analysis period from Jan/1979 to Dec/2014 (35 years). To (i) standardize the spatial resolution of the GCMs and to (ii) analyze its effect on the diagnostic, the models were bilinearly interpolated to four spatial resolutions (0.25°, 0.50°, 0.75° and 1.00°). The observational dataset was also aggregated to those spatial resolutions.

Table 1: Models included in the analysis.

ID	Model	Institution(s)	Latitudinal resolution (°)	Longitudinal resolution (°)
1	ACCES.CM2	CSIRO-ARCCSS	1.25	1.875
2	ACCESS.ESM1.5	CSIRO	1.25	1.875
3	BCC.CSM2	BCC	1.125	1.125
4	AWI.ESM.1.1.LR	AWI	1.85	1.875
5	CanESM5	CCCMa	2.8	2.8
6	CMCC	CMCC	0.95	1.25
7	EC.Earth3.AerChem	EC-Earth consortium	0.7	0.7
8	EC.Earth3.CC		0.7	0.7
9	EC.Earth3.Veg.LR		1.125	1.125
10	EC.Earth3.Veg		0.7	0.7
11	EC.Earth3		0.7	0.7
12	FGOALS.f3.L	CAS	2	2
13	FGOALS.g3		2	2
14	GFDL.CM4	NOAA-GFDL	1	1.25
15	GFDL.ESM4		1	1.25
16	IITM.ESM	CCCR-IITM	1.9	1.875
17	INM.CM4.8	INM	1.5	2
18	INM.CM5		1.5	2
19	IPSL.CM6A	IPSL	1.25	2.5
20	KIOST.ESM	KIOST	1.875	1.875
21	MIROC6	JAMSTEC, AORI, NIES, RCCS	1.4	1.4
22	MPI.ESM.1.2.HAM	HAMMOZ-Consortium	1.5	1.5
23	MPI.ESM1.2.HR	MPI-M	0.935	0.9375
24	MPI.ESM1.2.LR	MPI-M	1.8	1.875
25	NorESM2	CICERO-METNorway-NILU-UNI-UiB	1.9	2.5
26	NESM3	NUIST	1.865	1.875
27	TaiESM1	AS-RCEC	0.9	1.25
28	MMA	-	-	

4. Methods

Our analysis is based on three main components (Figure 2): (i) spatial and temporal (dis)aggregation of modeled and observed datasets; (ii) GCM diagnostics and ranking of raw model outputs (i.e., prior to the downscaling process); (iii) two statistical downscaling methods (SDMs) at 0.25° horizontal resolution; and (iv) re-visit GCM diagnostics and ranking after the downscaling process. The same metrics were used for diagnosing the GCMs in steps (ii) and (iv) (Figure 2).

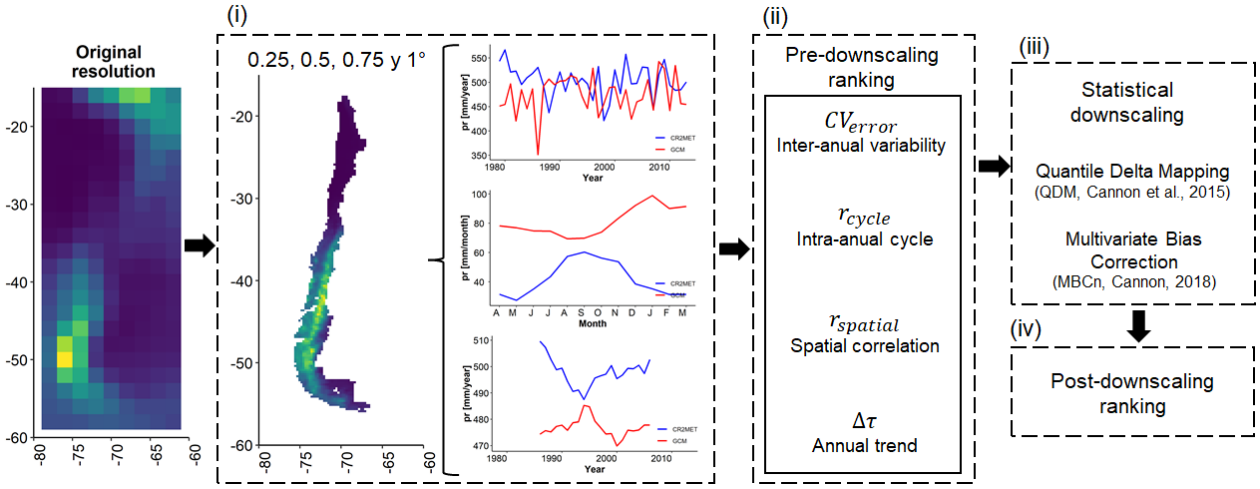

Figure 2: Schematics of the methodology: (i) spatial and temporal data aggregation (blue and red lines represents observed and modeled time series for an arbitrary pixel respectively); (ii) diagnostics prior to the downscaling process; (iii) application of Quantile Delta Mapping (QDM) and Multivariate Bias-Correction (MBCn); and (iv) diagnostics after downscaling process).

4.1.Spatiotemporal aggregation

Model outputs are bilinearly interpolated and observations are spatially aggregated to 0.25°, 0.50°, 0.75° and 1.00° latitude–longitude resolutions to (i) harmonize differences between GCM outputs and observations and (ii) to examine the effects of horizontal resolution on GCM evaluation results and ranking. The resulting spatially (dis)aggregated daily time series of precipitation and extreme temperatures from GCMs and observations are temporally aggregated (i.e., added in the case of precipitation, and averaged in the case of maximum and minimum temperature) to monthly and annual time steps. Additionally, climatological monthly averages are calculated for each month to evaluate the ability of GCMs to replicate the seasonality of precipitation and extreme temperatures.

4.2. GCM diagnostics

Since GCMs are not developed to ensure that historical modeled fields and observations are correlated in space and time (Dettinger et al., 2004; Gutmann et al., 2014; Pierce et al., 2014), with the potential of inducing systematic errors, our choice of metrics aims to assess the ability of GCMs to characterize climatic patterns from seasonal to larger timescales (Flato et al., 2013; Maraun, 2016; Randall et al., 2007), and preserve the spatial variability of observations.

First, we use the coefficient of variation (CV) to assess the ability of GCMs to simulate the observed interannual variability. Hence, the coefficient of variation from modeled (CV_{GCM}) and observed (CV_{OBS}) annual time series is calculated for each grid cell using annual amounts for Pr and annual averages for Tmax and Tmin, respectively. Then, we compute the difference$CV_{error} = CV_{GCM} - CV_{OBS}$, where values close to 0 indicate that the GCM reproduces adequately the interannual variability, while positive (negative) values indicate an overestimation (underestimation).

We diagnose the ability of GCMs to reproduce trends in observed annual time series using the non-parametric Mann Kendall's test (Mann, 1945). First, modeled and observed annual time series at each grid cell are smoothed using a moving average of 10, 15 and 20 water years. Then, we apply the Mann Kendall's test (using *trend* package available in R; Pohlert, 2020), obtaining τ_{GCM} and τ_{OBS}.The values for the τ statistic range between -1 and 1, where positive (negative) values close to 1 (-1) indicate that there is a monotonically increasing (decreasing) trend, while values close to 0 reflect the absence of trends. Using these outputs, we compute the second metric:

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

Where $sgn(\tau_{\mathrm{OBS}})$ and $sgn(\tau_{\mathrm{GCM}})$ are the sign of τ_{OBS} and τ_{GCM}, respectively. Positive values of $\Delta\tau$ indicate that the GCM adequately reproduces the sign of an observed trend, while negative values show the opposite. Additionally, $\Delta\tau \sim 0$ indicate that the magnitude of the modeled trend is similar to the observed one. Hence, small and positive $\Delta\tau$ values are preferred.

We also compute the Pearson correlation coefficient (r_{cycle}) between modeled and observed mean monthly values (i.e., using 12 pairs of model outputs and observations) to assess the ability of GCMs to reproduce seasonal variations in climatic variables. The r_{cycle} is estimated for each grid cell.

Finally, we evaluate the capability of GCMs to reproduce observed spatial patterns across the study domain. First, we calculate the climatological mean for the entire period (from annual amounts in the case of Pr, and annual averages for Tmax and Tmin) and each grid cell, obtaining maps of mean values for all GCMs and the observations. Then, we rearrange these maps as vectors, and compute the Pearson correlation coefficient $r_{spatial}$ between model simulations and observations. A model who perfectly reproduce the spatial distribution at the climatological scale will yield a value of $r_{spatial} = 1$.

4.3. Ranking approach

The metrics described in section 4.2 are computed for each combination of GCM, variable, and horizontal resolution (0.25°, 0.50°, 0.75° and 1.00°). These results are used to rank the GCMs based on their historical performance across the domain for each metric. For the case of $r_{spatial}$, the GCM that yields the value closest to 1 is assigned the first position, and the model that yields the lowest value gets the 28th position. For the rest of the metrics, given a pair of variable and resolution, we develop a ranking as follows (Figure *3*): (a) each metric is calculated for each grid cell; (b) an empirical cumulative distribution function (CDF) is obtained using metric values from all grid cells; and (c) the area between

the CDF and the optimal value is calculated. The GCM that yields the smallest area provides the overall best performance across the domain of interest, and therefore gets the first position in the ranking for each metric and each variable. Once the rankings for the four metrics are constructed, (d) a ranking for each variable (Pr, Tmax and Tmin) is computed as the mean among each metric, assigning equal weight for each pair of variable-metric. Finally, (e) the ranking of GCMs (total ranking) is calculated ordering from lowest to highest the average ranking among all variables.

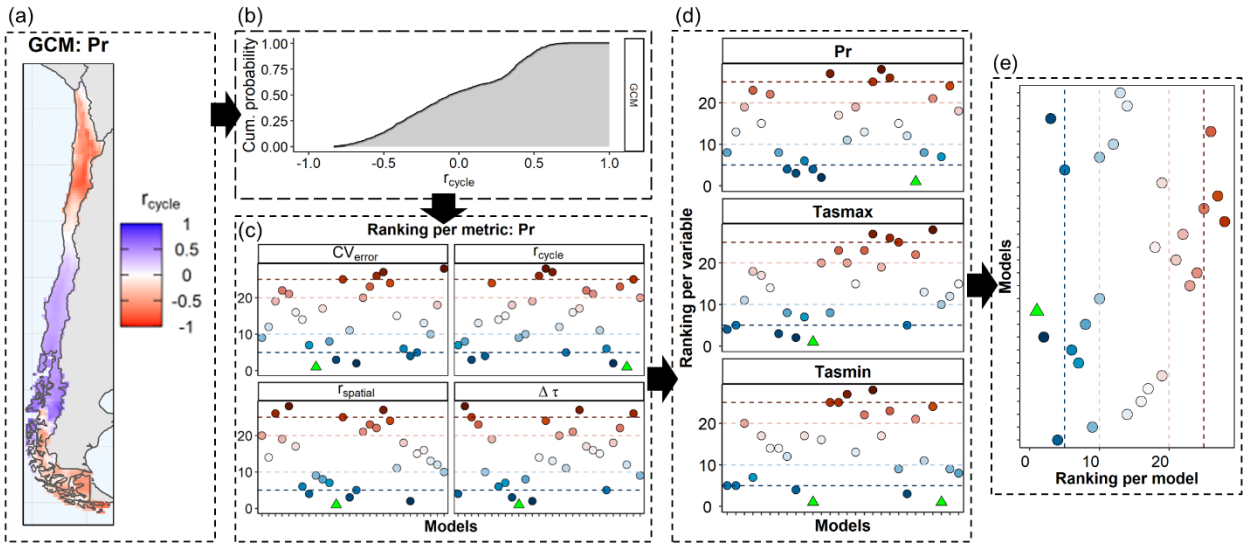

Figure 3: Illustration example of our ranking methodology. (a) shows r_{cycle} values obtained for an arbitrary model. (b) shows empirical CDF from the results of (a), and the area between the CDF and the desire value of the metric is calculated (gray area). (c) Ranking calculation for all metrics. (d) the averaged ranking per variable for each metric, and finally, (e) the total ranking is obtained as the mean of ranks per variable. In the ranking, blue (red) dots represent the best (worst) models, and green triangles represent the best model.

4.4. Statistical downscaling techniques

To test the effects of including the downscaling process in the selection of GCMs, we apply two SDMs: (i) the Quantile Delta Mapping (QDM) method (Cannon et al., 2015) and (ii) the N-dimensional Multivariate Bias Correction (MBCn) method (Cannon, 2018). To apply QDM, a cumulative distribution function is first computed using validation model time series $x_{m,p}$, and a non-exceedance probability $\tau_{m,p}(t)$ associated with the value at time t:

$$\boldsymbol{\tau_{m,p}(t) = F_{m,p}^{(t)}[x_{m,p}(t)]} \quad (2)$$

Where $F_{m,p}^{(t)}[]$ is the cumulative distribution function of the time series in the validation period. The value in the calibration period associated with that non exceedance probability ($\tau_{m,p}(t)$) is obtained (i.e., $F_{m,h}^{-1}[\tau_{m,p}(t)]$) and the relative change in quantiles between the calibration and validation periods is computed following:

$$\boldsymbol{\Delta_m(t) = \frac{x_{m,p}(t)}{F_{m,h}^{-1}[\tau_{m,p}(t)]}} \quad (3)$$

The same procedure is applied to the observed time series in the calibration period to compute bias-corrected value:

$$\hat{\mathbf{x}}_{\mathbf{o,h}}(\mathbf{t}) = \mathbf{F}_{\mathbf{o,h}}^{-\mathbf{1}}[\boldsymbol{\tau}_{\mathbf{m,p}}(\mathbf{t})] \quad (4)$$

Finally, the bias-corrected time series in the validation period at time t is given by applying the relative change $\Delta_m(t)$ (multiplicatively for precipitation, additively for temperature) to the bias-corrected value in the calibration period. For example, the precipitation in the validation period is obtained as:

$$\hat{\boldsymbol{x}}_{\boldsymbol{m,p}} = \hat{\boldsymbol{x}}_{\boldsymbol{o,h}} \cdot \boldsymbol{\Delta}_{\boldsymbol{m}}(\boldsymbol{t}) \quad (5)$$

The MBCn method is a generalization of QDM, allowing the simultaneous bias-correction of multiple variables in two steps: (i) application of an orthogonal rotation to the target data (i.e., the modeled fields); and (ii) correction of the marginal distributions via QDM. These steps are repeated iteratively until the multivariate distribution matches the distribution of observed historical fields (Cannon, 2018). The main benefit of MBCn is that it allows correcting the intra-variables dependence and is not restricted to correct a specific measure of joint dependence. A full description of both QDM and MBCn methods can be found in Cannon et al. (2015) and Cannon (2018), respectively, and the R package implementing the QDM an d MBCn algorithm is available for download in https://cran.r-project.org/web/packages/MBC (Cannon, 2020).

We apply the above techniques at a daily scale following a k-fold cross-validation approach, using six non-overlapping periods of six years each (over the period April/1979-May/2015), leaving the remaining 30 years for training.

5. Results

5.1. Ranking sensitivity to horizontal resolution

The ranking of GCMs before the downscaling process is shown in Figure 4. The best GCM in the overall ranking is the EC.Earth3 model (model ID 11 in Figure 4a), regardless of the horizontal resolution to which it was re-gridded . However, this model does not get the best position in any metric calculated for Pr (Figure 4b), although it gets the first position in the ranking for CV_{error} and $\Delta\tau$ in the case of Tmax (Figure S1), and for $r_{spatial}$ and $\Delta\tau$ in the case of Tmin (Figure S2). For the rest of the metrics, the EC.Earth3 model yields a moderate performance, obtaining positions between 5 and 12, and reaching the 23rd position for CV_{error} calculated for Tmin at a 1° resolution (Figure S2). Similar results are obtained for the following best models (models ID 1, 9 and 26 for ACCES.CM2, EC.Earth.Veg.LR and NESM3, respectively); although they all get a good position in the overall ranking (from 2nd to 4th places), only in a few cases these models are among the four highest ranks and, in general, these GCMs are associated with moderate performance.

On the other hand, INM.CM5 (model ID 18) gets the 28th position of the overall ranking, with the worst position for most combinations of metrics and variables, excepting r_{cycle} for Pr and CV_{error} for Tmax and Tmin, for which moderate performance is obtained. Similar performances are obtained with IPSL.CM6A (model ID 19), KIOST.ESM (model

ID 20) and NorESM2 (model ID 25), which yield general ranks between 24 and 27 in the total ranking (Figure 4a).

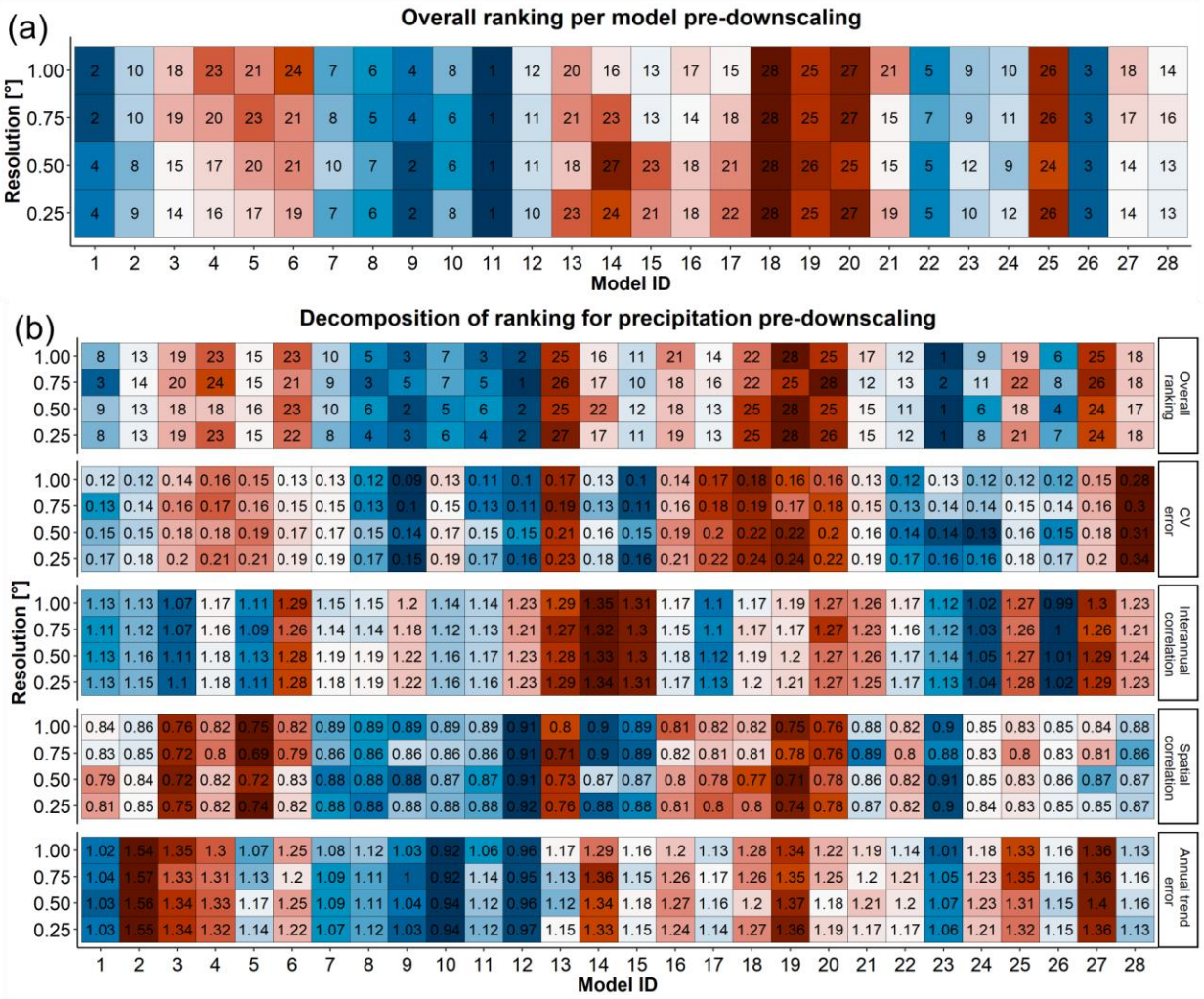

Figure 4: Ranking pre-downscaling for each resolution: (a) overall ranking per model and (b) overall ranking (first row) and ranking per metric (last four rows) calculated for precipitation. In (a) and the first row of (b), colors and numbers represent the position of the ranking for a given resolution, GCM and metric. For CV error, Interannual correlation and Annual trend error rows (b), numbers indicate the area between the CDF (Figure 3b) obtained per model, while the numbers in Spatial correlation row indicate the Pearson's correlation between GCMs and observations.

5.2. Effects of statistical downscaling on GCM rankings

After the downscaling process using both MBCn and QDM methods, the four diagnostic metrics are computed at a 0.25° spatial resolution, and a new ranking is generated. The results are similar for SDMs (Figure 5a): the best model after downscaling is ACCES.CM2 (model ID 1), with INM.CM4.8 (model ID 17) and MPI.ESM.1.2.HAM (model ID 22) in the 2nd and 3rd places, and the worst models are NorESM2 and INM.CM5 (models ID 25 and 18, respectively), with the latter obtaining the 28th position.

Figure 5a shows pronounced differences between the overall ranking prior and posterior SDMs, with a Pearson's correlation coefficient of 0.14 and 0.17 when is calculated between the overall ranking of raw outputs and statistical downscaled outputs (with QDM and MBCn, respectively). This disagreement is also observed in the ranking per variable (first row of Figure 5b): the model with the best overall ranking for precipitation (MPI.ESM1.2.HR, model ID 23) is relegated to the 6th position after a SDM is applied; the best model after SDM (ACCES.CM2, model ID 1) had 8th position before downscaling; and CanESM5 (model ID 5), a model with medium performance in the overall ranking of precipitation before downscaling, gets a rank of good performance after SDM (rank 4). Similar results are obtained for overall ranking of Tmax and Tmin (first rows in Figure S3 and Figure S4), and for ranking positions in CV_{error}, r_{cycle} and $r_{spatial}$ (second to fourth rows in Figure 5b, Figure S3 and Figure S4). However, an agreement is observed between the rankings for Δτ before and after applying SDMs: a Pearson's correlation coefficient of 0.76 between raw and downscaled outputs when is calculated for Pr (Figure 5b, last row); a correlation of 0.9 in Δτ ranking for Tmax (last row in Figure S3); and a correlation 0.71 in Δτ ranking for Tmin (last row in Figure S4).

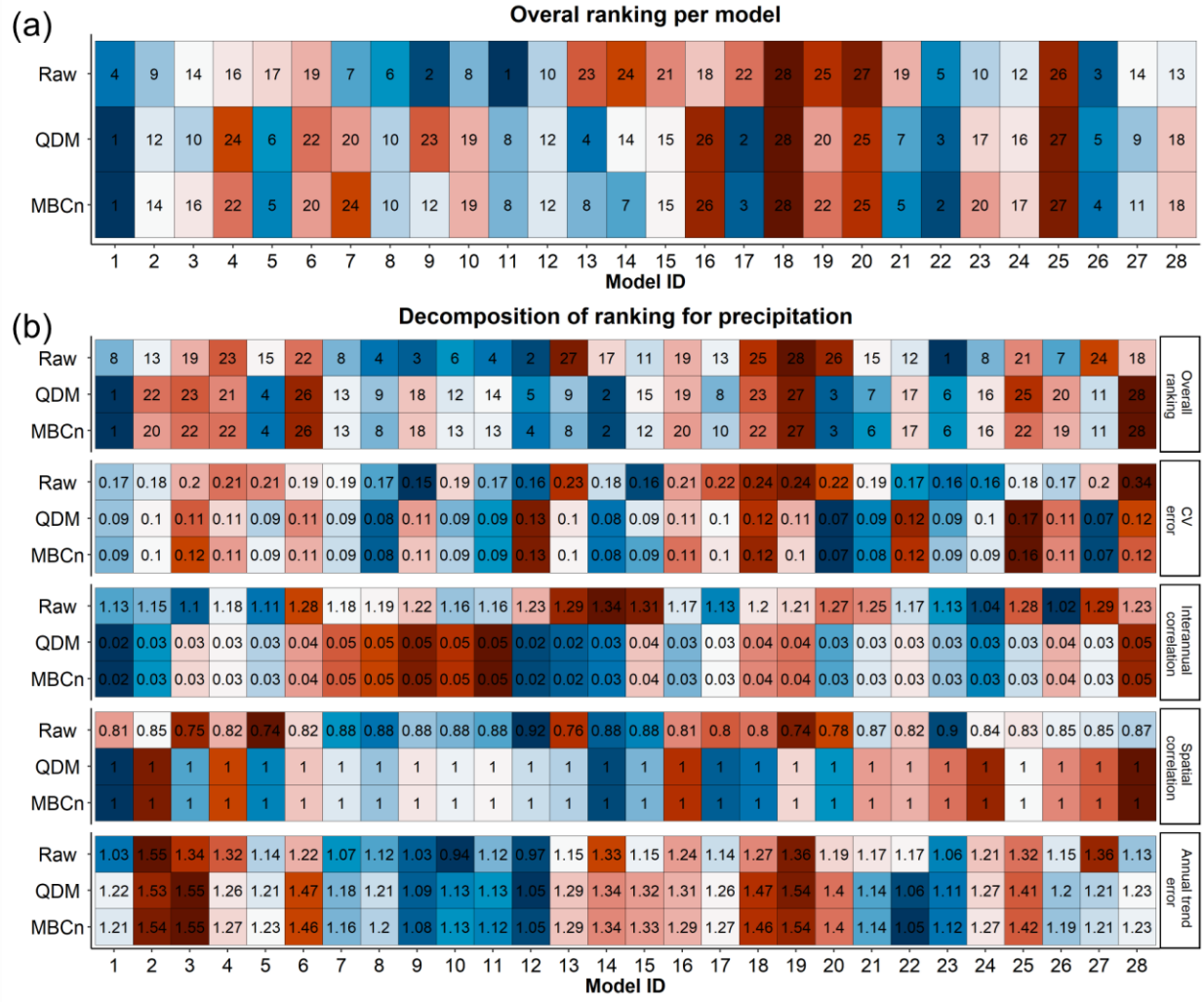

Figure 5: (a) Overall ranking per model before (raw) and after (QDM and MBCn) the downscaling process. (b) Overall ranking (first row) and ranking per metric (last four rows) calculated for precipitation, before (raw) and after the downscaling process (QDM

and MBCn). In (a) and the first row of (b), colors and numbers represent the position of the ranking for a given resolution, GCM and metric. For CV error, Interannual correlation and Annual trend error rows (b), numbers indicate the area between the CDF (Figure 3b) obtained per model, while the numbers in Spatial correlation row indicate the Pearson's correlation between GCMs and observations.

To disentangle the effects of statistical downscaling on GCM output and diagnostic metrics, we select three GCMs based on their overall ranking before the downscaling process (Figure 4): EC.Earth3 (rank 1, model ID 11), TaiESM1 (rank 14, model ID 27) and INM.CM5 (rank 28, model ID 18). The results are displayed for models prior to the downscaling process (raw models) and for statistically downscaled (through the MBCn method) GCM outputs.

Figure 6 shows the spatial distributions of mean annual Pr (period 01/1979-01/2014) for observations, raw and statistically downscaled GCM outputs. In general, the raw GCMs are able to reproduce the latitudinal precipitation gradient (Figure S*5*); however, only EC.Earth3 adequately simulates the amounts observed in central-southern Chile (EC.Earth3 in Figure 6b). Similar results are obtained for maximum and minimum temperature (Figure S*6* and Figure S*7*), for which the GCMs adequately represent the latitudinal gradient, but only a set of models reproduce the longitudinal variation. Figure 6 also shows that the downscaling process yields considerable improvements in the reproduction of the climatology for Pr (Figure 6c), Tmax and Tmin (not shown). Considering all the 28 models, the minimum value of $r_{spatial}$ obtained prior to the downscaling process are 0.69, 0.57 and 0.51, for Pr, Tmax and Tmin, respectively, while the minimum value is 0.99 in all the variables when is calculated posterior to downscaling (for both MBCn and QDM methods).

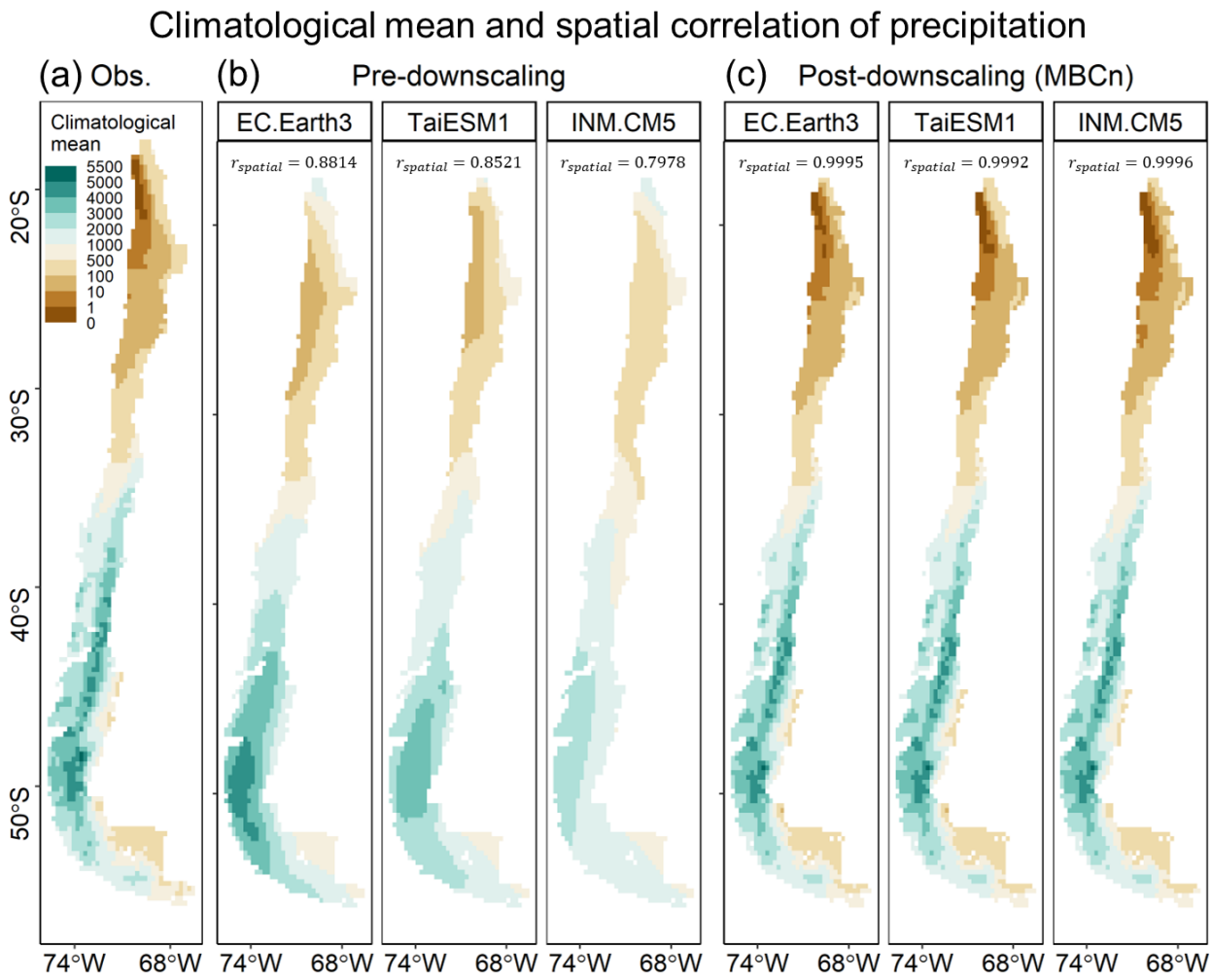

Figure 6: (a) Observed climatological mean of annual precipitation at a 0.25° horizontal resolution for (b) raw GCMs and (c) statistically downscaled GCM output using the MBCn method. Results are displayed for three models sorted by their position in the raw GCM ranking: EC.Earth3 (first position in the overall ranking and first position in the ranking of $r_{spatial}$); TaiESM1 (14th position in the overall ranking and 12th position in the ranking of $r_{spatial}$); and INM.CM5 (28th position in the overall ranking and 22th position in the ranking of $r_{spatial}$).

In general, all the models reproduce adequately the interannual variability of Tmax in the coastal and central zone, before (Figure S9) and after SDMs (not shown); Figure 7a illustrates this with three selected models (Figure 7a), and similar results were obtained for Tmin (Figure S10). There are discrepancies between GCMs in the mountain range area, where the models with higher position in the ranking usually underestimate the interannual variability of extreme temperatures (negatives values of CV_{error}). After statistical downscaling is performed, such underestimation is reduced (Figure 7b), and thus the ranking of GCMs for extreme temperatures changes compared with that obtained for raw GCM output (Figure S3 and Figure S4). For example, the INM.CM4.8 (Model ID 17) model obtain the 23rd position for the CV_{error} calculated for Tmax prior to downscaling and obtain the 1st after downscaling. Similar results are obtained for Pr (Figure 6b). It should be noted that the MMA model yields the worst position in this metric when

calculated for precipitation and it still has poor performance after downscaling, regardless of the method used (Figure 6b).

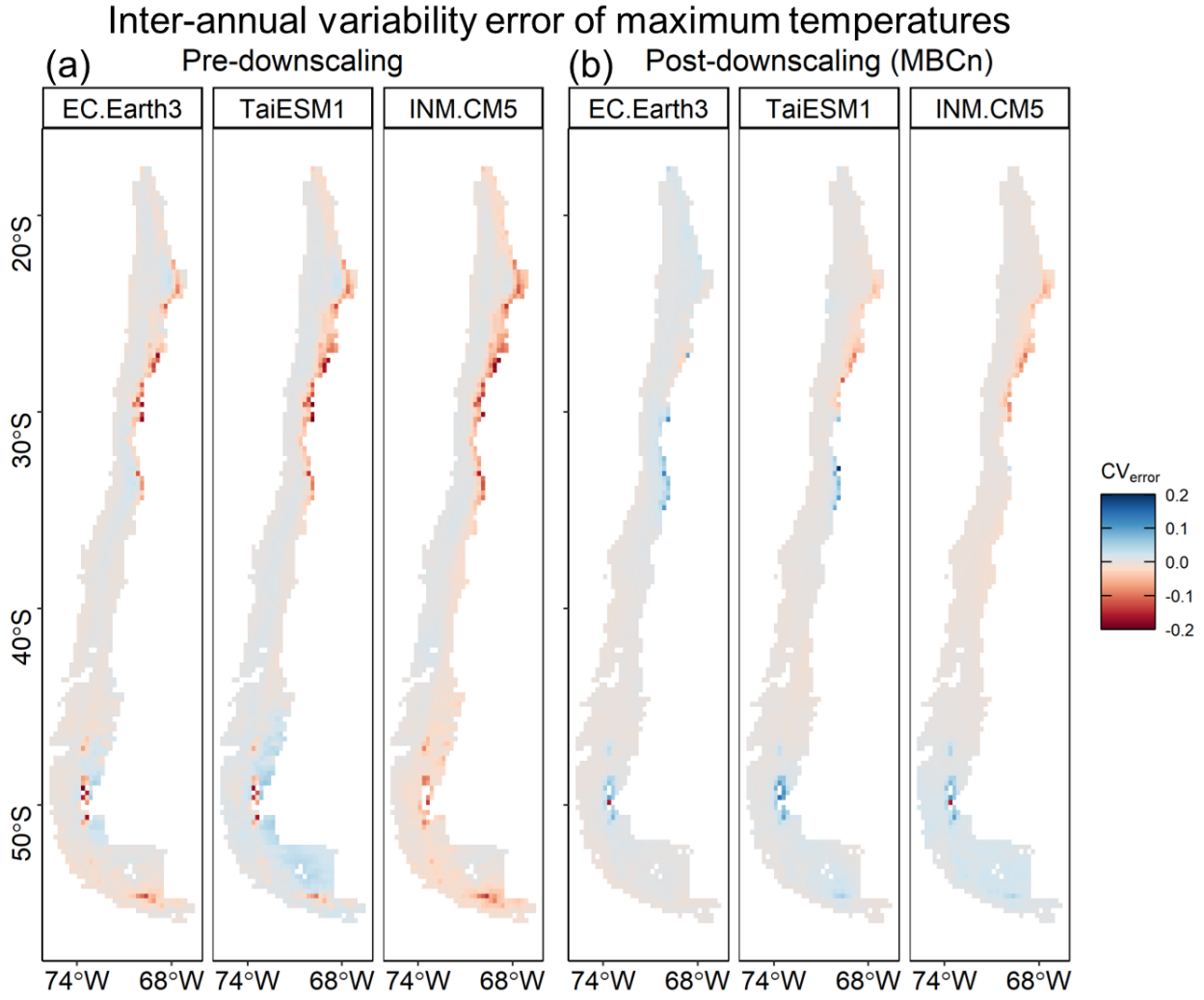

Figure 7: Error in the coefficient of variation of Tmax (CV_{error}) at a 0.25° for (a) raw GCMs and (b) statistically downscaled GCM output using the MBCn method. Results are displayed for three models sorted by their position in the raw GCM ranking: EC.Earth3 (first position in the overall ranking and first position in the ranking of CV_{error}); TaiESM1 (14th position in the overall ranking and 13th position in the ranking of CV_{error}); and INM.CM5 (28th position overall ranking and 15th position in the ranking of CV_{error}).

Figure 8a shows that the three models yield biases in trends of mean annual Tmax along the coastal zone between 15°S and 43°S, and this behavior is also obtained with the 25 remaining models (Figure S*15*). Indeed, modeled trends differ in sign with the observed ones (i.e., negatives values of Δτ), and similar results are obtained around latitude 50°S. In the case of Pr, the spatial distribution of this metric depends on the GCM (Figure S*14*); however, 22 out of the 28 models yield negatives values of Δτ in the western zone between latitudes -50° and -56° (along the coast and Cordillera de la Costa). Additionally, the percentage of pixels where opposite trends are obtained (i.e., with $\Delta\tau \leq 0$) ranges from 7.4% (EC.Earth3.Veg) to 21.8% (ACCES.ESM1.5) for Pr; from 9.4% (BCC.CSM2) to 15% (IITM.ESM) for Tmax; and from 12.1% (INM.CM4.8) to 16.2% (IITM.ESM) in the case of

Tmin. This mismatch in trends persists after applying SDMs (Figure 8b) to both precipitation and extreme temperatures.

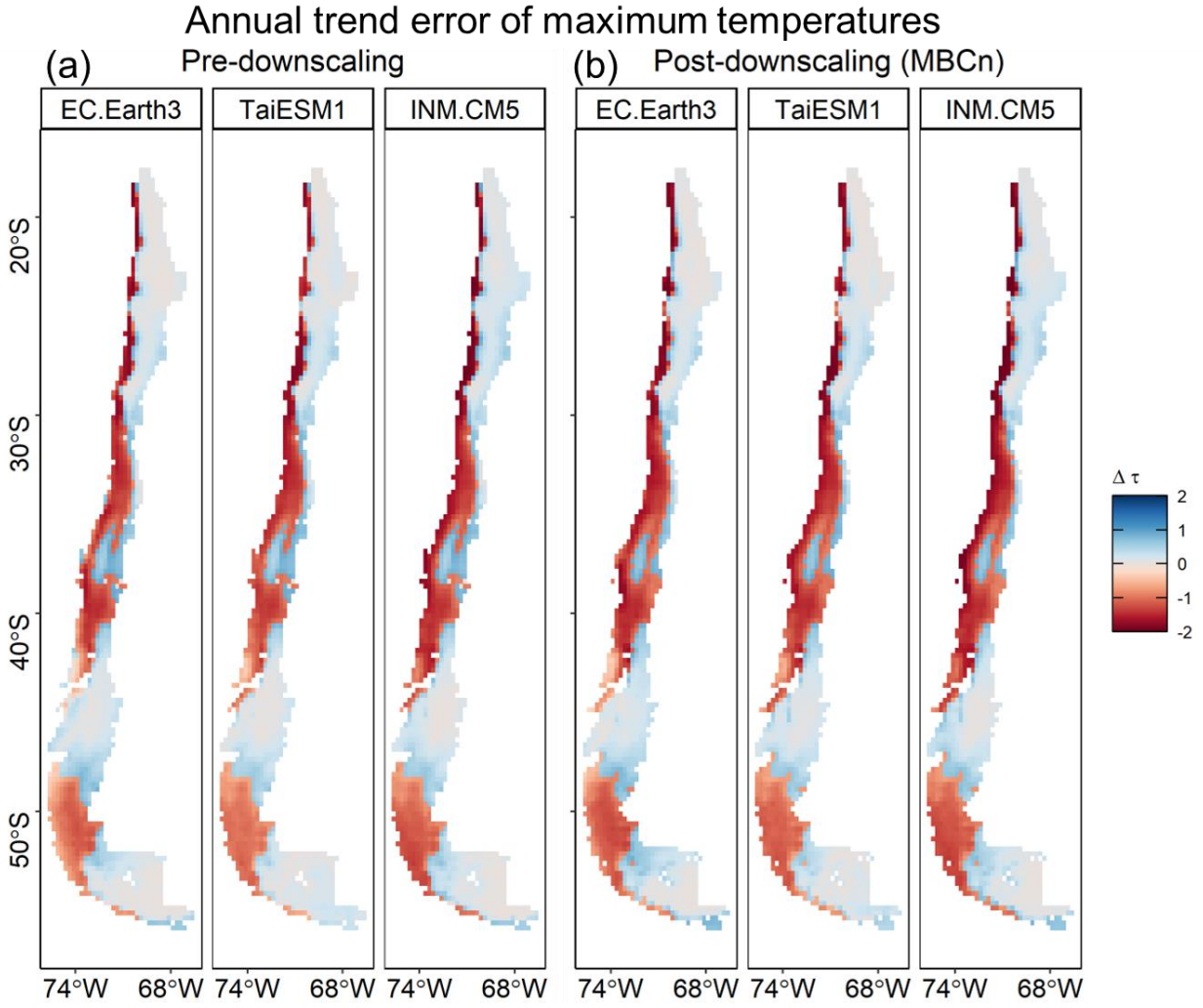

Figure 8: Difference between simulated and observed historical annual trends ($\Delta\tau$) for annually averaged maximum temperature. The results are shown at a 0.25° for (a) raw GCMs and (b) statistically downscaled GCM output using the MBCn method. Results are displayed for three models sorted by their position in the raw GCM ranking: EC.Earth3 (first position in the overall ranking and 5th position in the ranking for $\Delta\tau$); TaiESM1 (14th position in the overall ranking and 10th position in the ranking for $\Delta\tau$); and INM.CM5 (28th position in the overall ranking and 20th position in the ranking for $\Delta\tau$).

A clear difference model performance before and after downscaling can be observed in Figure 9, which shows the correlation between simulated and observed monthly means for extreme temperatures. The best model (EC.Earth3) shows positive values of r_{cycle} for most of the study domain, while the INM.CM5 model shows a disagreement in nearly all continental Chile (Figure 9a). This is reflected in the ranking calculated for this metric, and similar results were obtained for precipitation (Figure 4b). The results obtained for this metric calculated before downscaling are also shown in Figure S*11*, Figure S*12* and Figure S*13*. After the downscaling process, all the models provide a considerable

improvement for all variables, obtaining values of $r_{cycle} \geq 0.9$ in nearly all continental Chile.

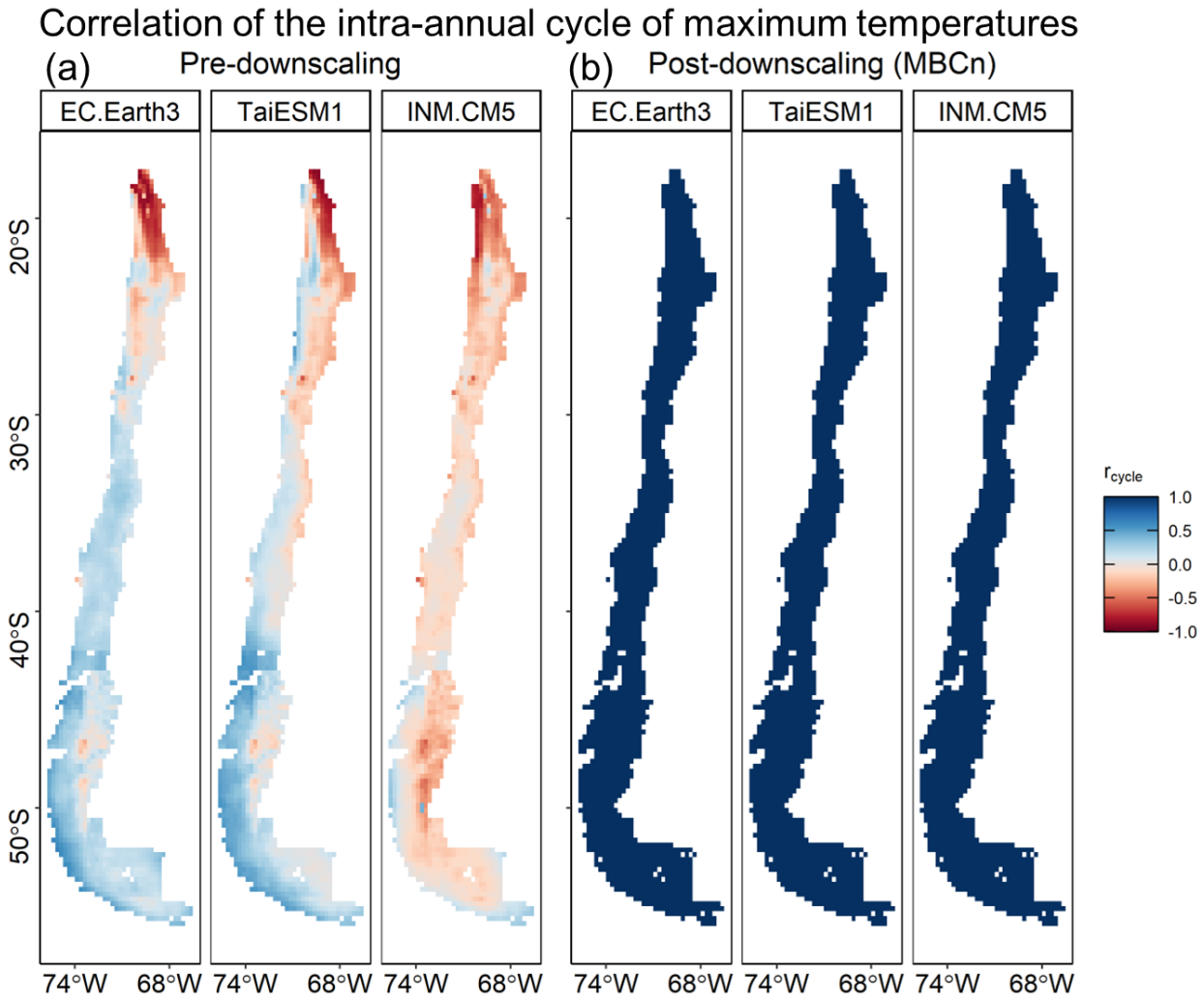

Figure 9: Correlation of the intra-annual cycle (r_{cycle}) for Tmax at a 0.25° for (a) raw GCMs and (b) statistically downscaled GCM output using the MBCn method. Results are displayed for three models sorted by their position in the raw GCM ranking: EC.Earth3 (first position in the overall ranking and 12th position in the ranking for r_{cycle}); TaiESM1 (14th position in the overall ranking and 17th position in the ranking for r_{cycle}); and INM.CM5 (28th position in the overall ranking and 28th position in the ranking for r_{cycle}).

6. Discussion

6.1. Diagnostic metrics

We used the coefficient of variation of the annual timeseries as a measure of the interannual variability (Chiew et al., 2009; Pierce et al., 2009). The long-term variability is a crucial characteristic that affects many aspects of the hydrometeorology, like the streamflow regime (Carrillo et al., 2011; Rubio-Álvarez & McPhee, 2010); snow processes (López-Moreno et al., 2013; McCreight & Small,

2014); and reflects the influence of large scale patterns like El Niño Southern Oscillation (Montecinos & Aceituno, 2003; Poveda et al., 2020) or Pacific Decadal Oscillation (Sagarika et al., 2015). Our results indicate that models reproduce the interannual variability of extreme temperatures better than the precipitation variability, which is consistent with previous studies (e.g.: Blázquez & Nuñez, 2013; Vaittinada Ayar et al., 2016; Kang et al., 2015; Webb et al., 2017). Additionally, a worst representation of the long-term variability for extreme temperatures was identified in the Andes Cordillera regardless the choice of GCM, indicating a poor model performance in zones with complex orography, maybe due to a poor representation of key processes like local circulation or temperature gradients (Espinoza et al., 2020).

The spatial coherence between the observed and modeled variables were tested through the spatial correlation of the climatological mean, $r_{spatial}$. This aspect is especially important for hydrology studies, since the spatial distribution of forcings affects the reliability of the spatial patterns of the hydrological fluxes (Dembélé et al., 2020) and, consequently, impacts the water balance (Clark et al., 2006). Previous studies on GCM diagnostics also recommended to evaluate spatial aspects (Chiew et al., 2009; Karmalkar et al., 2019; Nashwan & Shahid, 2020; Pierce et al., 2009), since spatial correlation of the climatological mean was identified as a *robust* metric that allows to discriminate models with poor performance, independent of the ensemble member of the GCM (Rupp et al., 2013). Our results show that there is no superior model for the three variables.

GCM ranking depends on the selected metrics to describe GCM performance, and the results show that a good raking position in one metric does not necessarily yield a good position in other metrics. Ahmed et al. (2019) ranked 36 CMIP5 GCMs using six spatial metrics, identifying that the best model for one variable (e.g., minimum temperature) can reach the 34th position in another (e.g., precipitation). In our case, according with the $r_{spatial}$ ranking of raw GCMs, the best models for precipitation, maximum and minimum temperature are FGOALS.f3.L (with a horizontal resolution of 90 km), EC.Earth3.AerChem (80 km) and Ec.Earth3 (80 km); while the GCMs with worst position are CanESM5 (250 km), MIROC6 (120 km) and NESM3 (170 km), respectively, suggesting that GCM with finer horizontal resolutions may provide a good representation of the spatial patterns of atmospherics fields. This result aligns well with Liang-Liang et al. (2021), where high resolution models (HighResMIP) were tested in Asia and compared the outputs with their corresponding low resolution GCMs, finding that the bias is smaller for the high resolution models.

For hydrological modelling purposes, it is necessary that GCMs reproduce the seasonality of the variables, particularly precipitation and temperature. This is especially important in a climate change scenario, where the hydrologic cycle can be shifted due to warmer temperatures, changing the partitioning of precipitation between rainfall and snowfall (Alder & Hostetler, 2019; Araya-Osses et al., 2020). An adequate reproduction of the intra-annual cycle in the historical period would allow a reduction in dispersion of projected changes. In this study, we assess the GCMs capability to simulate seasonal patterns through the correlation between the observed and modeled mean monthly values (r_{cycle}). Despite other studies have preferred to calculate the bias of the climatological annual cycle by seasons to evaluate this aspect (e.g., Karmalkar et al., 2019), we dismiss

this option because this is typically corrected by statistical downscaling techniques (Maraun, 2016). Liang-Liang et al. (2021) used this approach to assess the GCMs ability to simulate the mean intra-annual cycle of precipitation in Asia, identifying that the orography influence considerably the performance of GCMs. Similar results are obtained for continental Chile for extreme temperatures, with better agreement between the mean annual cycle of the GCMs and observations in the coastal zone when compared against the mountain range zone (Figure 9a). Additionally, our results show that GCMs yield a better simulation of mean annual cycle of extreme temperatures in comparison with the mean annual cycle of precipitation (Figure S*11*), in agreement with Lynch et al. (2016) who found that CMIP5 model ensemble simulates the annual cycle quite well in Northeast United States, while an overestimation of the amplitude of the annual cycle was found for precipitation.

The GCM's ability to reproduce historical trends of annual series was tested using (to the best of our knowledge) a novel metric based on the Mann Kendall's statistic. Previous studies diagnosed this property by calibrating a linear regression to the annual timeseries (Brunner et al., 2020; Rupp et al., 2013; Tokarska et al., 2020). However, Rupp et al. (2013) found that metrics evaluating modeled trends with this approach are highly sensitive to the ensemble member. On the other hand, the Mann Kendall's test is usually used for timeseries analysis, with several applications in hydrology (e.g., Cheng et al., 2014; Demaria et al., 2013; Luke et al., 2017) or climate studies (Burger et al., 2018; Valdés-Pineda et al., 2014; Vergara et al., 2020; Wi et al., 2016) and it has been recently used in GCM diagnostics (Shen et al., 2018; Zolghadr-Asli et al., 2019). The advantage of using the Mann-Kendall's statistic over the linear trend is (i) the possibility of analyzing the statistical significance of the trend and (ii) not having to assume that the trend follows a linear behavior. Additionally, the sign of the statistic indicates if the trend is positive or negative, which is useful to contrast modeled and observed trends.

When using the Mann-Kendall's metric, a larger number of pixels with opposite trends to those observed were identified for precipitation in comparison with extreme temperatures, indicating a discrepancy when modelling precipitation. This uncertainty is also greater spatially: while all the models perform poorly in representing the extreme temperatures trends in coastal zones of continental Chile, there is no clear patterns in the errors observed for precipitation.

Extreme temperature observations show a decreasing trend in most of the Coast on Continental Chile, which is consistent with previous studies (Burger et al. 2018; Falvey & Garreaud 2009). However, all the GCMs considered in this study show a positive trend. This systematic error identified could be associated with a poor representation of regional processes (e.g., an inadequate discretization of the continental-oceanic edge) which differs with the coarse horizontal resolution of the GCMs (Maraun, 2016).

6.2. Ranking sensitivity to horizontal resolution

The different spatial resolutions of the native GCMs, and the need of a systematic approach to test these models, makes the choice of diagnostic resolution important, although it generally remains unexplored. Indeed, previous studies diagnosing the historical performance of GCMs usually considered only one spatial resolution (Brunner et al., 2020; Rupp et al., 2013), and the impacts associated to this decision isn't quantified.

This is important for a country like Chile, whose territory presents a considerable latitudinal and longitudinal variations for both temperature and precipitation (Aceituno et al., 2021; Garreaud, 2009). To quantify the impacts of this decision, the first step of the proposed methodology is to aggregate/interpolate GCMs outputs and observations to four mid-scale resolutions both observational dataset and the modeled fields to be diagnosed with the aim of diagnose the models at different resolutions. Although differences are found in the rankings for different resolutions, an agreement for the best (i.e., first five) and worst (i.e., last five) models is observed (Figure 4). A lower level of agreement can be found for models with a medium ranking position, indicating that the methodology is able to identify the best and worst models regardless of the horizontal resolution.

6.3. Effects of statistical downscaling on GCM rankings

The effect of downscaling process on the ranking was assessed through the calculation of four metrics, computed before and after SDMs. The results show no considerable differences between the rankings obtained after applying QDM or MBCn methods (Figure 5a). These may be due the similarities of both methods - MBCn is generalization of QDM for N variables (Cannon, 2018) .

The results presented here show that the downscaling process strongly affects the ranking. The model with the best position on the overall ranking prior to downscaling are relegated to 8th position after downscaling, regardless of the SDM used (Figure 5a). This is mainly because the downscaling process improves the simulation of the intra-annual cycle (r_{cycle}) and the climatological mean - what is reflected in the $r_{spatial}$ metric - considerably: this improvement produces that models without an adequate performance in these aspects now do, matching the performance of the models with better positions in the ranking prior to downscaling (which also improves in these aspects). This is identified by obtaining values greater than 0.99 and 0.90 for $r_{spatial}$ (Figure 6c) and r_{cycle} (Figure 9b) for all the models after the downscaling process. Because of this, the superiority that was clear in the pre-downscaling ranking is less clear in the post-downscaling ranking. Additionally, it should be noted that applying QDM or MBCn doesn't correct errors in modeled trends neither for precipitation (last row in Figure 5b) nor for extreme temperatures (last rows in Figure S3 and Figure S4), indicating that this aspect should be considered in a process of selection of GCMs because is a structural aspect that is not corrected by statistical downscaling methods.

6.4. Limitations of the approach and opportunities

The metrics proposed here were considered with the aim to (i) test different aspects of GCMs results, (ii) avoid selecting models that adequately simulate only one aspect or variable, but fail to reproduce others (Brunner et al., 2020), and (iii) focused on diagnose GCMs variables at the timescale for which the models were created, i.e., from seasonal to climatological scale (Flato et al., 2013; IPCC, 2021; Karmalkar et al., 2019; Rupp et al., 2013). However, only four metrics were used to conduct the diagnosis, and the opportunity of using more than one metric per aspect is proposed for future research. Additional to spatial correlation, the variability or reproduction of amounts could be diagnosed trough the calculation of novel metrics like Spatial KGE (Nashwan & Shahid, 2020), Spatial Efficiency Metric (Demirel et al., 2018) or Cramer's V (Ahmed et al., 2019).

It should be noted that there are aspects that we did not evaluate in this article, such as extreme events or regional teleconnections. Therefore, the inclusion of metrics focused on diagnose these aspects, or the application of metrics like the proposed by the Expert Team on Climate Change Detection and Indices (ETCCDI) is proposed for future research. One aspect not considered in our study is the weight associated with each metric and variable to build the overall ranking per model: the same weight was assigned to each metric-variable combination, but this weight could vary depending on the final use given to the selected GCMs.

Vaittinada Ayar et al. (2016) classified the statistical downscaling methods in four categories; (i) based on transfer functions (TF: Stoner et al., 2013; Thrasher et al., 2012); (ii) stochastics weather generators (SWG: Wilks, 2010; Zorita & Von Storch, 1999); (iii) weather typing methods (WT: Hidalgo et al., 2008; Maurer et al., 2010; Pierce et al., 2014); and (iv) model outputs statistics (MOS), where QDM and MBCn can be found as examples. The SDMs applied in this study correspond to the same classification which is, perhaps, the reason of the similarity obtained for the ranking after downscaling. Therefore, it is proposed for future work to investigate the effect of statistical downscaling methods on GCM rankings considering methods associated to other classifications, with the aim to identify what aspects they influence (spatial, variability, trends, or others) and how this affects the ranking compared to the one made before downscaling. An example of this type of analysis was presented by Gutmann et al. (2014), who identified that the performance for a given diagnostic metric depends on the category of the statistical method.

Finally, the resolutions used in this study range from 0.25° to 1°, which allows the identification of mid-scales climatological patterns. Considering that, in general, the observational datasets have a finer resolution, an opportunity for future work is to diagnose the GCMs at finer resolutions enabling the inclusion of local patterns.

7. Conclusion

In this study, we diagnose 28 CMIP6 GCMs (27 models and the multimodel average) for continental Chile based on their historical performance over the period 1979-2014. The methodology focuses on diagnosing four aspects of the modeled time series of precipitation and extreme temperatures: (i) the interannual variability, (ii) the spatial correlation of the climatological mean, (iii) the temporal reproduction of the mean annual cycle and (iv) the annual trend at climatological scale. This set of metrics combines the criteria used in previous studies for selecting GCMs and it is focused on testing the structural aspects that are not necessarily corrected by statistical downscaling methods. Our main conclusions are:

- None of the GCMs considered is superior in all the variables/metrics simultaneously, before or after the statistical downscaling. Those models with a poor performance in all the metrics should be filtered before downscaling process.
- The horizontal resolution used for model diagnostics does not affect the resulting best and worst GCMs, i.e., the best and worst models retain their position independently if the metrics are calculated at 0.25°, 0.50°, 0.75° or 1.00° horizontal resolutions.

- The multimodel average (calculated from the average of all GCMs diagnosed) showed a medium-low performance before and after the statistical downscaling process because of (i) a systematic underestimation of the interannual variability and (ii) an inadequate reproduction of the spatial patterns, for both precipitation and extreme temperatures.
- The statistical downscaling methods used here (Quantile Delta Mapping and N-dimensional Multivariate Bias Correction) showed similar results for the metrics proposed, improving the climatological annual cycle and spatial correlation. This generates a considerable variation of the overall ranking when the results of raw models are compared with the downscaled outputs. However, none of the methods fix interannual variability or the simulated trends that are contrary to the observed.
- For continental Chile, the modeled annual timeseries showed opposite trends to those observed in up to 21.4%, 15% and 16.2% of the territory for precipitation, maximum and minimum temperatures, respectively. In the case of extreme temperatures, opposite trends are mainly observed in the coastal zone and could be associated with a poor representation of regional processes (e.g., an inadequate discretization of the continental-oceanic edge) which differs with the coarse horizontal resolution of the GCMs.

For future studies, we recommend the selection of GCMs on the diagnostic of various aspects of the climatology, with the aim of avoid selecting models which adequately performance only one aspect but fails to reproduce others. Additionally, it should be considered that latter a downscaling process will be applied, which will surely modify the performance of the models and therefore will influence the ranking of raw GCMs. Because of this, a two-step selection is proposed for future studies: (1) a first selection focused on filtering according to performance in aspects that are inherent to the model (such as trends or interannual variability); and (2) a second filter of models based on their performance after downscaling.

Appendix

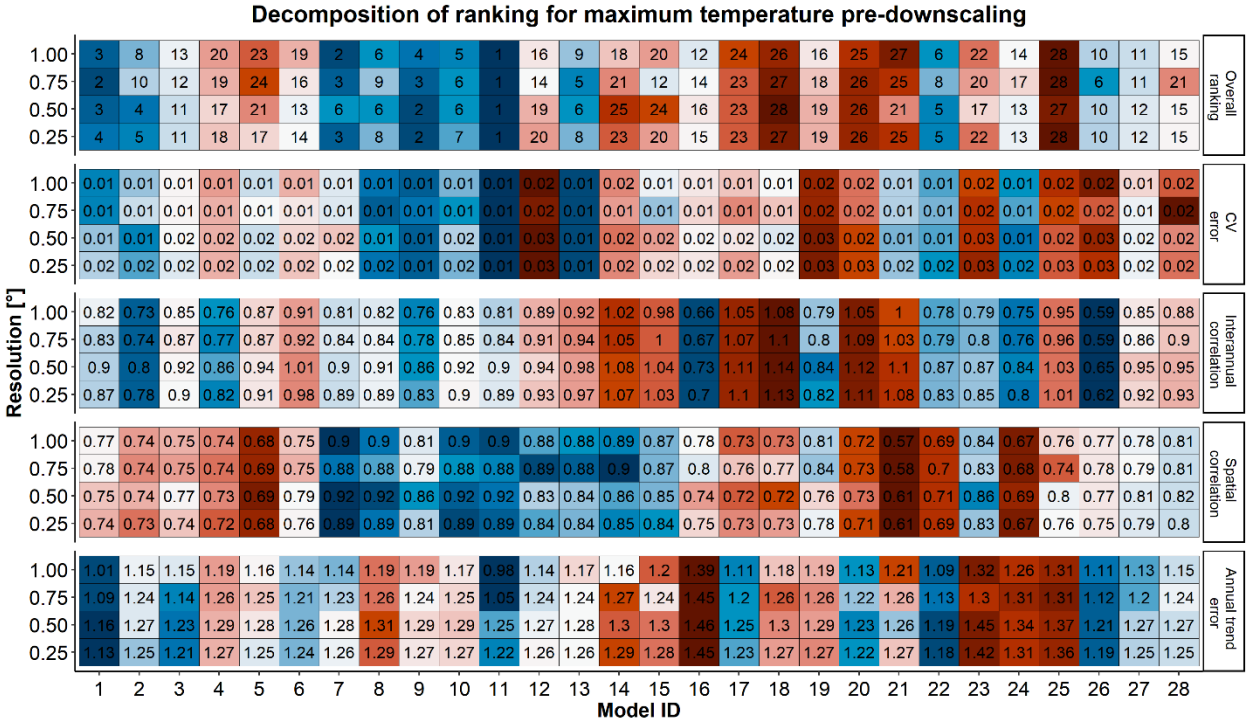

Figure S1: Overall ranking (first row) and ranking per metric (last four rows) calculated for maximum temperature, prior to downscaling process. Colors and numbers in the first row (Overall ranking) indicate the ranking of each GCM. For CV error, Interannual correlation and Annual trend error rows, numbers indicate the area between the CDF (Figure 3b) obtained per model, while the numbers in Spatial correlation row indicate the Pearson's correlation between GCMs and observations.

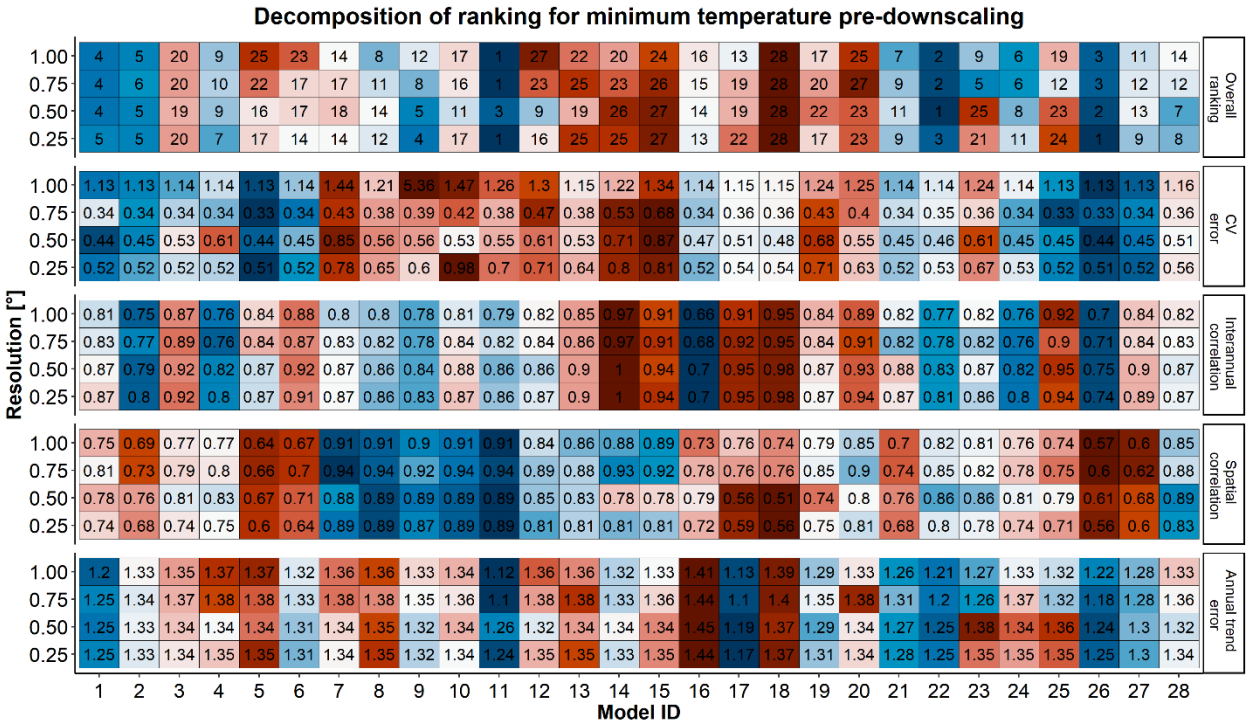

Figure S2: Overall ranking (first row) and ranking per metric (last four rows) calculated for minimum temperature, prior to downscaling process. Colors and numbers in the first row (Overall ranking) indicate the ranking of each GCM. For CV error, Interannual correlation and Annual trend error rows, numbers indicate the area between the CDF (Figure 3b) obtained per model, while the numbers in Spatial correlation row indicate the Pearson's correlation between GCMs and observations.

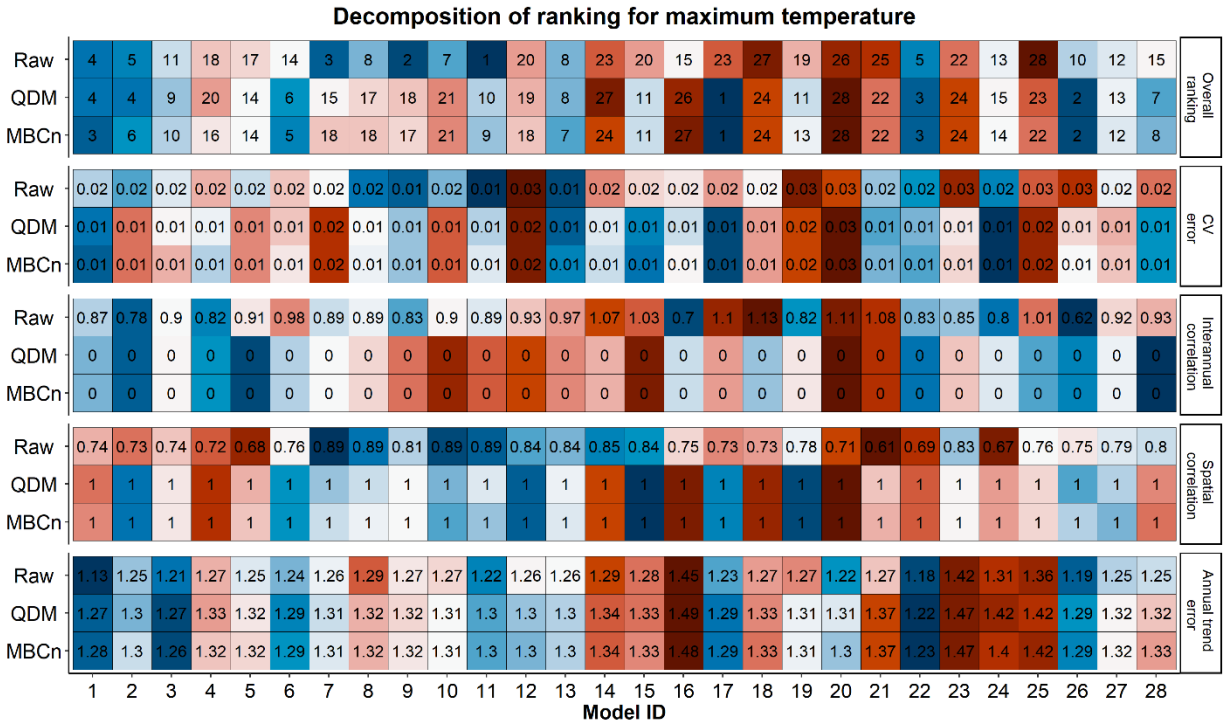

Figure S3: Overall ranking (first row) and ranking per metric (last four rows) calculated for maximum temperature, prior (Raw) and posterior to downscaling process (QDM and MBCn). Colors and numbers in the first row (Overall ranking) indicate the ranking of each GCM. For CV error, Interannual correlation and Annual trend error rows, numbers indicate the area between the CDF (Figure 3b) obtained per model, while the numbers in Spatial correlation row indicate the Pearson's correlation between GCMs and observations.

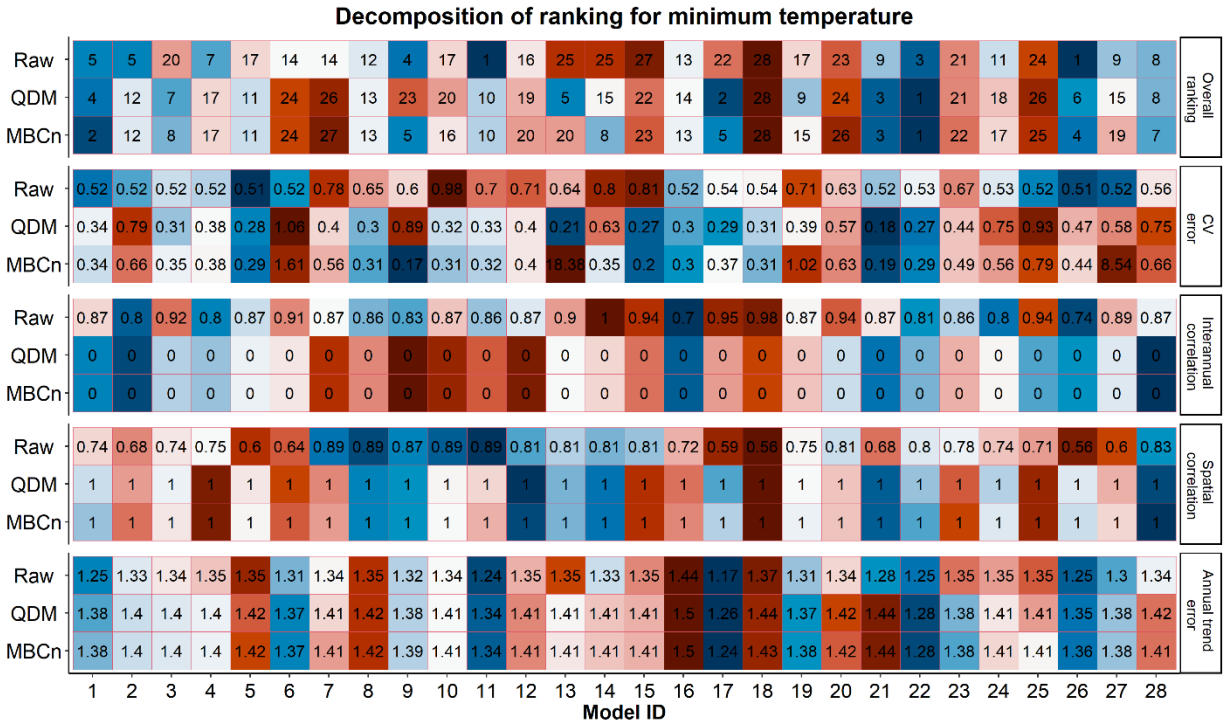

Figure S4: Overall ranking (first row) and ranking per metric (last four rows) calculated for minimum temperature, prior (Raw) and posterior to downscaling process (QDM and MBCn). Colors and numbers in the first row (Overall ranking) indicate the ranking of each GCM. For CV error, Interannual correlation and Annual trend error rows, numbers indicate the area between the CDF (Figure 3b) obtained per model, while the numbers in Spatial correlation row indicate the Pearson's correlation between GCMs and observations.

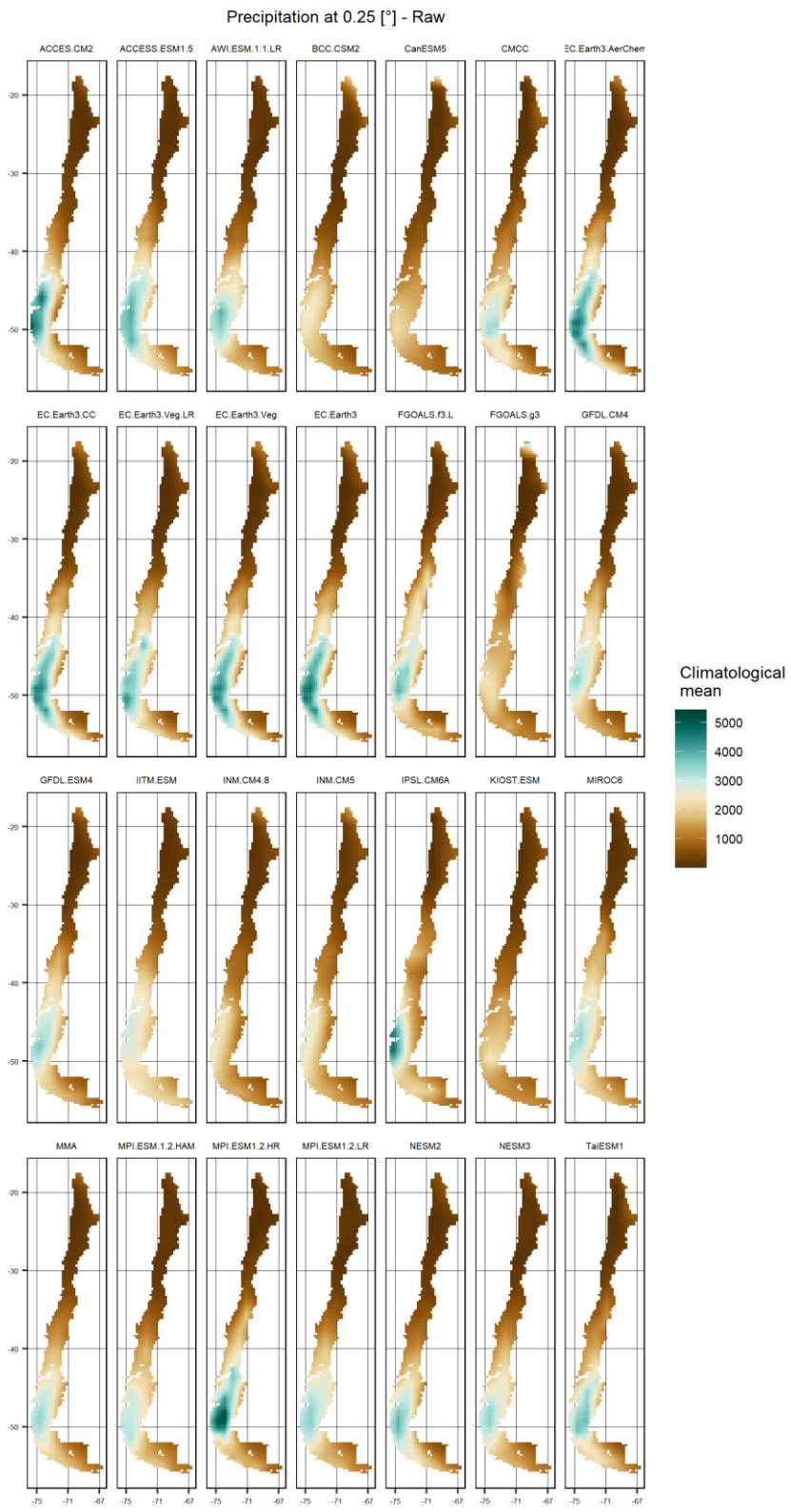

Figure S5: Spatial distribution of the climatological mean of precipitation (mm/yr) simulated by the 28 raw GCMs diagnosed.

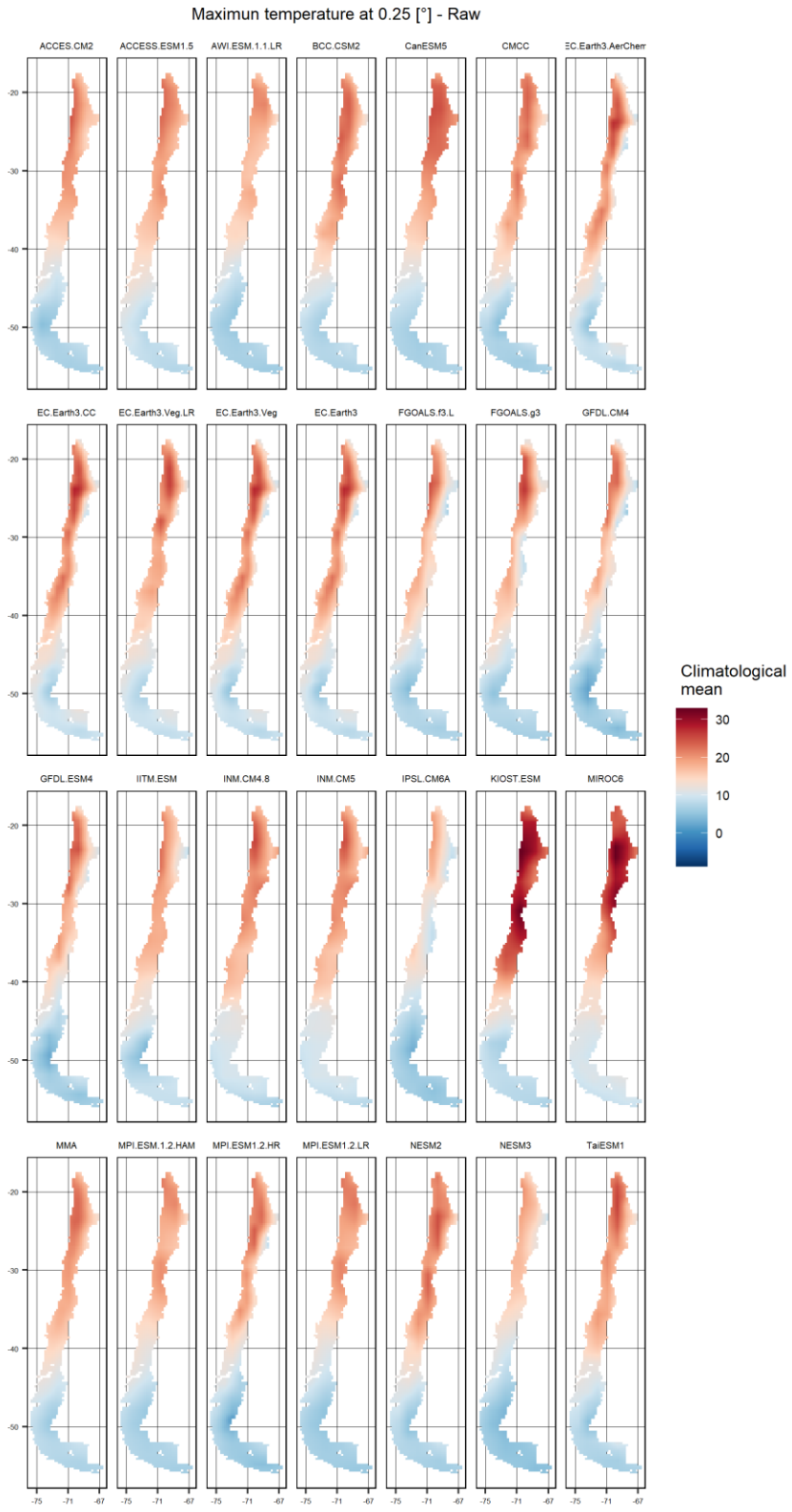

Figure S6: Spatial distribution of the climatological mean of maximum temperature (°C) simulated by the 28 raw GCMs diagnosed.

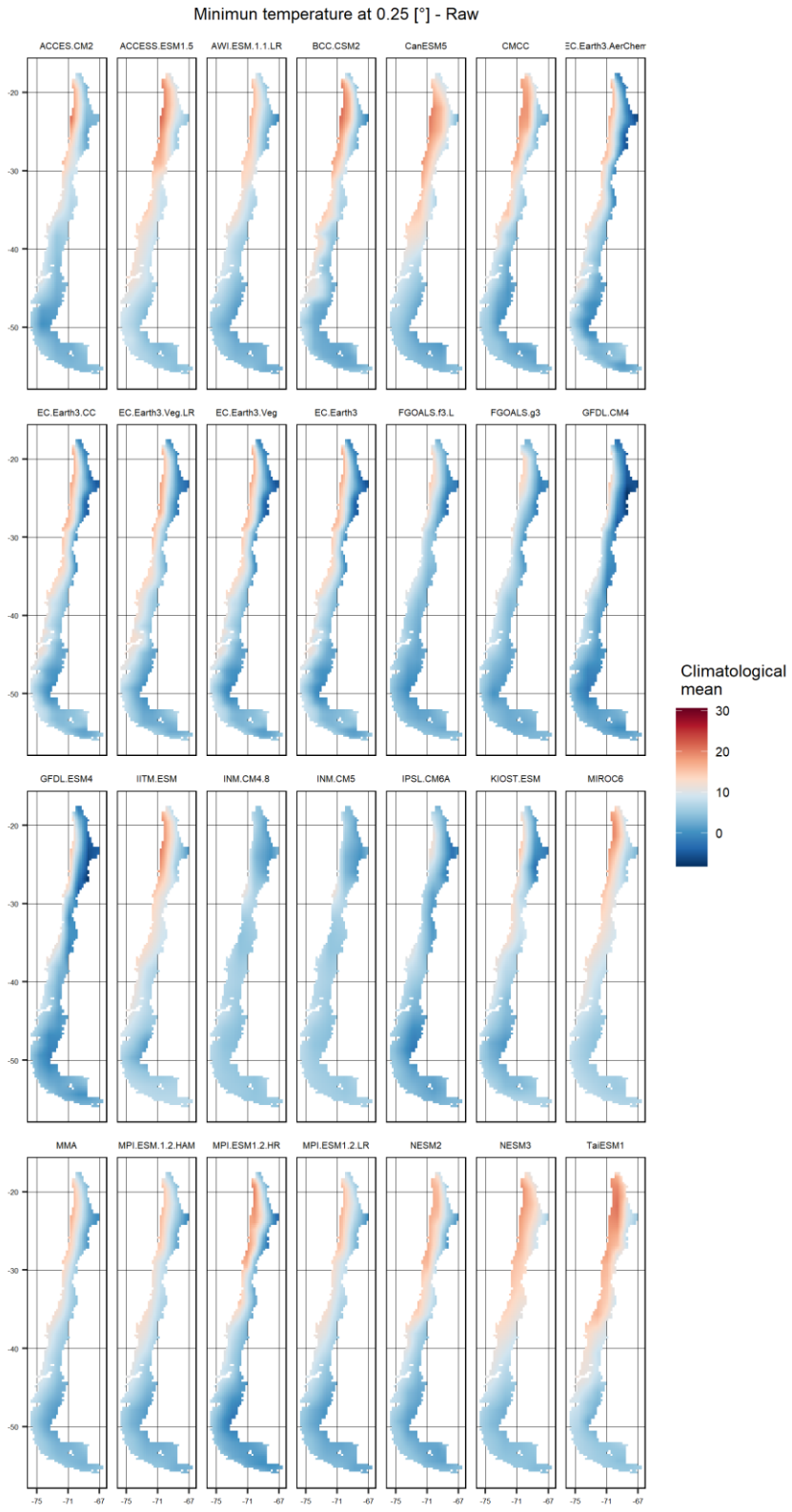

Figure S7: Spatial distribution of the climatological mean of minimum temperature (°C) simulated by the 28 raw GCMs diagnosed.

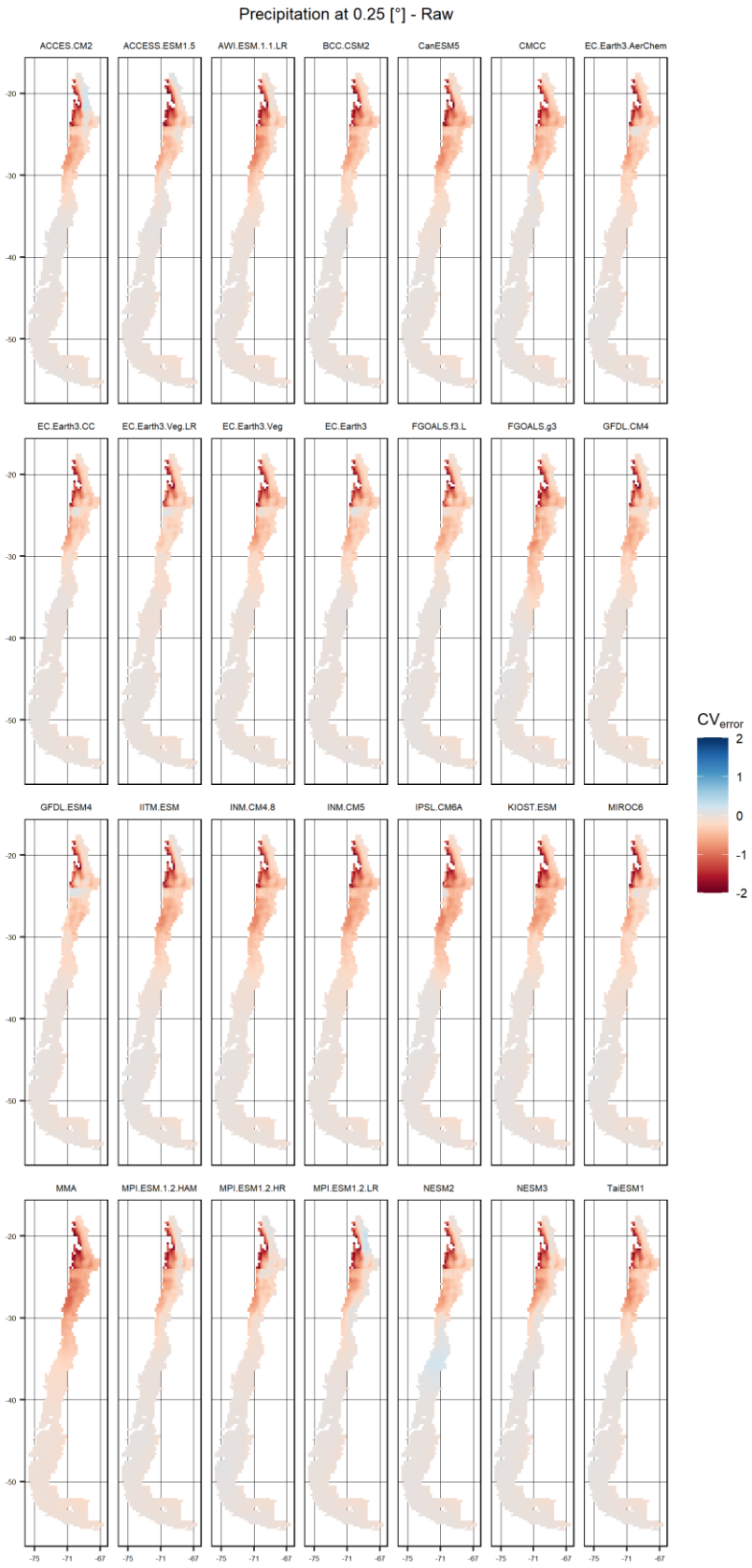

Figure S8: Spatial distribution of the CV error (-) calculated for precipitation simulated by the 28 raw GCMs diagnosed.

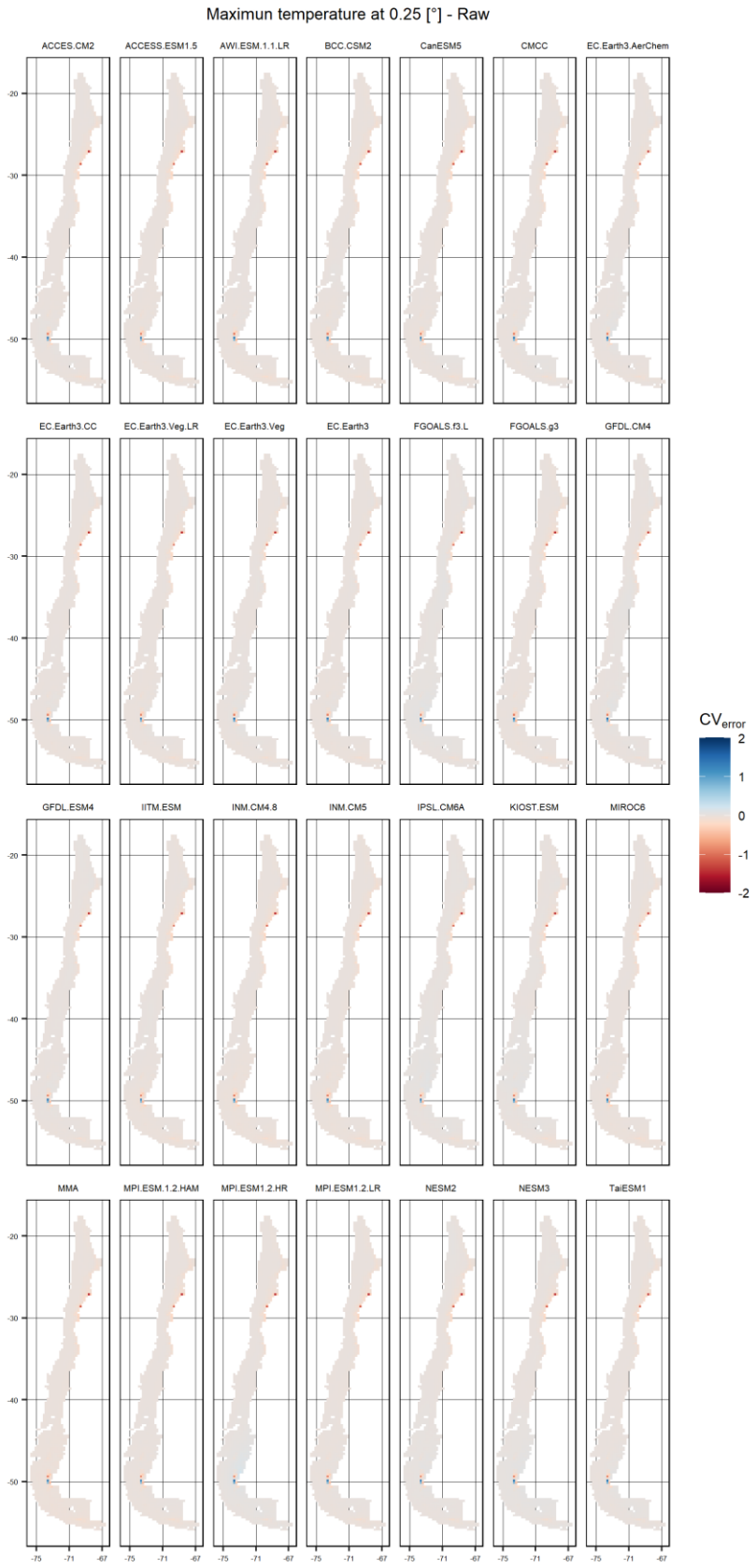

Figure S9: Spatial distribution of the CV error (-) calculated for maximum temperatures simulated by the 28 raw GCMs diagnosed.

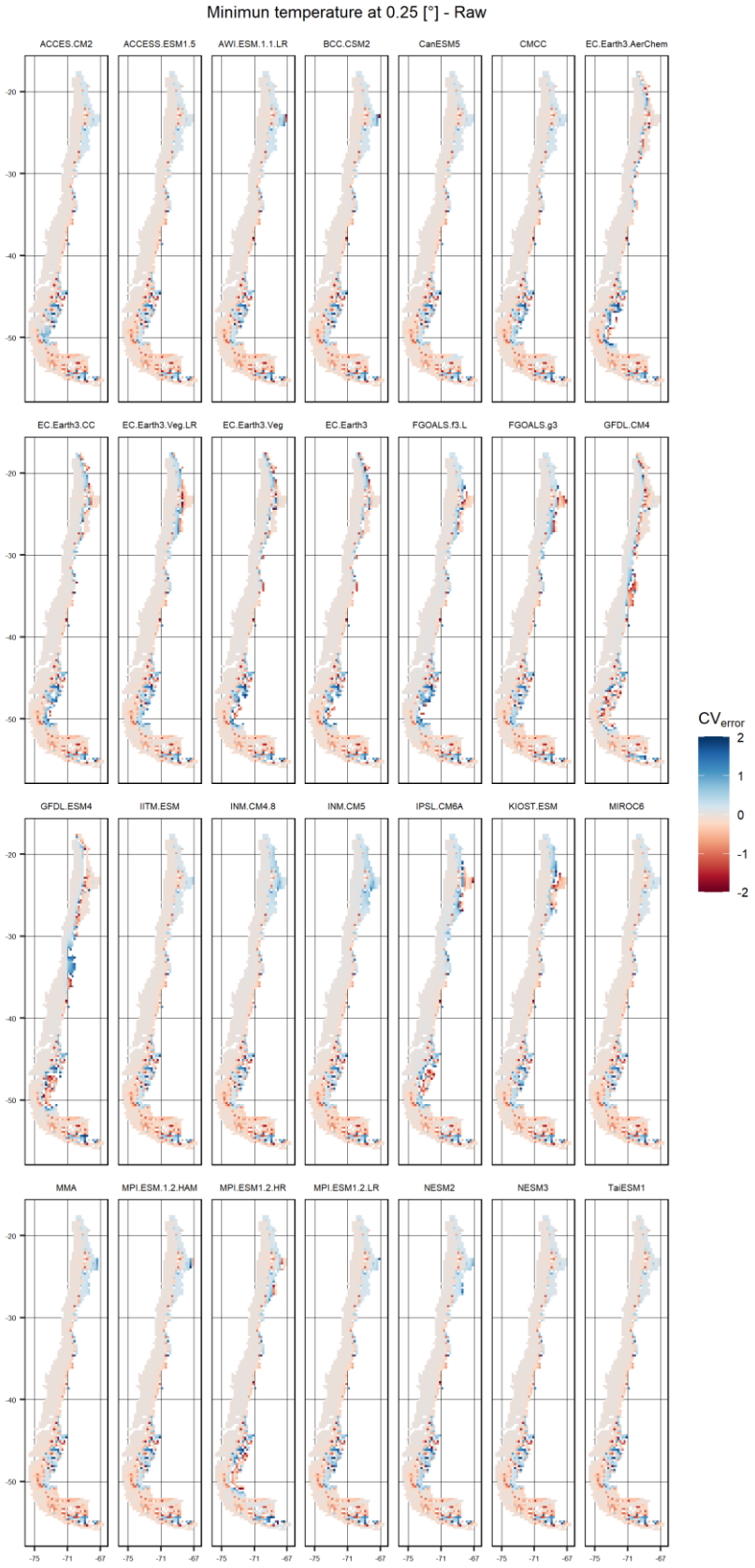

Figure S10: Spatial distribution of the CV error (-) calculated for minimum temperatures simulated by the 28 raw GCMs diagnosed.

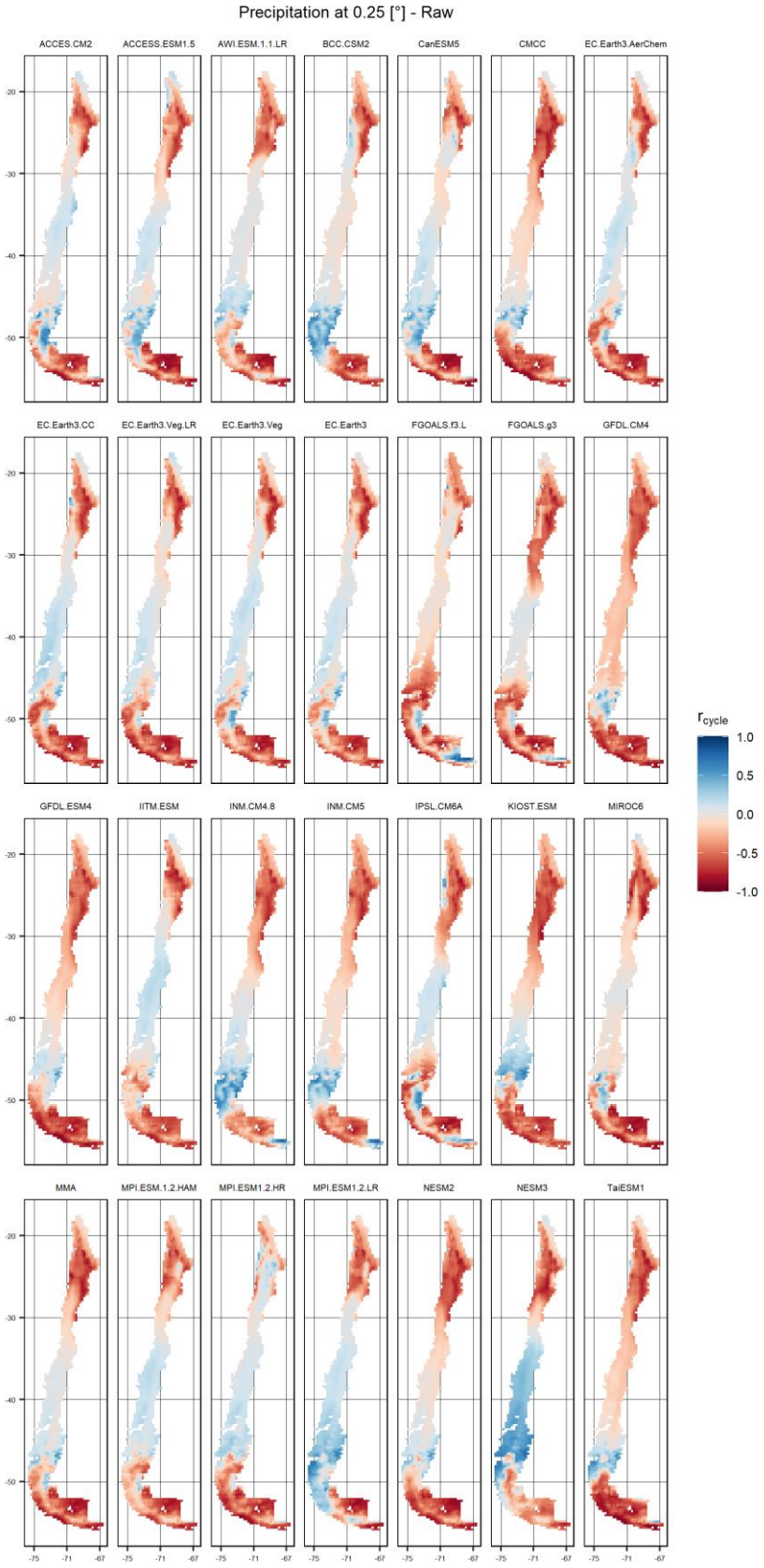

Figure S11: Spatial distribution of the interannual correlation (-) calculated for precipitation simulated by the 28 raw GCMs diagnosed.

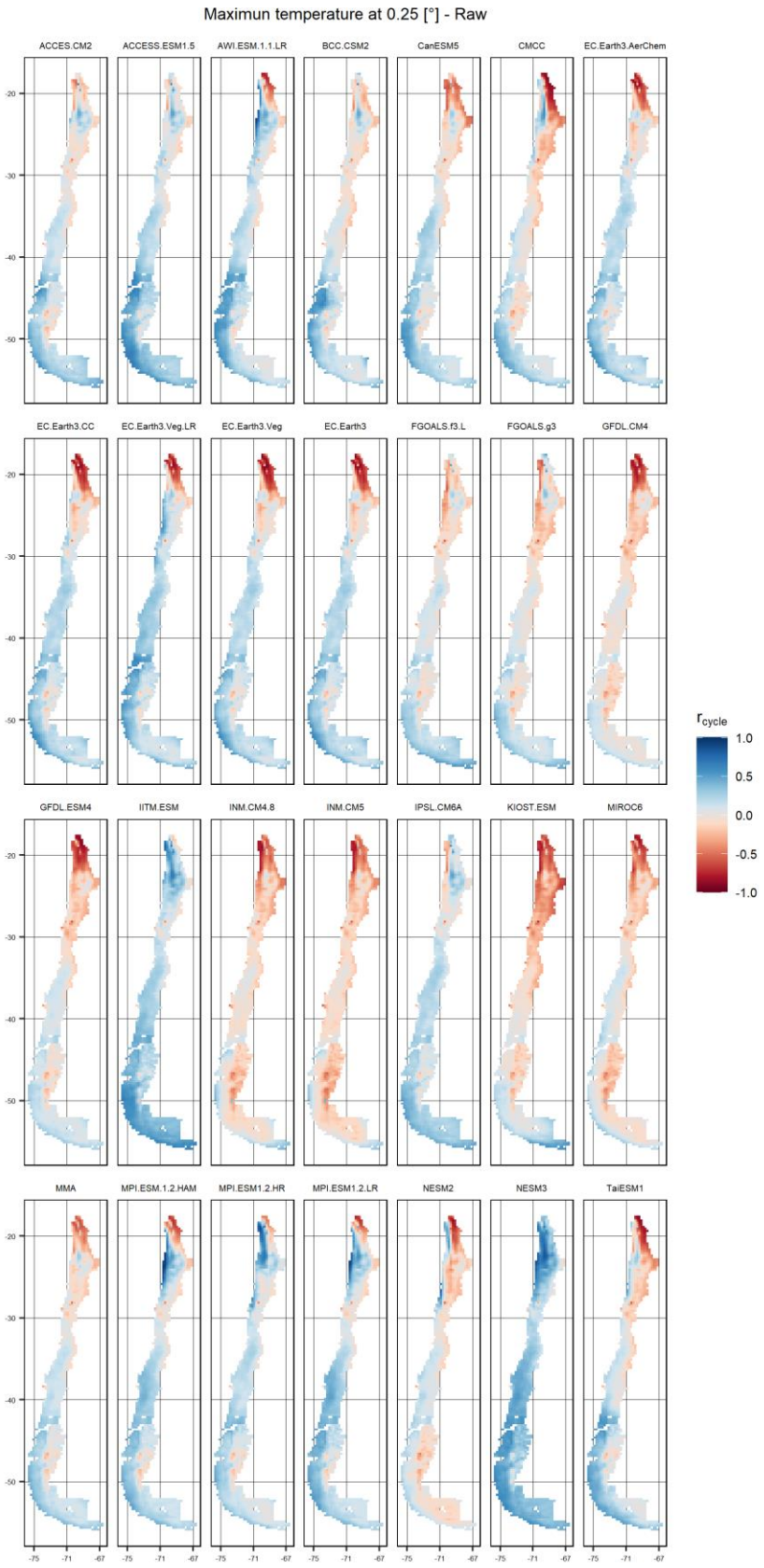

Figure S12: Spatial distribution of the interannual correlation (-) calculated for maximum temperatures simulated by the 28 raw GCMs diagnosed.

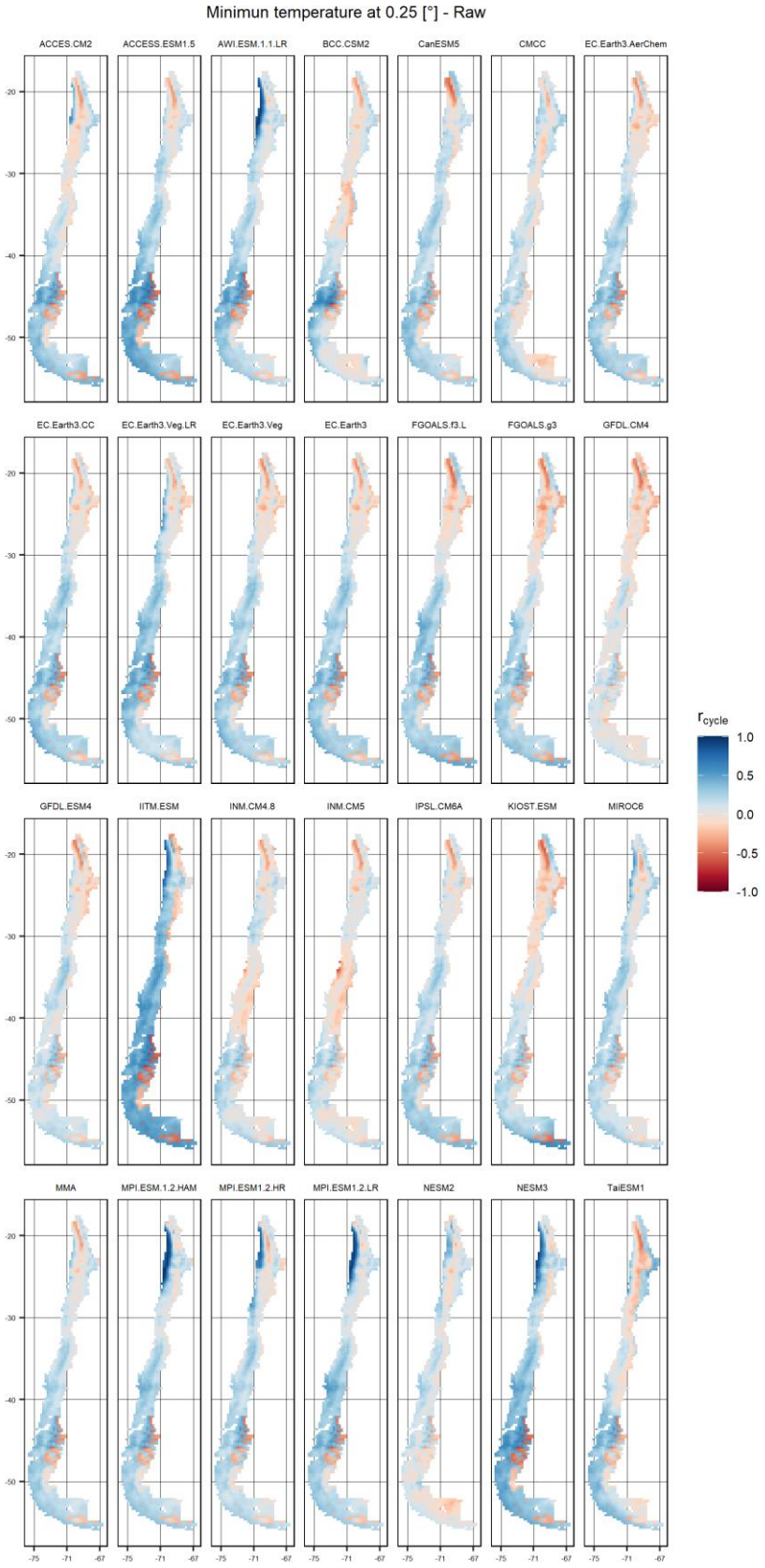

Figure S13: Spatial distribution of the interannual correlation (-) calculated for minimum temperatures simulated by the 28 raw GCMs diagnosed.

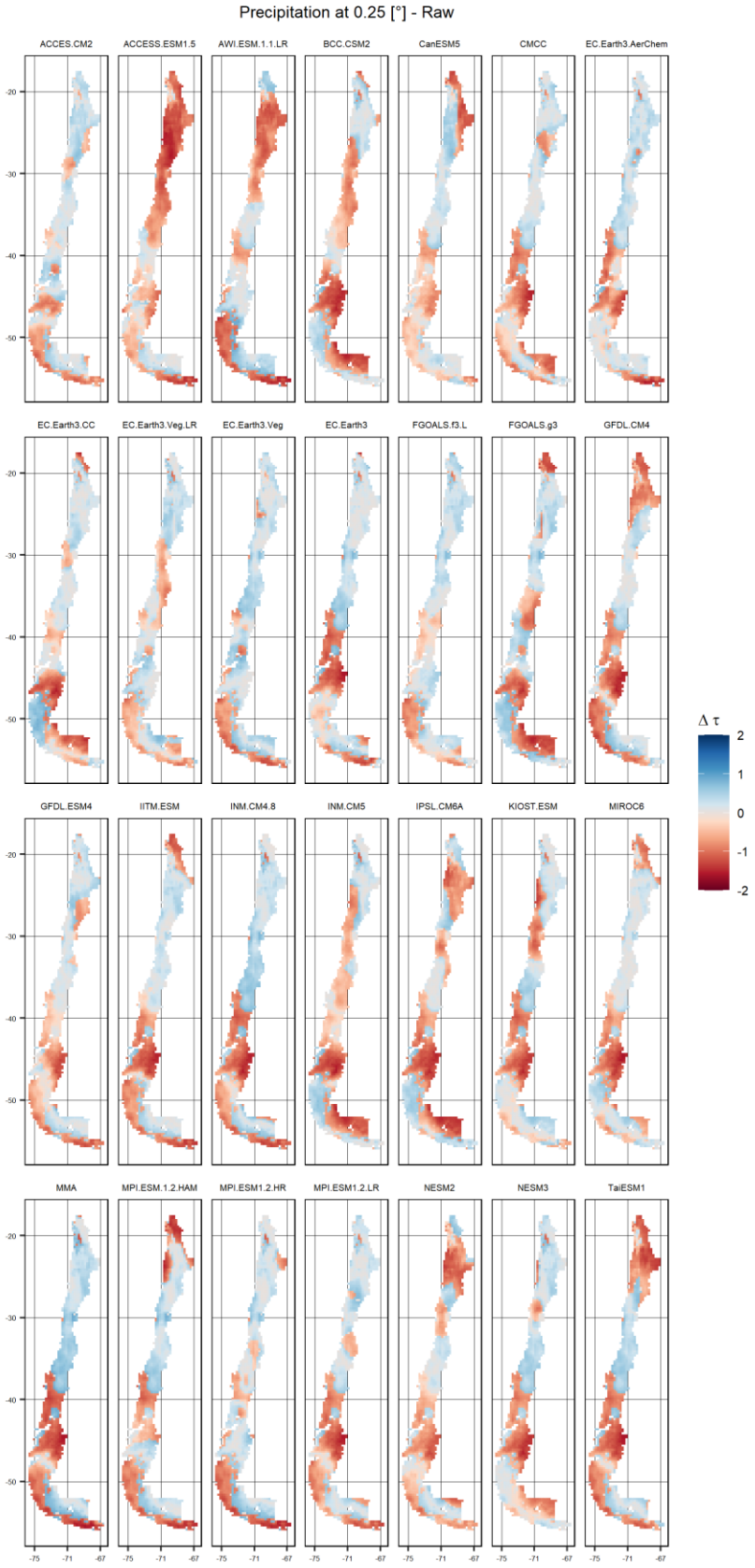

Figure S14: Spatial distribution of the annual trend error (-) calculated for precipitation simulated by the 28 raw GCMs diagnosed.

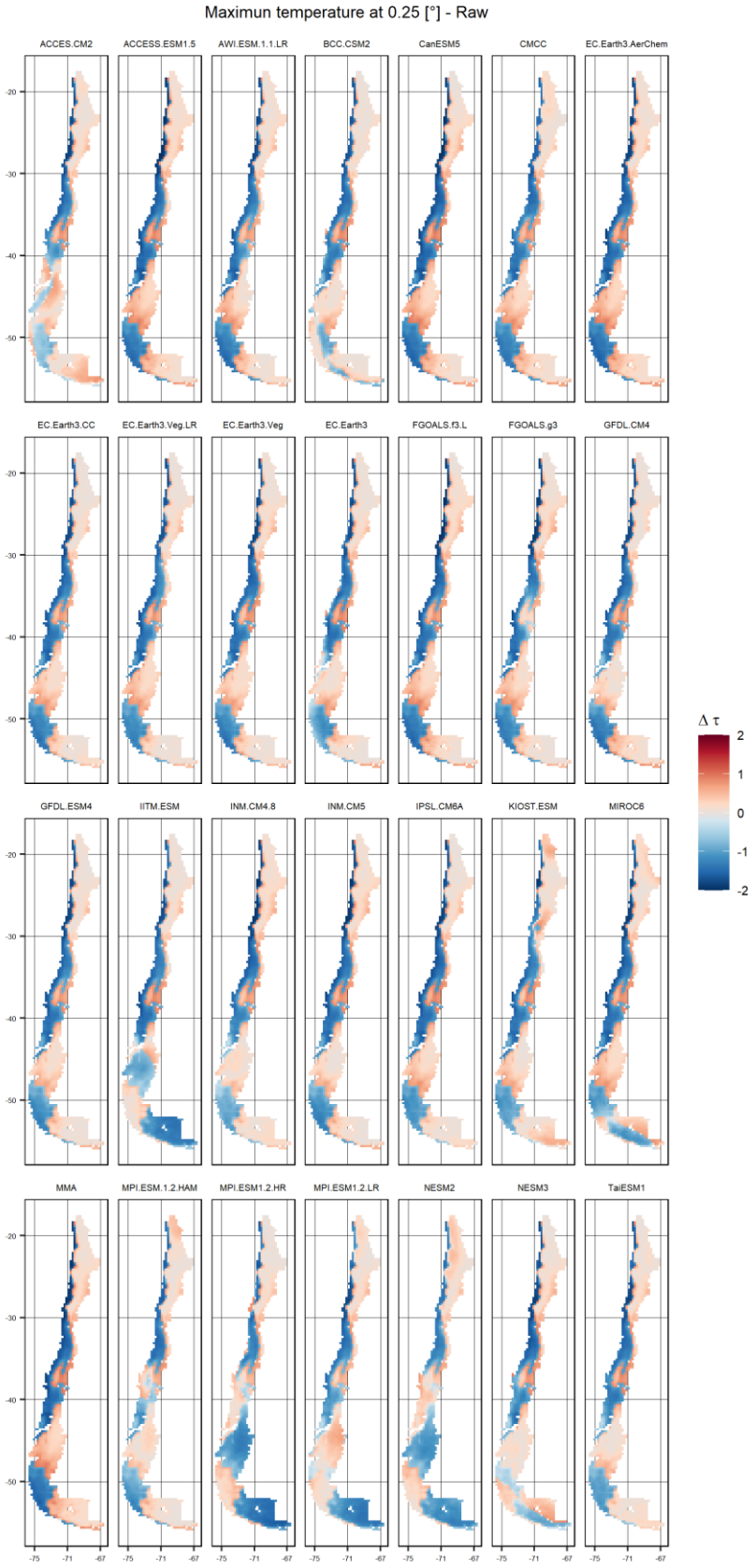

Figure S15: Spatial distribution of the annual trend error (-) calculated for maximum temperatures simulated by the 28 raw GCMs diagnosed.

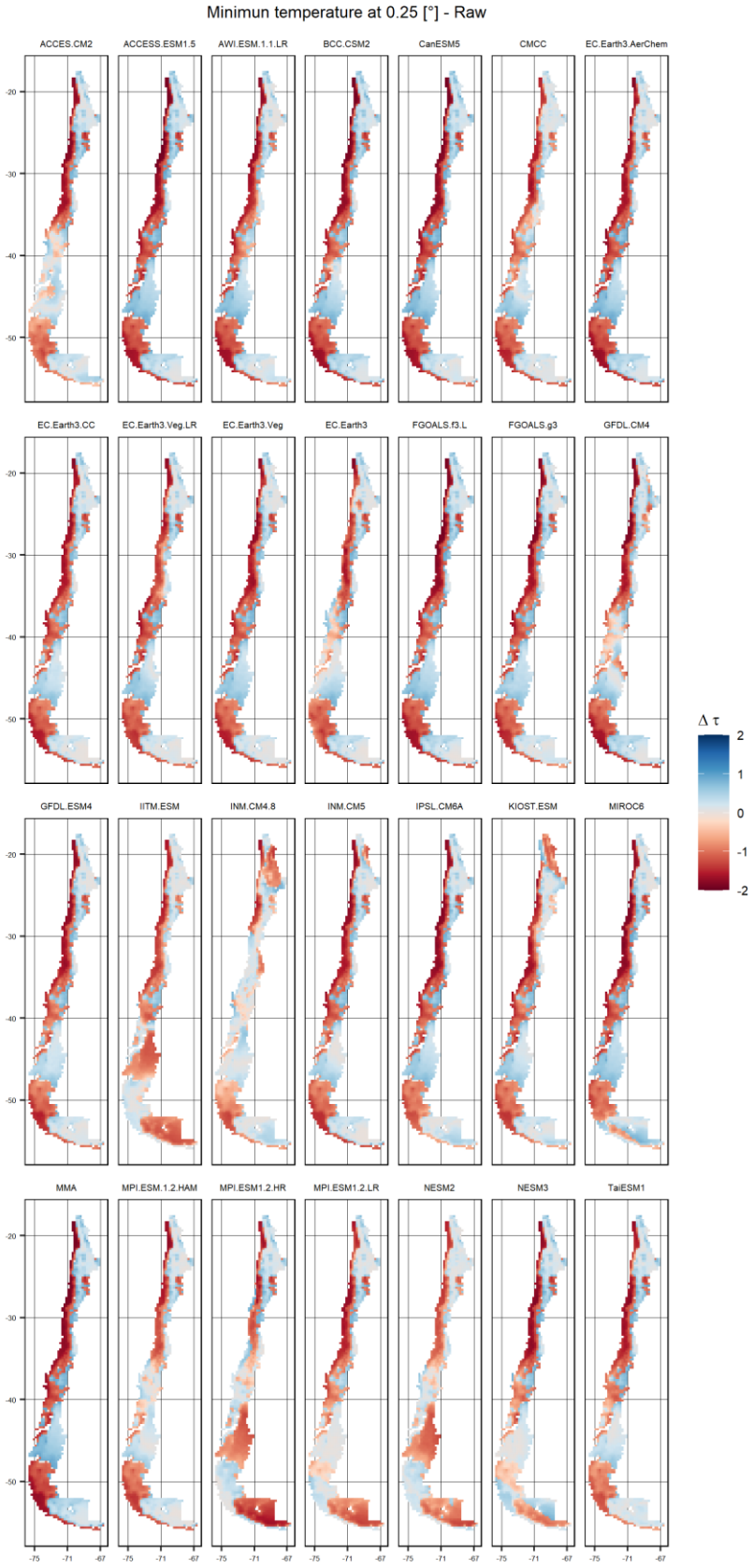

Figure S16: Spatial distribution of the annual trend (-) error calculated for minimum temperatures simulated by the 28 raw GCMs diagnosed.

References

Aceituno, P., Boisier, J. P., Garreaud, R., Rondanelli, R., & Rutllant, J. A. (2021). Climate and Weather in Chile, 7–29. https://doi.org/10.1007/978-3-030-56901-3_2

Ahmed, K., Sachindra, D. A., Shahid, S., Demirel, M. C., & Chung, E. S. (2019). Selection of multi-model ensemble of general circulation models for the simulation of precipitation and maximum and minimum temperature based on spatial assessment metrics. *Hydrology and Earth System Sciences*, *23*(11), 4803–4824. https://doi.org/10.5194/hess-23-4803-2019

Alder, J. R., & Hostetler, S. W. (2019). The Dependence of Hydroclimate Projections in Snow-Dominated Regions of the Western United States on the Choice of Statistically Downscaled Climate Data. *Water Resources Research*, *55*(3), 2279–2300. https://doi.org/10.1029/2018WR023458

Alvarez-Garreton, C., Mendoza, P. A., Pablo Boisier, J., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., et al. (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

Araya-Osses, D., Casanueva, A., Román-Figueroa, C., Uribe, J. M., & Paneque, M. (2020). Climate change projections of temperature and precipitation in Chile based on statistical downscaling. *Climate Dynamics*, *54*(9–10), 4309–4330. https://doi.org/10.1007/s00382-020-05231-4

Blázquez, J., & Nuñez, M. N. (2013). Analysis of uncertainties in future climate projections for South America: Comparison of WCRP-CMIP3 and WCRP-CMIP5 models. *Climate Dynamics*, *41*(3–4), 1039–1056. https://doi.org/10.1007/s00382-012-1489-7

Boisier, J. P., Alvarez-Garretón, C., Cepeda, J., Osses, A., Vásquez, N., & Rondanelli, R. (2018). CR2MET: A high-resolution precipitation and temperature dataset for hydroclimatic research in Chile. *Eguga*, *20*(Vic), 19739. Retrieved from https://ui.adsabs.harvard.edu/abs/2018EGUGA..2019739B/abstract

Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., & Knutti, R. (2020). Reduced global warming from CMIP6 projections when weighting models by performance and independence. *Earth System Dynamics*, *11*(4), 995–1012. https://doi.org/10.5194/esd-11-995-2020

Burger, F., Brock, B., & Montecinos, A. (2018). Seasonal and elevational contrasts in temperature trends in Central Chile between 1979 and 2015. *Global and Planetary Change*, *162*(January), 136–147. https://doi.org/10.1016/j.gloplacha.2018.01.005

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

Cannon, A. J. (2020). Package “MBC,” *29*(19), 7045–7064. https://doi.org/10.1175/JCLI-D-15-0679.1

Cannon, A. J., Sobie, S. R., & Murdock, T. Q. (2015). Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate*, *28*(17), 6938–6959. https://doi.org/10.1175/JCLI-D-14-00754.1

Carrillo, G., Troch, P. A., Sivapalan, M., Wagener, T., Harman, C., & Sawicz, K. (2011). Catchment classification: Hydrological analysis of catchment behavior through process-based modeling along a climate gradient. *Hydrology and Earth System Sciences*, *15*(11),

3411–3430. https://doi.org/10.5194/hess-15-3411-2011

Chegwidden, O. S., Nijssen, B., Rupp, D. E., Arnold, J. R., Clark, M. P., Hamman, J. J., et al. (2019). How Do Modeling Decisions Affect the Spread Among Hydrologic Climate Change Projections? Exploring a Large Ensemble of Simulations Across a Diversity of Hydroclimates. *Earth's Future*, *7*(6), 623–637. https://doi.org/10.1029/2018EF001047

Chen, J., Brissette, F. P., & Lucas-Picher, P. (2015). Assessing the limits of bias-correcting climate model outputs for climate change impact studies. *Journal of Geophysical Research: Atmospheres*, *120*(3), 1123–1136. https://doi.org/10.1002/2014JD022635

Cheng, L., AghaKouchak, A., Gilleland, E., & Katz, R. W. (2014). Non-stationary extreme value analysis in a changing climate. *Climatic Change*, *127*(2), 353–369. https://doi.org/10.1007/s10584-014-1254-5

Chiew, F. H. S., Teng, J., Vaze, J., & Kirono, D. G. C. (2009). Influence of global climate model selection on runoff impact assessment. *Journal of Hydrology*, *379*(1–2), 172–180. https://doi.org/10.1016/j.jhydrol.2009.10.004

Clark, M. P., Slater, A. G., Barrett, A. P., Hay, L. E., McCabe, G. J., Rajagopalan, B., & Leavesley, G. H. (2006). Assimilation of snow covered area information into hydrologic and land-surface models. *Advances in Water Resources*, *29*(8), 1209–1221. https://doi.org/10.1016/j.advwatres.2005.10.001

Clark, M. P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, *47*(9), 1–16. https://doi.org/10.1029/2010WR009827

Clark, M. P., Wilby, R. L., Gutmann, E. D., Vano, J. A., Gangopadhyay, S., Wood, A. W., et al. (2016). Characterizing Uncertainty of the Hydrologic Impacts of Climate Change. *Current Climate Change Reports*, *2*(2), 55–64. https://doi.org/10.1007/s40641-016-0034-x

Demaria, E. M. C., Maurer, E. P., Thrasher, B., Vicuña, S., & Meza, F. J. (2013). Climate change impacts on an alpine watershed in Chile: Do new model projections change the story? *Journal of Hydrology*, *502*, 128–138. https://doi.org/10.1016/j.jhydrol.2013.08.027

Dembélé, M., Hrachowitz, M., Savenije, H. H. G., Mariéthoz, G., & Schaefli, B. (2020). Improving the Predictive Skill of a Distributed Hydrological Model by Calibration on Spatial Patterns With Multiple Satellite Data Sets. *Water Resources Research*, *56*(1), 1–26. https://doi.org/10.1029/2019WR026085

Demirel, M. C., Mai, J., Mendiguren, G., Koch, J., Samaniego, L., & Stisen, S. (2018). Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model. *Hydrology and Earth System Sciences*, *22*(2), 1299–1315. https://doi.org/10.5194/hess-22-1299-2018

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

Espinoza, J. C., Garreaud, R., Poveda, G., Arias, P. A., & Crespo, P. J. (2020). Hydroclimate of the Andes Part I : Main Climatic Features, *8*(March), 1–20. https://doi.org/10.3389/feart.2020.00064

Falvey, M., & Garreaud, R. D. (2009). Regional cooling in a warming world: Recent temperature trends in the southeast Pacific and along the west coast of subtropical South America (1979-2006). *Journal of Geophysical Research Atmospheres*, *114*(4), 1–16.

https://doi.org/10.1029/2008JD010519

Flato, G., Marotzke, J., Abiodun, B., Chou, S. C., Collins, W., Cox, P., et al. (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

Garreaud, R. (2009). The Andes Climate and Wheater. *Advances in Geosciences*, 1–9.

Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. *Journal of Geophysical Research Atmospheres*, *113*(6), 1–20. https://doi.org/10.1029/2007JD008972

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

Hakala, K., Addor, N., & Seibert, J. (2018). Hydrological modeling to evaluate climate model simulations and their bias correction. *Journal of Hydrometeorology*, *19*(8), 1321–1337. https://doi.org/10.1175/JHM-D-17-0189.1

Hidalgo, H. G., D., D. M., & Cayan, D. R. (2008). Downscaling with constructed analogues: Daily precipitation and temperature fields over the United States. *California Climate Change*, (January), 48. Retrieved from http://tenaya.ucsd.edu/wawona-m/downscaled/supporting_materials/CEC-500-2007-123.pdf

IPCC. (2021). Assessment Report 6 Climate Change 2021: The Physical Science Basis. Retrieved from https://www.ipcc.ch/report/ar6/wg1/

Kang, I. S., Yang, Y. M., & Tao, W. K. (2015). GCMs with implicit and explicit representation of cloud microphysics for simulation of extreme precipitation frequency. *Climate Dynamics*, *45*(1–2), 325–335. https://doi.org/10.1007/s00382-014-2376-1

Karmalkar, A. V., Thibeault, J. M., Bryan, A. M., & Seth, A. (2019). Identifying credible and diverse GCMs for regional climate change studies—case study: Northeastern United States. *Climatic Change*, *154*(3–4), 367–386. https://doi.org/10.1007/s10584-019-02411-y

Knoben, W. J. M., Freer, J. E., Peel, M. C., Fowler, K. J. A., & Woods, R. A. (2020). A Brief Analysis of Conceptual Model Structure Uncertainty Using 36 Models and 559 Catchments. *Water Resources Research*, *56*(9), 1–23. https://doi.org/10.1029/2019WR025975

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., & Brunner, L. (2020). Partitioning climate projection uncertainty with multiple large ensembles and CMIP5 / 6, 491–508.

Liang-Liang, L., Jian, L., & Ru-Cong, Y. (2021). Evaluation of CMIP6 HighResMIP models in simulating precipitation over Central Asia. *Advances in Climate Change Research*, (xxxx). https://doi.org/10.1016/j.accre.2021.09.009

Liang, Y., Gillett, N. P., & Monahan, A. H. (2020). Climate Model Projections of 21st Century Global Warming Constrained Using the Observed Warming Trend. *Geophysical Research Letters*, *47*(12), 1–10. https://doi.org/10.1029/2019GL086757

Liu, Z., Wang, Y., Xu, Z., & Duan, Q. (2017). Conceptual Hydrological Models. *Handbook of Hydrometeorological Ensemble Forecasting*, 1–23. https://doi.org/10.1007/978-3-642-40457-3_22-1

López-Moreno, J. I., Pomeroy, J. W., Revuelto, J., & Vicente-Serrano, S. M. (2013). Response of snow processes to climate change: Spatial variability in a small basin in the Spanish

Pyrenees. *Hydrological Processes*, *27*(18), 2637–2650. https://doi.org/10.1002/hyp.9408

Luke, A., Vrugt, J. A., AghaKouchak, A., Matthew, R., & Sanders, B. F. (2017). Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the United States. *Water Resources Research*, *53*(7), 5469–5494. https://doi.org/10.1002/2016WR019676

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

Lynch, C., Seth, A., & Thibeault, J. (2016). Recent and projected annual cycles of temperature and precipitation in the Northeast United States from CMIP5. *Journal of Climate*, *29*(1), 347–365. https://doi.org/10.1175/JCLI-D-14-00781.1

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

McCreight, J. L., & Small, E. E. (2014). Modeling bulk density and snow water equivalent using daily snow depth observations. *Cryosphere*, *8*(2), 521–536. https://doi.org/10.5194/tc-8-521-2014

Mendoza, P. A., Clark, M. P., Mizukami, N., Newman, A. J., Barlage, M., Gutmann, E. D., et al. (2015). Effects of hydrologic model choice and calibration on the portrayal of climate change impacts. *Journal of Hydrometeorology*, *16*(2), 762–780. https://doi.org/10.1175/JHM-D-14-0104.1

Montecinos, A., & Aceituno, P. (2003). Seasonality of the ENSO-related rainfall variability in central Chile and associated circulation anomalies. *Journal of Climate*, *16*(2), 281–296. https://doi.org/10.1175/1520-0442(2003)016<0281:SOTERR>2.0.CO;2

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

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

Pierce, D. W., Cayan, D. R., & Thrasher, B. L. (2014). Statistical downscaling using localized constructed analogs (LOCA). *Journal of Hydrometeorology*, *15*(6), 2558–2585. https://doi.org/10.1175/JHM-D-14-0082.1

Pohlert, T. (2020). Package ‘ trend .’

Poveda, G., Espinoza, J. C., Zuluaga, M. D., Solman, S. A., Garreaud, R., & van Oevelen, P. J. (2020). High Impact Weather Events in the Andes. *Frontiers in Earth Science*, *8*(May), 1–32. https://doi.org/10.3389/feart.2020.00162

Randall, D. A., Wood, R. A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., et al. (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

Rubio-Álvarez, E., & McPhee, J. (2010). Patterns of spatial and temporal variability in streamflow records in south central Chile in the period 1952-2003. *Water Resources Research*, *46*(5), 1–16. https://doi.org/10.1029/2009WR007982

Rupp, D. E., Abatzoglou, J. T., Hegewisch, K. C., & Mote, P. W. (2013). Evaluation of CMIP5 20th century climate simulations for the Pacific Northwest USA. *Journal of Geophysical Research Atmospheres*, *118*(19), 10,884-10,906. https://doi.org/10.1002/jgrd.50843

Sa'adi, Z., Shiru, M. S., Shahid, S., & Ismail, T. (2020). Selection of general circulation models for the projections of spatio-temporal changes in temperature of Borneo Island based on CMIP5. *Theoretical and Applied Climatology*, *139*(1–2), 351–371. https://doi.org/10.1007/s00704-019-02948-z

Sagarika, S., Kalra, A., & Ahmad, S. (2015). Interconnections between oceanic-atmospheric indices and variability in the U.S. streamflow. *Journal of Hydrology*, *525*, 724–736. https://doi.org/10.1016/j.jhydrol.2015.04.020

Shen, M., Chen, J., Zhuan, M., Chen, H., Xu, C. Y., & Xiong, L. (2018). Estimating uncertainty and its temporal variation related to global climate models in quantifying climate change impacts on hydrology. *Journal of Hydrology*, *556*, 10–24. https://doi.org/10.1016/j.jhydrol.2017.11.004

Stoner, A. M. K., Hayhoe, K., Yang, X., & Wuebbles, D. J. (2013). An asynchronous regional regression model for statistical downscaling of daily climate variables. *International Journal of Climatology*, *33*(11), 2473–2494. https://doi.org/10.1002/joc.3603

Thrasher, B., Maurer, E. P., McKellar, C., & Duffy, P. B. (2012). Technical Note: Bias correcting climate model simulated daily temperature extremes with quantile mapping. *Hydrology and Earth System Sciences*, *16*(9), 3309–3314. https://doi.org/10.5194/hess-16-3309-2012

Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., & Knutti, R. (2020). Past warming trend constrains future warming in CMIP6 models. *Science Advances*, *6*(12), 1–14. https://doi.org/10.1126/sciadv.aaz9549

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

Uribe, J. M., Cabrera, R., de la Fuente, A., & Paneque, M. (2012). Atlas Bioclimáti co de Chile Atlas Bioclimático de Chile.

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

Valdés-Pineda, R., Pizarro, R., García-Chevesich, P., Valdés, J. B., Olivares, C., Vera, M., et al. (2014). Water governance in Chile: Availability, management and climate change. *Journal of Hydrology*, *519*(PC), 2538–2567. https://doi.org/10.1016/j.jhydrol.2014.04.016

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

Vergara, I., M. Moreiras, S., Araneo, Di., & Garreaud, R. (2020). Geo-climatic hazards in the eastern subtropical Andes: Distribution, climate drivers and trends. *Natural Hazards and Earth System Sciences*, *20*(5), 1353–1367. https://doi.org/10.5194/nhess-20-1353-2020

Viale, M., Bianchi, E., Cara, L., Ruiz, L. E., Villalba, R., Pitte, P., et al. (2019). Contrasting climates at both sides of the Andes in Argentina and Chile. *Frontiers in Environmental Science*, *7*(May), 1–15. https://doi.org/10.3389/fenvs.2019.00069

Wang, H. M., Chen, J., Xu, C. Y., Chen, H., Guo, S., Xie, P., & Li, X. (2019). Does the weighting of climate simulations result in a better quantification of hydrological impacts? *Hydrology and Earth System Sciences*, *23*(10), 4033–4050. https://doi.org/10.5194/hess-23-4033-2019

Wang, H. M., Chen, J., Xu, C. Y., Zhang, J., & Chen, H. (2020). A Framework to Quantify the Uncertainty Contribution of GCMs Over Multiple Sources in Hydrological Impacts of Climate Change. *Earth's Future*, *8*(8). https://doi.org/10.1029/2020EF001602

Webb, M. J., Andrews, T., Bodas-Salcedo, A., Bony, S., Bretherton, C. S., Chadwick, R., et al. (2017). The Cloud Feedback Model Intercomparison Project (CFMIP) contribution to CMIP6. *Geoscientific Model Development*, *10*(1), 359–384. https://doi.org/10.5194/gmd-10-359-2017

Wi, S., Valdés, J. B., Steinschneider, S., & Kim, T. W. (2016). Non-stationary frequency analysis of extreme precipitation in South Korea using peaks-over-threshold and annual maxima. *Stochastic Environmental Research and Risk Assessment*, *30*(2), 583–606. https://doi.org/10.1007/s00477-015-1180-8

Wilks, D. S. (2010). Use of stochastic weather generators for precipitation downscaling. *Wiley Interdisciplinary Reviews: Climate Change*, *1*(6), 898–907. https://doi.org/10.1002/wcc.85

Wood, A. W., Leung, L. R., Sridhar, V., & Lettenmaier, D. P. (2004). Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change*, *62*(1–3), 189–216. https://doi.org/10.1023/B:CLIM.0000013685.99609.9e

Zhang, R., Cheng, L., Liu, P., Huang, K., Gong, Y., Qin, S., & Liu, D. (2021). Effect of GCM credibility on water resource system robustness under climate change based on decision scaling. *Advances in Water Resources*, *158*(September), 104063. https://doi.org/10.1016/j.advwatres.2021.104063

Zolghadr-Asli, B., Bozorg-Haddad, O., Sarzaeim, P., & Chu, X. (2019). Investigating the variability of gcms' simulations using time series analysis. *Journal of Water and Climate Change*, *10*(3), 449–463. https://doi.org/10.2166/wcc.2018.099

Zorita, E., & Von Storch, H. (1999). The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. *Journal of Climate*, *12*(8 PART 2), 2474–2489. https://doi.org/10.1175/1520-0442(1999)012<2474:tamaas>2.0.co;2

III. Conclusiones

En el presente trabajo se diagnosticó el desempeño histórico de 28 modelos (27 modelos más uno generado a partir del promedio de todos, llamado MMA) asociados a la 6ta Fase del Proyecto de Intercomparación de Modelos Acoplados (CMIP6). El diagnostico fue aplicado a todo Chile continental para el periodo 1979-2014, utilizando el producto grillado CR2METv2.0 como set de datos observacional. La evaluación consideró cuatro aspectos de las series modeladas de precipitaciones y temperaturas extremas: (i) la variabilidad interanual; (ii) la correlación espacial de la media climatológica; (iii) la reproducción temporal del ciclo medio anual; y (iv) la tendencia anual a escala climatológica. Para ello, se construyeron cuatro métricas con el objetivo de aunar los criterios utilizados en estudios previos de selección de GCMs, las cuales se enfocan en evaluar los aspectos estructurales de los GCM que no son necesariamente corregidos por métodos estadísticos de escalamiento (i.e., *downscaling*). Dichas métricas fueron calculadas para cada variable-modelo, y se compararon con los valores *observados*, obteniendo un *ranking* de desempeño para combinación de modelo, métrica y variable.

Adicionalmente, se analizó el efecto de la resolución horizontal a la que se diagnostican los modelos. Para ello, se repitió el proceso de diagnóstico a cuatro resoluciones horizontales distintas (0.25°, 0.50°, 0.75° y 1.00°) y analizó la variación del *ranking* de modelos.

Finalmente, se estudió la influencia que tiene el *downscaling* en el rendimiento de los modelos al aplicarse dos métodos de escalamiento estadístico a los 28 GCMs diagnosticados (Quantile Delta Mapping y N-dimensional Multivariate Bias Correction) para luego recalcular el *ranking* de modelos y compararlo con el obtenido previo al proceso de escalamiento. Las principales conclusiones del estudio son las siguientes:

- **Ningún modelo demuestra una superioridad en todas las variables y métricas a la vez**, ya sea antes o después del escalamiento estadístico. Debido a esto, se deben filtrar aquellos modelos que presenten un mal desempeño en todas las métricas, asegurándose de seleccionar modelos que, por separado, tengan un buen desempeño en cada uno de los aspectos evaluados.
- **No se identifican variaciones considerables en el *ranking* de modelos – previo al escalamiento – producto de la variación de resoluciones de diagnóstico**, ya que los mejores y peores modelos conservan su posición independiente si son evaluados a 0.25°, 0.50°, 0.75° o 1.00°. Estos resultados permiten deducir que la incertidumbre asociada a la resolución de diagnóstico es más bien despreciable, por lo que utilizar una resolución de mediana escala (i.e., cercana a 0.50°) reduciría el gasto computacional del trabajo con GCMs, entregando resultados similares a los que se obtendrían a partir del trabajo con modelos a resolución más fina.
- **El MMA presenta un desempeño medio-bajo antes y después del escalamiento estadístico.** Esto se debe a (i) una subestimación sistemática de la variabilidad interanual y (ii) una inadecuada reproducción de los patrones espaciales, tanto de precipitación como de temperaturas extremas.

- **Los dos métodos de escalamiento estadístico aplicados generan una variación considerable de la clasificación general de los GCMs.** Tanto el método Quantile Delta Mapping como N-dimensional Multivariate Bias Correction producen una mejoría del ciclo anual climatológico y la correlación espacial, independiente del GCM al que se apliquen. Esto produce que todos los modelos presenten un buen desempeño en esos aspectos, independiente de la variable analizada, y que, en consecuencia, GCMs que tenían un desempeño bajo o medio adquieran una mejor posición en el *ranking* calculado posteriormente al escalado.
- **Ninguno de los dos métodos de escalamiento estadístico aplicados logra corregir errores en las tendencias históricas.** Para Chile continental, las series temporales anuales modeladas por los GCMs mostraron tendencias contrarias a las observadas en hasta un 21,4%, 15% y 16,2% del territorio para precipitación, temperaturas máximas y mínimas, respectivamente, lo cual se observa antes y después del proceso de escalamiento. En el caso de las temperaturas extremas, las tendencias contrarias se observan principalmente en la zona costera y podrían estar asociadas a una pobre representación de los procesos regionales (por ejemplo, una inadecuada discretización del borde continental-oceánico) que difiere con la resolución horizontal gruesa de la GCM.

Para futuros estudios se recomienda basar la selección de GCMs en el diagnóstico de varios aspectos de la climatología, con el objetivo de evitar seleccionar modelos que reproduzcan adecuadamente solo un aspecto, pero fallen al reproducir otros. Adicionalmente, se debe considerar que al aplicar un proceso de escalamiento estadístico se modificará el desempeño de los modelos, lo que influirá en la clasificación de los GCM. Por ello, para futuros estudios se propone una selección en dos pasos: (1) una primera selección enfocada en filtrar según el desempeño en aspectos que son inherentes al modelo (como tendencias o variabilidad interanual); y (2) un segundo filtro de modelos basado en su rendimiento después del proceso de escalamiento.

IV. Bibliografía

Aceituno, P., Boisier, J. P., Garreaud, R., Rondanelli, R., & Rutllant, J. A. (2021). Climate and Weather in Chile, 7–29. https://doi.org/10.1007/978-3-030-56901-3_2

Ahmed, K., Sachindra, D. A., Shahid, S., Demirel, M. C., & Chung, E. S. (2019). Selection of multi-model ensemble of general circulation models for the simulation of precipitation and maximum and minimum temperature based on spatial assessment metrics. *Hydrology and Earth System Sciences*, *23*(11), 4803–4824. https://doi.org/10.5194/hess-23-4803-2019

Alder, J. R., & Hostetler, S. W. (2019). The Dependence of Hydroclimate Projections in Snow-Dominated Regions of the Western United States on the Choice of Statistically Downscaled Climate Data. *Water Resources Research*, *55*(3), 2279–2300. https://doi.org/10.1029/2018WR023458

Alvarez-Garreton, C., Mendoza, P. A., Pablo Boisier, J., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., et al. (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

Araya-Osses, D., Casanueva, A., Román-Figueroa, C., Uribe, J. M., & Paneque, M. (2020). Climate change projections of temperature and precipitation in Chile based on statistical downscaling. *Climate Dynamics*, *54*(9–10), 4309–4330. https://doi.org/10.1007/s00382-020-05231-4

Blázquez, J., & Nuñez, M. N. (2013). Analysis of uncertainties in future climate projections for South America: Comparison of WCRP-CMIP3 and WCRP-CMIP5 models. *Climate Dynamics*, *41*(3–4), 1039–1056. https://doi.org/10.1007/s00382-012-1489-7

Boisier, J. P., Alvarez-Garretón, C., Cepeda, J., Osses, A., Vásquez, N., & Rondanelli, R. (2018). CR2MET: A high-resolution precipitation and temperature dataset for hydroclimatic research in Chile. *Eguga*, *20*(Vic), 19739. Retrieved from https://ui.adsabs.harvard.edu/abs/2018EGUGA..2019739B/abstract

Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., & Knutti, R. (2020). Reduced global warming from CMIP6 projections when weighting models by performance and independence. *Earth System Dynamics*, *11*(4), 995–1012. https://doi.org/10.5194/esd-11-995-2020

Burger, F., Brock, B., & Montecinos, A. (2018). Seasonal and elevational contrasts in temperature trends in Central Chile between 1979 and 2015. *Global and Planetary Change*, *162*(January), 136–147. https://doi.org/10.1016/j.gloplacha.2018.01.005

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

Cannon, A. J. (2020). Package "MBC," *29*(19), 7045–7064. https://doi.org/10.1175/JCLI-D-15-0679.1

Cannon, A. J., Sobie, S. R., & Murdock, T. Q. (2015). Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate*, *28*(17), 6938–6959. https://doi.org/10.1175/JCLI-D-14-00754.1

Carrillo, G., Troch, P. A., Sivapalan, M., Wagener, T., Harman, C., & Sawicz, K. (2011). Catchment classification: Hydrological analysis of catchment behavior through process-based modeling along a climate gradient. *Hydrology and Earth System Sciences*, *15*(11),

3411–3430. https://doi.org/10.5194/hess-15-3411-2011

Chegwidden, O. S., Nijssen, B., Rupp, D. E., Arnold, J. R., Clark, M. P., Hamman, J. J., et al. (2019). How Do Modeling Decisions Affect the Spread Among Hydrologic Climate Change Projections? Exploring a Large Ensemble of Simulations Across a Diversity of Hydroclimates. *Earth's Future*, *7*(6), 623–637. https://doi.org/10.1029/2018EF001047

Chen, J., Brissette, F. P., & Lucas-Picher, P. (2015). Assessing the limits of bias-correcting climate model outputs for climate change impact studies. *Journal of Geophysical Research: Atmospheres*, *120*(3), 1123–1136. https://doi.org/10.1002/2014JD022635

Cheng, L., AghaKouchak, A., Gilleland, E., & Katz, R. W. (2014). Non-stationary extreme value analysis in a changing climate. *Climatic Change*, *127*(2), 353–369. https://doi.org/10.1007/s10584-014-1254-5

Chiew, F. H. S., Teng, J., Vaze, J., & Kirono, D. G. C. (2009). Influence of global climate model selection on runoff impact assessment. *Journal of Hydrology*, *379*(1–2), 172–180. https://doi.org/10.1016/j.jhydrol.2009.10.004

Clark, M. P., Slater, A. G., Barrett, A. P., Hay, L. E., McCabe, G. J., Rajagopalan, B., & Leavesley, G. H. (2006). Assimilation of snow covered area information into hydrologic and land-surface models. *Advances in Water Resources*, *29*(8), 1209–1221. https://doi.org/10.1016/j.advwatres.2005.10.001

Clark, M. P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, *47*(9), 1–16. https://doi.org/10.1029/2010WR009827

Clark, M. P., Wilby, R. L., Gutmann, E. D., Vano, J. A., Gangopadhyay, S., Wood, A. W., et al. (2016). Characterizing Uncertainty of the Hydrologic Impacts of Climate Change. *Current Climate Change Reports*, *2*(2), 55–64. https://doi.org/10.1007/s40641-016-0034-x

Demaria, E. M. C., Maurer, E. P., Thrasher, B., Vicuña, S., & Meza, F. J. (2013). Climate change impacts on an alpine watershed in Chile: Do new model projections change the story? *Journal of Hydrology*, *502*, 128–138. https://doi.org/10.1016/j.jhydrol.2013.08.027

Dembélé, M., Hrachowitz, M., Savenije, H. H. G., Mariéthoz, G., & Schaefli, B. (2020). Improving the Predictive Skill of a Distributed Hydrological Model by Calibration on Spatial Patterns With Multiple Satellite Data Sets. *Water Resources Research*, *56*(1), 1–26. https://doi.org/10.1029/2019WR026085

Demirel, M. C., Mai, J., Mendiguren, G., Koch, J., Samaniego, L., & Stisen, S. (2018). Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model. *Hydrology and Earth System Sciences*, *22*(2), 1299–1315. https://doi.org/10.5194/hess-22-1299-2018

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

Espinoza, J. C., Garreaud, R., Poveda, G., Arias, P. A., & Crespo, P. J. (2020). Hydroclimate of the Andes Part I : Main Climatic Features, *8*(March), 1–20. https://doi.org/10.3389/feart.2020.00064

Falvey, M., & Garreaud, R. D. (2009). Regional cooling in a warming world: Recent temperature trends in the southeast Pacific and along the west coast of subtropical South America (1979-2006). *Journal of Geophysical Research Atmospheres*, *114*(4), 1–16.

https://doi.org/10.1029/2008JD010519

Flato, G., Marotzke, J., Abiodun, B., Chou, S. C., Collins, W., Cox, P., et al. (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

Garreaud, R. (2009). The Andes Climate and Wheater. *Advances in Geosciences*, 1–9.

Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. *Journal of Geophysical Research Atmospheres*, *113*(6), 1–20. https://doi.org/10.1029/2007JD008972

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

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

Hakala, K., Addor, N., & Seibert, J. (2018). Hydrological modeling to evaluate climate model simulations and their bias correction. *Journal of Hydrometeorology*, *19*(8), 1321–1337. https://doi.org/10.1175/JHM-D-17-0189.1

Hidalgo, H. G., D., D. M., & Cayan, D. R. (2008). Downscaling with constructed analogues: Daily precipitation and temperature fields over the United States. *California Climate Change*, (January), 48. Retrieved from http://tenaya.ucsd.edu/wawona-m/downscaled/supporting_materials/CEC-500-2007-123.pdf

IPCC. (2021). Assessment Report 6 Climate Change 2021: The Physical Science Basis. Retrieved from https://www.ipcc.ch/report/ar6/wg1/

Kang, I. S., Yang, Y. M., & Tao, W. K. (2015). GCMs with implicit and explicit representation of cloud microphysics for simulation of extreme precipitation frequency. *Climate Dynamics*, *45*(1–2), 325–335. https://doi.org/10.1007/s00382-014-2376-1

Karmalkar, A. V., Thibeault, J. M., Bryan, A. M., & Seth, A. (2019). Identifying credible and diverse GCMs for regional climate change studies—case study: Northeastern United States. *Climatic Change*, *154*(3–4), 367–386. https://doi.org/10.1007/s10584-019-02411-y

Kling, H., Fuchs, M., & 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/10.1016/j.jhydrol.2012.01.011

Knoben, W. J. M., Freer, J. E., Peel, M. C., Fowler, K. J. A., & Woods, R. A. (2020). A Brief Analysis of Conceptual Model Structure Uncertainty Using 36 Models and 559 Catchments. *Water Resources Research*, *56*(9), 1–23. https://doi.org/10.1029/2019WR025975

Knutti, R. (2010). The end of model democracy? *Climatic Change*, *102*(3), 395–404. https://doi.org/10.1007/s10584-010-9800-2

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., & Brunner, L. (2020). Partitioning climate projection uncertainty with multiple large ensembles and CMIP5 / 6, 491–508.

Liang-Liang, L., Jian, L., & Ru-Cong, Y. (2021). Evaluation of CMIP6 HighResMIP models in simulating precipitation over Central Asia. *Advances in Climate Change Research*, (xxxx).

https://doi.org/10.1016/j.accre.2021.09.009

Liang, Y., Gillett, N. P., & Monahan, A. H. (2020). Climate Model Projections of 21st Century Global Warming Constrained Using the Observed Warming Trend. *Geophysical Research Letters*, *47*(12), 1–10. https://doi.org/10.1029/2019GL086757

Liu, Z., Wang, Y., Xu, Z., & Duan, Q. (2017). Conceptual Hydrological Models. *Handbook of Hydrometeorological Ensemble Forecasting*, 1–23. https://doi.org/10.1007/978-3-642-40457-3_22-1

López-Moreno, J. I., Pomeroy, J. W., Revuelto, J., & Vicente-Serrano, S. M. (2013). Response of snow processes to climate change: Spatial variability in a small basin in the Spanish Pyrenees. *Hydrological Processes*, *27*(18), 2637–2650. https://doi.org/10.1002/hyp.9408

Luke, A., Vrugt, J. A., AghaKouchak, A., Matthew, R., & Sanders, B. F. (2017). Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the United States. *Water Resources Research*, *53*(7), 5469–5494. https://doi.org/10.1002/2016WR019676

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

Lynch, C., Seth, A., & Thibeault, J. (2016). Recent and projected annual cycles of temperature and precipitation in the Northeast United States from CMIP5. *Journal of Climate*, *29*(1), 347–365. https://doi.org/10.1175/JCLI-D-14-00781.1

Mann, H. B. (1945). Nonparametric Tests Against Trend. *Econometrica*, *13*(3), 245. https://doi.org/10.2307/1907187

Maraun, D. (2013). Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *Journal of Climate*, *26*(6), 2137–2143. https://doi.org/10.1175/JCLI-D-12-00821.1

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

McCreight, J. L., & Small, E. E. (2014). Modeling bulk density and snow water equivalent using daily snow depth observations. *Cryosphere*, *8*(2), 521–536. https://doi.org/10.5194/tc-8-521-2014

Mendoza, P. A., Clark, M. P., Mizukami, N., Newman, A. J., Barlage, M., Gutmann, E. D., et al. (2015). Effects of hydrologic model choice and calibration on the portrayal of climate change impacts. *Journal of Hydrometeorology*, *16*(2), 762–780. https://doi.org/10.1175/JHM-D-14-0104.1

Montecinos, A., & Aceituno, P. (2003). Seasonality of the ENSO-related rainfall variability in central Chile and associated circulation anomalies. *Journal of Climate*, *16*(2), 281–296. https://doi.org/10.1175/1520-0442(2003)016<0281:SOTERR>2.0.CO;2

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

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

Pierce, D. W., Cayan, D. R., & Thrasher, B. L. (2014). Statistical downscaling using localized constructed analogs (LOCA). *Journal of Hydrometeorology*, *15*(6), 2558–2585. https://doi.org/10.1175/JHM-D-14-0082.1

Pohlert, T. (2020). Package ‘ trend .’

Poveda, G., Espinoza, J. C., Zuluaga, M. D., Solman, S. A., Garreaud, R., & van Oevelen, P. J. (2020). High Impact Weather Events in the Andes. *Frontiers in Earth Science*, *8*(May), 1–32. https://doi.org/10.3389/feart.2020.00162

Randall, D. A., Wood, R. A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., et al. (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

Rubio-Álvarez, E., & McPhee, J. (2010). Patterns of spatial and temporal variability in streamflow records in south central Chile in the period 1952-2003. *Water Resources Research*, *46*(5), 1–16. https://doi.org/10.1029/2009WR007982

Rupp, D. E., Abatzoglou, J. T., Hegewisch, K. C., & Mote, P. W. (2013). Evaluation of CMIP5 20th century climate simulations for the Pacific Northwest USA. *Journal of Geophysical Research Atmospheres*, *118*(19), 10,884-10,906. https://doi.org/10.1002/jgrd.50843

Sa’adi, Z., Shiru, M. S., Shahid, S., & Ismail, T. (2020). Selection of general circulation models for the projections of spatio-temporal changes in temperature of Borneo Island based on CMIP5. *Theoretical and Applied Climatology*, *139*(1–2), 351–371. https://doi.org/10.1007/s00704-019-02948-z

Sagarika, S., Kalra, A., & Ahmad, S. (2015). Interconnections between oceanic-atmospheric indices and variability in the U.S. streamflow. *Journal of Hydrology*, *525*, 724–736. https://doi.org/10.1016/j.jhydrol.2015.04.020

Salazar Morey, A. A. (2017). Predicción De Forzantes Meteorológicas a Escala Local En Un Clima No Estacionario . Aplicación a Cuencas De Chile Central, 93. Retrieved from http://repositorio.uchile.cl/bitstream/handle/2250/144470/Determinación-de-una-metodología-para-la-predicción-de-forzantes-meteorológicas-a-escala-local-en-un-clima-no.pdf?sequence=1&isAllowed=y

Shen, M., Chen, J., Zhuan, M., Chen, H., Xu, C. Y., & Xiong, L. (2018). Estimating uncertainty and its temporal variation related to global climate models in quantifying climate change impacts on hydrology. *Journal of Hydrology*, *556*, 10–24. https://doi.org/10.1016/j.jhydrol.2017.11.004

Sota Christie, J., Lagos-Zúñiga, M. Á., Mendoza Zúñiga, P., & Rondanelli Rojas, R. (2022). Proyecciones de eventos hidro-meteorológicos extremos en una cuenca del Altiplano chileno, bajo el escenario de cambio climático SSP5-8.5.

Stoner, A. M. K., Hayhoe, K., Yang, X., & Wuebbles, D. J. (2013). An asynchronous regional regression model for statistical downscaling of daily climate variables. *International Journal of Climatology*, *33*(11), 2473–2494. https://doi.org/10.1002/joc.3603

Thrasher, B., Maurer, E. P., McKellar, C., & Duffy, P. B. (2012). Technical Note: Bias correcting climate model simulated daily temperature extremes with quantile mapping. *Hydrology*

and Earth System Sciences, *16*(9), 3309–3314. https://doi.org/10.5194/hess-16-3309-2012

Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., & Knutti, R. (2020). Past warming trend constrains future warming in CMIP6 models. *Science Advances*, *6*(12), 1–14. https://doi.org/10.1126/sciadv.aaz9549

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

Uribe, J. M., Cabrera, R., de la Fuente, A., & Paneque, M. (2012). Atlas Bioclimáti co de Chile Atlas Bioclimático de Chile.

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

Valdés-Pineda, R., Pizarro, R., García-Chevesich, P., Valdés, J. B., Olivares, C., Vera, M., et al. (2014). Water governance in Chile: Availability, management and climate change. *Journal of Hydrology*, *519*(PC), 2538–2567. https://doi.org/10.1016/j.jhydrol.2014.04.016

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

Vergara, I., M. Moreiras, S., Araneo, Di., & Garreaud, R. (2020). Geo-climatic hazards in the eastern subtropical Andes: Distribution, climate drivers and trends. *Natural Hazards and Earth System Sciences*, *20*(5), 1353–1367. https://doi.org/10.5194/nhess-20-1353-2020

Viale, M., Bianchi, E., Cara, L., Ruiz, L. E., Villalba, R., Pitte, P., et al. (2019). Contrasting climates at both sides of the Andes in Argentina and Chile. *Frontiers in Environmental Science*, *7*(May), 1–15. https://doi.org/10.3389/fenvs.2019.00069

Wang, H. M., Chen, J., Xu, C. Y., Chen, H., Guo, S., Xie, P., & Li, X. (2019). Does the weighting of climate simulations result in a better quantification of hydrological impacts? *Hydrology and Earth System Sciences*, *23*(10), 4033–4050. https://doi.org/10.5194/hess-23-4033-2019

Wang, H. M., Chen, J., Xu, C. Y., Zhang, J., & Chen, H. (2020). A Framework to Quantify the Uncertainty Contribution of GCMs Over Multiple Sources in Hydrological Impacts of Climate Change. *Earth's Future*, *8*(8). https://doi.org/10.1029/2020EF001602

Webb, M. J., Andrews, T., Bodas-Salcedo, A., Bony, S., Bretherton, C. S., Chadwick, R., et al. (2017). The Cloud Feedback Model Intercomparison Project (CFMIP) contribution to CMIP6. *Geoscientific Model Development*, *10*(1), 359–384. https://doi.org/10.5194/gmd-10-359-2017

Wi, S., Valdés, J. B., Steinschneider, S., & Kim, T. W. (2016). Non-stationary frequency analysis of extreme precipitation in South Korea using peaks-over-threshold and annual maxima. *Stochastic Environmental Research and Risk Assessment*, *30*(2), 583–606. https://doi.org/10.1007/s00477-015-1180-8

Wilks, D. S. (2010). Use of stochastic weather generators for precipitation downscaling. *Wiley Interdisciplinary Reviews: Climate Change*, *1*(6), 898–907. https://doi.org/10.1002/wcc.85

Wood, A. W., Leung, L. R., Sridhar, V., & Lettenmaier, D. P. (2004). Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change*, *62*(1–3), 189–216. https://doi.org/10.1023/B:CLIM.0000013685.99609.9e

Zhang, R., Cheng, L., Liu, P., Huang, K., Gong, Y., Qin, S., & Liu, D. (2021). Effect of GCM credibility on water resource system robustness under climate change based on decision scaling. *Advances in Water Resources*, *158*(September), 104063. https://doi.org/10.1016/j.advwatres.2021.104063

Zolghadr-Asli, B., Bozorg-Haddad, O., Sarzaeim, P., & Chu, X. (2019). Investigating the variability of gcms' simulations using time series analysis. *Journal of Water and Climate Change*, *10*(3), 449–463. https://doi.org/10.2166/wcc.2018.099

Zorita, E., & Von Storch, H. (1999). The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. *Journal of Climate*, *12*(8 PART 2), 2474–2489. https://doi.org/10.1175/1520-0442(1999)012<2474:tamaas>2.0.co;2