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

UNRAVELING THE EFFECTS OF MODELING DECISIONS ON HYDROCLIMATIC PROJECTIONS IN CHILE

TESIS PARA OPTAR AL GRADO DE
DOCTOR EN INGENIERÍA CIVIL

NICOLÁS ANDRÉS VÁSQUEZ PLACENCIA

PROFESOR GUÍA:
PABLO MENDOZA ZÚÑIGA

MIEMBROS DE LA COMISIÓN:
CAMILA ÁLVAREZ GARRETÓN
JAMES MCPHEE TORRES
NAOKI MIZUKAMI
MAURICIO ZAMBRANO BIGIARINI

Este trabajo ha sido parcialmente financiado por la beca ANID Doctorado Nacional No. 21230289 y proyecto Fondecyt No. 11200142

SANTIAGO DE CHILE
2024

RESUMEN DE TESIS PARA OPTAR AL GRADO DE DOCTOR EN INGENIERÍA CIVIL
POR: NICOLÁS ANDRÉS VÁSQUEZ PLACENCIA
FECHA: 2024
PROF. GUÍA: PABLO A. MENDOZA ZÚÑIGA

Desentrañando los Efectos de Decisiones de Modelación en Proyecciones Hidroclimáticas en Chile

Resumen

La estimación de cambio hidrológico se realiza, generalmente, mediante el uso de modelos hidrológicos, los que son calibrados y luego forzados considerando distintas condiciones futuras. Este proceso requiere la toma de distintas decisiones metodológicas, las cuales repercuten en las proyecciones hidrológicas. Debido a la secuencialidad de la toma de decisiones, la estimación de cambio hidrológico se ampara en el contexto de cascada de incertidumbre, es decir, decisiones sucesivas que incrementan la dispersión de los resultados. Entre las decisiones relevantes se encuentran la elección de escenarios de emisiones de gases de efecto invernadero, modelos que proyecten el clima futuro, escalamiento espacial y corrección de sesgo de dichos modelos climáticos, elección de modelos hidrológicos y estimación de sus parámetros, entre otros.

Este trabajo se aboca a analizar el impacto de algunas de estas decisiones en la simulación hidrológica histórica y proyecciones hidrológicas futuras bajo un clima cambiante en 120 cuencas de Chile continental. Para las proyecciones de cambio climático, se consideran 30 Modelos de Circulación General (GCM) bajo el escenario de emisiones SSP5-8.5, siete métodos de corrección de sesgo y tres estratificaciones temporales para aplicar los métodos de corrección de sesgo. Los resultados permiten obtener series de precipitación y temperaturas extremas diarias para todo Chile continental, desde el año 1960 al año 2100.

Por otra parte, la modelación hidrológica se realiza mediante el modelo hidrológico *Variable Infiltration Capacity* (VIC), la cual tiene dos objetivos: (i) identificar el efecto de distintas decisiones de calibración (e.g., elección de los parámetros a regularizar y la función objetivo) en la simulación de variables hidrológicas (e.g., evapotranspiración) y (ii) realizar proyecciones de cambio hidrológico.

Los resultados permiten identificar la relevancia de decisiones que, tradicionalmente, se consideran inofensivas. Por ejemplo, la estratificación temporal a la cual se aplican los métodos de corrección de sesgo impacta en la magnitud y signo de los cambios proyectados, principalmente, de precipitación. Más aún, la estratificación temporal impacta directamente en la estacionalidad promedio proyectada de la precipitación. Respecto a las proyecciones de cambio climático, se esperan disminuciones de precipitación desde la región de Valparaíso al sur, reducciones severas en la caída de nieve en todo el territorio nacional y aumentos en la isoterma 0°C en días con precipitación que favorecerá los riesgos de crecidas e inundaciones. No obstante la dispersión de las proyecciones entre GCMs, hay un alto nivel de acuerdo ($> 90\%$) en la distribución espacial futura del tipo de clima según la clasificación de Köppen-Geiger.

Respecto a la modelación hidrológica, se identifican riesgos importantes en la calibración del modelo VIC considerando solo caudal, pudiendo errar severamente el balance de masa estacional a escala de cuenca a pesar de una buena simulación del caudal. Este tipo de errores afecta las proyecciones hidrológicas a escalas temporales anuales, estacionales y mensuales, así como la magnitud y signo de los cambios hidrológicos proyectados.

Summary

The hydrological change due to changing climatic conditions is estimated through the use of hydrological models, which are calibrated and then forced considering different future conditions (for example, changes in climatic conditions and/or in vegetation land cover). This process requires the definition of different methodological decisions that impact hydrological projections. Due to the sequentially of the decision-making process, the estimation of hydrological change follows the "cascade of uncertainty" framework, that is, successive decisions that increase the uncertainty of the results. Among the relevant decisions are the choice of greenhouse emissions scenarios, models projecting the future climate, downscaling and bias correction of climate models outputs, choice of hydrological models, and estimation of their parameters, among others.

This work focuses on analyzing the impact of the choice of global climate model, choice of bias correction method and the temporal stratification used to apply them, and choice of hydrological model parameters on historical hydrological simulation and future hydrological projections under a changing climate in 120 basins of continental Chile. For climate change projections, 30 General Circulation Models (GCMs), seven bias correction methods are considered and three temporal stratifications used to apply the bias correction methods. The latter decision is usually overlooked within the cascade of uncertainty framework. The results allow us to obtain daily precipitation and extreme temperature series for continental Chile from 1960 to 2100.

On the other hand, hydrological modeling is carried out using the hydrological model *Variable Infiltration Capacity* (VIC), which has two objectives: (i) identify the effect of different calibration decisions of the model parameters on the simulation of different hydrological variables (e.g., evapotranspiration) and (ii) make projections of hydrological change.

The results allow the identification of the relevance of decisions traditionally considered innocuous within the hydrological community. For example, the temporal stratification to which the bias correction methods are applied impacts the magnitude and sign of the projected changes, particularly for precipitation. Furthermore, temporal stratification directly impacts the projected average seasonality of precipitation. Regarding climate change projections under the SSP5-8.5 scenario, reductions in annual precipitation are expected from the Valparaíso Region (32°S-33°S) to the south, severe reductions in snowfall throughout the national territory and a rise in the 0°C isotherm on days with precipitation that will favor the risks of floods and floods. Despite the dispersion of projections between GCMs, there is a high level of agreement in the spatial distribution of the climate type according to the Köppen-Geiger classification.

Regarding hydrological modeling, important pitfalls are identified in the calibration of the VIC model considering only streamflow: the seasonal mass balance at the basin scale can be incorrect despite a good streamflow simulation. This type of error affects hydrological projections at annual, seasonal, and monthly time scales, both in magnitude and sign of the projected changes.

A Miguel, mis padres y mis hermanas.

Acknowledgments

Quisiera partir agradeciendo a mis mentores este camino de la Hidrología: los profesores Ximena Vargas y Pablo Mendoza, quienes me han apoyado hace ya varios años y cuya guía profesional y moral me conducen hasta el día de hoy. Demás está decir que me siento privilegiado de haber sido su alumno. En particular, agradezco al profesor Mendoza todas las horas de trabajo dedicadas a reuniones, redacciones y ediciones de documentos mientras fui estudiante del programa de Doctorado, lo que me ayudó a *estar bien, con ánimo y energía*. Sin su apoyo este trabajo hubiese sido imposible. También agradezco a los profesores James McPhee y Miguel Lagos, con quienes he tenido la oportunidad de trabajar y aprender durante mi paso por el Departamento de Ingeniería Civil. Agradezco profundamente al profesor Martyn Clark, quien apoyó mi postulación a la beca ELAP y supervisó mi pasantía en Canadá. Sus comentarios influyeron profundamente en esta tesis.

Agradezco a los miembros de la comisión revisora: Camila Álvarez, Mauricio Zambrano, Naoki Mizukami y James McPhee, por el tiempo dedicado a la revisión de este trabajo en sus distintas etapas. Sus comentarios contribuyeron a mejorar la calidad de esta tesis.

También agradezco a quienes me recibieron por mi paso en Canadá: Alex, Terava, Maddie, Louise, Caroline, Gaby y Wouter, quienes me hicieron sentir parte del grupo de investigación del *Coldwater Laboratory* y cuya ayuda valoro enormemente. También a agradezco a Lucía por su enorme apoyo al llegar a Canadá.

A todos mis compañeros de oficinas, con quienes hemos gastado varias horas de conversación: Natalia, Julio, Sebastián, Elizabeth y Yerel; y a los funcionarios de la Facultad, cuya ayuda ha sido indispensable en distintas etapas del programa.

Luego de tantos años en la Universidad de Chile, he logrado conocer personas que se han vuelto fundamentales desde hace ya varios años: Jacqueline, Javier, Tomás, Jaime, Paci, Vicky, Felipe, Edu y Carolina. También agradezco a mis amigos de toda la vida: Marina, Rocío y Víctor. Espero que sepan cuánto los aprecio.

Agradezco a mi Miguel, mis padres y mis hermanas por todo el amor, apoyo, guía y motivación que me brindan día a día, y de cuyo afecto y entrega solo puedo sentirme bendecido.

Finalmente, quisiera agradecer el financiamiento que permitió culminar este trabajo: (i) beca de reducción de arancel otorgada por el programa de Doctorado en Ingeniería Civil y la Escuela de Postgrado de la FCFM de la U. de Chile, (ii) Proyecto Fondecyt No. 11200142, el cual fue liderado por el profesor Pablo Mendoza, y en el cual se enmarca esta tesis, (iii) la beca *Emerging Leaders in the Americas Program* (ELAP), otorgada por el Gobierno de Canadá, la cual me permitió realizar una pasantía de seis meses en la Universidad de Saskatchewan, Canadá, (iv) la beca ANID Doctorado Nacional No. 21230289 y (v) a los profesores Pablo Mendoza y Ximena Vargas, que financiaron mi participación en varios congresos nacionales e internacionales.

Powered@NLHPC: Esta tesis fue parcialmente apoyada por la infraestructura de supercómputo del NLHPC (ECM-02).

Table of Content

List of Figures

List of Tables

Chapter 1

Introduction

Improving the predictability of the water cycle is critical for effective water resources management and planning, especially under scenarios driven by human consumption, land use and land cover changes, and climate variability and change (X. Li et al., 2018; Rijsberman, 2006). Although hydrological models enable the estimation of state variables and fluxes in areas without observations, they require (i) accounting the dominant hydrological processes for the system(s) of interest, (ii) estimates of forcing data, (iii) in-situ and/or remote-sensed measurements to assess their performance and (iv) reconciling process, data, and model scales (K. J. Beven, 1990; K. Beven, 1993; Blöschl & Sivapalan, 1995; Peel & Blöschl, 2011). Further, additional decisions (e.g., selection of future scenario, global climate models, bias corrections methods, etc.) are needed to estimate projected hydrologic changes through the "uncertainty cascade" paradigm (Wilby & Dessai, 2010; Clark et al., 2016), which provides an ensemble of opportunity based on the factorial combination of multiple modeling alternatives.

Recently, historical and future hydrological portrayals for continental Chile were produced through the National Water Balance project (DGA, 2017, 2018, 2019b, 2022). In these studies, the CR2MET meteorological product (Boisier, 2023) and the Variable Infiltration Capacity model (VIC; Liang et al., 1994) were used to characterize the hydrology of continental Chile from 1985 to 2015, and project climate change impacts for the period 2030-2060 using outputs from four General Circulation Models (GCMs) from the Fifth phase of the Coupled Models Intercomparison Project (CMIP5; K. Taylor et al., 2012) under the Representative Concentration Pathways 8.5 scenario (RCP 8.5; K. Taylor et al., 2012). However, several subjective modeling decisions were made (e.g., hydrological model, objective function, regularization approaches, downscaling techniques, choice of GCMs). This work aims to revisit methodological choices of previous efforts (DGA, 2017, 2018, 2019b, 2022) to improve historical and future water balance estimates across continental Chile, as well as to update the version of GCMs from CMIP5 to CMIP6.

1.1 Background

The past decade has seen tremendous advances towards more realistic representations in hydrology and land surface models (e.g., Clark et al., 2017; Fan et al., 2019; Fisher & Koven, 2020; Torres-Rojas et al., 2022; Hao et al., 2022), recognizing the role of uncertain parameters in model fidelity (Mendoza, Clark, Barlage, et al., 2015; Cuntz et al.,

2016; Cheng et al., 2023; Yan et al., 2023; Denager et al., 2023). Hence, understanding, quantifying, and reducing hydrological uncertainties is a crucial task, given the large effects that modeling decisions may have on projected climate change impacts (e.g., Addor et al., 2014; Mendoza, Clark, Barlage, et al., 2015; L. A. Melsen et al., 2018; Meresa et al., 2021; Senatore et al., 2022; Mehboob & Kim, 2024). Accordingly, robust model evaluation strategies that go beyond streamflow data and related hydrological signatures are needed (K. Fowler, Coxon, et al., 2018; Cinkus et al., 2023), using additional in-situ measurements when available (e.g., Nijzink et al., 2018; Széles et al., 2020; Pelletier & Andréassian, 2022) and/or satellite remote sensing (SRS) products (e.g., M. Demirel et al., 2018; Koch et al., 2018; Dembélé, Ceperley, et al., 2020; Széles et al., 2020; Pool et al., 2024). The latter has emerged as a powerful tool, providing information on the spatial heterogeneities of landscape properties and hydrological variables (e.g., Lettenmaier et al., 2015; McCabe et al., 2017). Despite the above progress, the assessment of climate change impacts is still challenged by the spread in hydrological projections, given multiple alternatives for modeling decisions (Mendoza, Clark, Barlage, et al., 2015; Clark et al., 2016; Hattermann et al., 2018; L. A. Melsen et al., 2019). Reducing such spread is critical for robust water resources planning and for the execution of adaptation/mitigation responses (Clark et al., 2016; Lehner et al., 2019).

In Chile, the most recent effort to provide hydrological projections at the national scale is the National Water Balance project (DGA, 2017). DGA (2018, 2019a, 2022) used four GCMs, one statistical downscaling method, and one hydrological model (VIC) to estimate projected hydrologic changes. DGA (2018, 2019b, 2022) contributed to improving hydrological understanding across Chile since (i) they set up a physically motivated hydrological model at a 0.05° horizontal resolution for the domain, and (ii) they analyzed ∼120 of catchments with extremely different hydroclimates and physiographic characteristics (e.g., topography, soils, vegetation).

1.2 Motivation

Despite the advances conducted by DGA (2022), other aspects could be explored to improve model fidelity. For example, Murillo et al. (2022) illustrated the added value of including vertical heterogeneity in each simulated unit when using the VIC model. Cortés-Salazar et al. (2023) implemented a routing software (mizuRoute; Mizukami et al., 2016) in more than 120 basins in continental Chile, demonstrating the importance of including the routing process when calibrating hydrological models. Regarding the choice of GCMs for hydrological projections, Gateño et al. (2024) evaluated the performance of 27 CMIP6 GCMs models in a historical (reference) period by proposing a new GCM performance metric (inspired by the Kling-Gupta efficiency metric; Gupta et al., 2009; Kling et al., 2012) that summarizes several climate features for continental Chile. This strategy can be complemented by considering several statistical downscaling methods (e.g., Chegwidden et al., 2019; Guo et al., 2020; Crow et al., 2024; Vásquez et al., 2024). Further, the diversity in landscape and climates across continental Chile provides the opportunity to explore how modeling choices affect water resources for different types of catchments.

This thesis aims to provide a robust portrayal of projected hydroclimatic changes across Chile by exploring the effects of different methodological decisions that are usually overlooked in hydrological modeling and climate change assessments and by including recent advances in continental Chile (Murillo et al., 2022; Cortés-Salazar et al., 2023;

Gateño et al., 2024; Vásquez & Mendoza, 2024) and hydrological modeling (Dembélé, Ceperley, et al., 2020; Clark et al., 2021; Cinkus et al., 2023). To this end, the "uncertainty cascade" paradigm - which consists of the examination of several decisions to provide hydrological projections - results critical and inspires the following research questions:

1. How relevant is the bias correction strategy used to replicate historically observed climatic features and explain the spread of the projected precipitation changes and seasonality?
2. What is the level of (dis)agreement among CMIP6 climate models in projecting the future spatial distribution of climate types, as well as precipitation, temperature, and snowfall changes in continental Chile?
3. Within the context of distributed hydrological modeling, how does the spatial distribution of specific parameters affect the streamflow performance and the catchment-scale seasonal water balance?
4. In the context of climate change impact assessments, how does the calibration/representation of average annual cycles of Q and ET affect the projected hydrological changes at the annual, seasonal, and monthly time scales?

1.3 Outline

This thesis is structured into six chapters: Chapter 1 is the introduction; Chapters 2 to 5, the main body of this thesis (conceptualized as separate but consecutive journal articles); and Chapter 6, corresponding to concluding remarks. Chapters 2 to 5 are summarized as follows:

- **Chapter 2**
  To address question number one, the following decisions within the climate change impact assessments are evaluated: (i) choice of CMIP6 GCMs, (ii) choice of bias correction method (BCM) and (iii) how BCMs are implemented. In this chapter, how bias correction methods are implemented is explored in detail. The results show that the temporal stratification, i.e., the temporal scale used to apply the bias correction method (for example, correcting each season or month), directly impacts the magnitude and signal of projected changes. Moreover, the choice of temporal stratification is crucial to replicate historical seasonalities —, which is essential for further hydrological analyses —and to define the future precipitation seasonality.
- **Chapter 3**
  Since the choice of GCMs is among the most relevant decisions explaining the dispersion of hydrological projections, this chapter addresses question number two, contributing to the existing literature by analyzing projected changes in precipitation, temperature, and snowfall under scenario SSP5-8.5 for 30 CMIP6-GCMs in continental Chile. The results indicate future snowfall reductions, with consequences in rainfall contributing areas to streamflow on rainy days due to a projected rise in isotherm 0°C. Additionally, and despite the spread in projections among GCMs, the spatial distribution of Köppen-Geiger climate groups is delineated, showing a high agreement among models with respect to the projected spatial distribution of such climate groups.

- **Chapter 4**
  The calibration of distributed hydrological models is usually conducted using streamflow (Q) information only. Moreover, a priori spatial distributions of parameters are required to run such models. In this chapter, and to address question number three, different VIC parameters are regularized, showing that its selection impacts not only streamflow-based metrics but also evapotranspiration (ET), fractional snow-covered area (fSCA), land surface temperature (LST), and soil moisture (SM). The main finding is that flux equifinality (from a streamflow point of view) may produce extremely different ET and SM seasonalities.

- **Chapter 5**
  To respond to question number four, hydrological projections in 120 basins in near-natural regimes in continental Chile are provided for two objective functions: (i) using only streamflow-based metrics and (ii) including ET in the calibration of the VIC model. The results suggest that similar simulated streamflow performances could be achieved for extremely different ET seasonalities. Although LST, fSCA, and SM are also analyzed, this chapter focuses on Q and ET since both variables control the overall catchment-scale water balance. The results suggest that, for similar simulated streamflow performance, differences in ET simulations largely impact the projected changes in Q and ET. Additionally, including ET in the objective function affects the relative importance of the fluxes contributing to total runoff and projected changes in total runoff at the annual, seasonal, and monthly time scales. The basins more sensitive to the inclusion of ET in the objective function are located in central and southern Chile (i.e., humid areas), with comparatively larger (lower) annual precipitation amounts (aridity index) and decoupled Q and ET seasonalities.

Chapter 2

The key role of temporal stratification for GCM bias correction in climate impact assessments

Summary

Characterizing climate change impacts on water resources typically relies on Global Climate Model (GCM) outputs that are bias-corrected using observational datasets. In this process, two pivotal decisions are (i) the Bias Correction Method (BCM) and (ii) how to handle the historically observed time series, which can be used as a continuous whole (i.e., without dividing it into sub-periods), or partitioned into monthly, seasonal (e.g., three months), or any other temporal stratification (TS). This chapter examines how the interplay between the choice of BCM, TS, and the raw GCM seasonality may affect historical portrayals and projected changes. To this end, outputs from 29 GCMs belonging to the CMIP6 under the Shared Socioeconomic Pathway 5–8.5 scenario are considered, as well as seven BCMs and three TSs (entire period, seasonal, and monthly). The results show that the effectiveness of BCMs in removing biases can vary depending on the TS and climate indices analyzed. Further, the choice of BCM and TS may yield different projected change signals and seasonality (especially for precipitation), even for climate models with low bias and a reasonable representation of precipitation seasonality during a reference period. Because some BCMs may be computationally expensive, the linear scaling method arises as a computationally cheap diagnostic tool to assess how the choice of TS may affect the projected precipitation seasonality of a specific GCM. More generally, the results presented here unveil trade-offs in how BCMs are applied, regardless of the climate regime, urging the hydroclimate community to carefully implement these techniques[1].

[1]This chapter is published in the Earth's Future Journal. It can be found in the following link: `https://doi.org/10.1029/2023EF004242`.

2.1 Introduction

Understanding and quantifying climate change impacts is crucial for long-term water resources planning and management. Such characterization typically involves hydrologic model simulations forced by an ensemble of scenario-driven meteorological time series obtained from Statistically Downscaled Bias-Corrected (SDBC) Global Climate Model (GCM) outputs (e.g., Addor et al., 2014; Hattermann et al., 2018; Her et al., 2019; Chen et al., 2021; Hanus et al., 2021; Vicuña et al., 2021). This approach usually requires the choice of emission scenario (e.g., Vano et al., 2015; Chegwidden et al., 2019), the choice of GCM (e.g., Hakala et al., 2018; Di Virgilio et al., 2022), the selection of Bias Correction Method (BCM; e.g., Werner & Cannon, 2016; Gutiérrez et al., 2019; Hess et al., 2023), and the choice of observational (or reference) dataset (e.g., Wootten et al., 2021; Rastogi et al., 2022), among others (e.g., hydrologic model and its parameters).

Among the above decisions, the selection and configuration of BCMs is a critical step given the risk of introducing artificial perturbations in GCM outputs (Hagemann et al., 2011; Maurer & Pierce, 2014; Wootten et al., 2021), generating a mismatch between simulated (i.e., obtained from bias-corrected GCMs) and observed (i.e., obtained from a reference dataset) annual cycles of climate variables (e.g., precipitation; Teutschbein & Seibert, 2010; Alder & Hostetler, 2019; Chen et al., 2021), with potential effects on projected climate change impacts and subsequent interpretations and adaptation strategies. A somewhat overlooked step is the strategy for handling the time series when applying BCMs, hereafter referred to as temporal stratification (TS). For example, the bias correction of simulated daily time series can be performed using all the historical period (i.e., a single application of the BCM; e.g., Aryal & Zhu, 2017; Ghimire et al., 2019) or sub-periods of the historical time series, such as seasons (e.g., four applications of the BCM; e.g., Ruffault et al., 2014; Teng et al., 2015; Maity et al., 2019), months (i.e., twelve applications of the BCM; e.g., H. Li et al., 2010; Pierce et al., 2015; Switanek et al., 2017; Matiu & Hanzer, 2022; Wu et al., 2022; Guo et al., 2020), or any other temporal window (e.g., Haerter et al., 2011; Reiter et al., 2018).

Despite the large body of work exploring modeling decisions at the top of the "cascade of uncertainty" (Wilby & Dessai, 2010), climate impact studies have typically relied on subjectively selected TSs. For example, Teng et al. (2015) compared four BCMs (all of them applied with a unique seasonal TS) for hydrological projections in southeastern Australia, concluding that the hydrological model amplifies biases in precipitation after applying the BCMs, and that the large spread in the projected signal of changes in precipitation extremes yields different impacts on runoff. Hakala et al. (2018) applied the quantile mapping (QM) method (using a seasonal TS) to assess whether a hydrological model, forced by SDBC GCMs, can replicate the hydrological climatology observed during a historical reference period, obtaining that, even after bias correction, biases in precipitation and streamflow seasonality persist. To analyze the effects of different observational datasets and BCMs on climate projections, Wootten et al. (2021) used three observational datasets to apply two bias correction methods: (i) the 'Delta' approach with a 3-month moving window, and (ii) the quantile delta mapping (QDM) method over four periods consisting of three non-overlapping months (i.e., a seasonal TS). They concluded that the selection of BCMs and observational datasets have different impacts on historical and projected time series for different variables, although they did not isolate the effect of the choice of TS.

Other studies have focused on the ability of different BCMs to reproduce historically observed climate indices (e.g., Gutmann et al., 2014; François et al., 2020; Xavier et al., 2022), or the effects on climate projections (e.g., Maurer & Pierce, 2014; L. A. Melsen et al., 2018). In this regard, different BCMs have been proposed to reduce the artifacts induced by quantile-based methods (Maraun, 2016), to preserve the trends of raw GCM projections (e.g., Hempel et al., 2013; Pierce et al., 2015), include non-rainy days (e.g., Maity et al., 2019) or the intermittent nature of precipitation (e.g., H. Li et al., 2010), without emphasizing the role of the TS and the evaluation timescale. More recently, Vogel et al. (2023) proposed a framework to evaluate downscaling and BCMs for climate change studies and demonstrated it over Australia using four GCMs, three BCMs, and two downscaling methods, considering different TS (monthly, 3-month, and multi-time scales) for the BCMs. They suggested that the TS may influence the analysis after bias correction and should be adequately selected after a careful bias assessment.

Although the preceding studies have covered domains with specific climate types, the trade-offs in selecting TS, BCMs, and GCMs for estimating historical biases (after applying BCMs) and projections across contrasting climates remain unclear. This chapter seeks to disentangle the relative contribution of the choice of GCMs, BCMs, and (particularly) TSs to the spread of bias-corrected time series at the annual, seasonal, and monthly timescales during historical and future periods rather than finding the 'best' configuration for the assessment of climate change impacts. It could be hypothesized that the choice of TS may have substantial impacts on historical and future climate projections. Specifically, the following research questions are addressed:

1. To what extent does the choice of bias correction method and temporal stratification alter historical GCM simulations across different climate regions?
2. What are the effects of bias correction methods and temporal stratification on the projected signal and seasonality of different climate variables?
3. Are there any connections between the effects of TS (on historical biases and projections) and the capability of raw GCM output to replicate historically observed climatology?

To seek answers, this chapter evaluates the performance of 29 SDBC GCMs from the sixth phase of the Coupled Models Intercomparison Project (CMIP6; O'Neill et al., 2016) over different climate groups in continental Chile. Seven methods (three univariate and four multivariate) are used to correct biases in precipitation and maximum and minimum temperature. All BCMs are applied considering three different TSs: (i) using the entire period (i.e., all daily data simultaneously used for one application of the BCM), (ii) seasonally (i.e., four applications of the BCM using four seasonally stratified time series), and (iii) monthly (i.e., twelve applications of the BCM for twelve monthly stratified time series).

2.2 Study area and datasets

2.2.1 Study area

The study domain is continental Chile, which is suitable for a comprehensive assessment of the GCM-BCM-TS interplay in very different climate types. Figure 2.1 shows the spatial distribution of mean annual precipitation, mean annual temperature, and three climate indices. The snowfall fraction $SF = Sn/P$ (Figure 2.1d) is the fraction of mean annual precipitation (P, Figure 2.1b) falling as snow (Sn). The aridity index (Figure 2.1e) is the ratio between mean annual potential evapotranspiration (PET) and mean annual precipitation. PET is computed using the Oudin et al. (2005) formula – available in the R Package airGR (Coron et al., 2017) - which requires air temperature (provided at daily time steps here) and latitude as inputs. To estimate Sn, it is considered that snowfall occurs when the mean daily temperature is below 2°C (Jennings et al., 2018; Han et al., 2019; Sepúlveda et al., 2022), and p-seasonality is computed with the formula proposed by Woods (2009, Eq. 14). Finally, the precipitation seasonality (p-seasonality, Figure 2.1f) indicates whether most precipitation falls during winter (negative values) or summer (positive values). In this paper, the names of the seasons are used within the context of the Southern Hemisphere (i.e., winter refers to months JJA, while summer to DJF).

In the northern area (17°S-25°S), two main climate zones can be identified: (i) the super-arid coastal area, with very low annual precipitation amounts (<50 mm/yr), and (ii) the Altiplano region, with lower temperatures due to increasing altitude and larger annual precipitation (∼200 mm/yr). The mean annual precipitation increases towards the south, although the Andes Cordillera generates a west-east gradient, with larger precipitation amounts and lower temperatures on the western slopes of the Andes Cordillera compared to the valleys. Moving south from ∼37°S, the altitude of Andean mountains progressively decreases, as well as the contribution of snowmelt to runoff, whereas precipitation increases. South from 45°S, a west-to-east precipitation gradient produces high precipitation amounts on the coast (>2,500 mm/yr), whereas a dry climate develops in Patagonia a few kilometers to the east, with decreasing precipitation amounts. In summary: (i) most snowfall occurs in the Andes Cordillera, though snowfall events can also occur in the valleys of Austral Chile (<45°S); (ii) the hydroclimate is water-limited ($PET/P > 1$) in approximately half of the Chilean territory, especially from ∼35°S to the north, whereas the hydroclimate of the south is energy limited ($PET/P < 1$); and (iii) most precipitation in Chile falls during the winter (red color in panel f), being the Altiplano (northern Chile) and Patagonia (∼50-55°S) two notable exceptions. For a more comprehensive review of the climate and weather of Chile, readers are referred to Aceituno et al. (2021) and Vásquez et al. (2021).

2.2.2 Datasets

The gridded meteorological product CR2MET v2.5 (Boisier et al., 2018; DGA, 2022) is considered as the observational baseline (hereafter reference dataset). CR2MET precipitation estimates (pr) are obtained through a combination of (i) logistic regression models and (ii) multiple linear regression models that use ERA5 reanalysis outputs (Hersbach et al., 2020) and geomorphological attributes as predictors and daily precipitation from

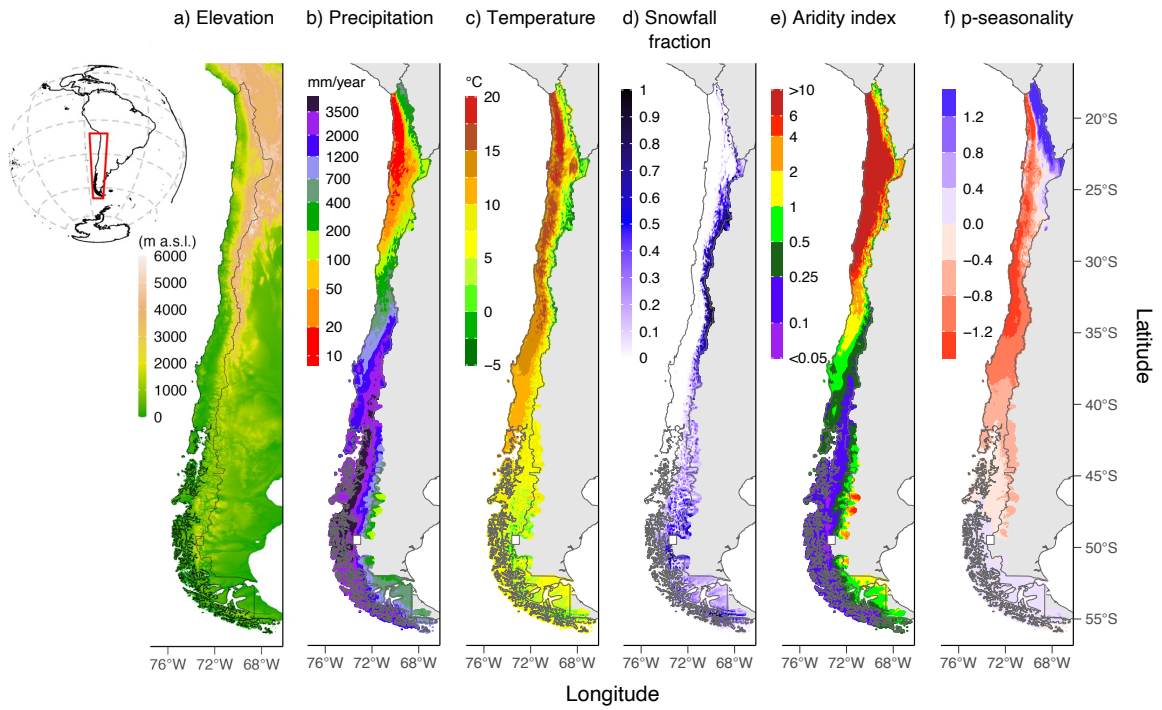

Figure 2.1: Main physiographic and climate attributes of continental Chile for the period Apr/1980-Mar/2014 (34 water years): (a) elevation, (b) mean annual precipitation, (c) mean annual temperature, (d) snowfall fraction, (e) aridity index, and (f) p-seasonality.

meteorological stations as predictands. For daily extreme temperatures (tmax and tmin), land surface temperature from MODIS AQUA and TERRA (Wan, 2014) are also included as predictors. All variables (pr, tmax, and tmin) are available at a daily time step for the period January/1979-March/2020, covering continental Chile at a horizontal resolution of 0.05° x 0.05°. The mean daily temperature is computed as the average between tmax and tmin (e.g., Demaria et al., 2013). It should be noted that CR2MET is, arguably, the most accurate meteorological dataset for continental Chile since its development incorporated local meteorological stations.

For climate projections, outputs from 29 GCMs from the CMIP6 are considered (O'Neill et al., 2016), based on the data availability for pr, tmax, and tmin during the historical and projected periods, and the SSP5-8.5 scenario for being the worst in terms of greenhouse emissions and the 'business as usual' development case. The name and horizontal resolution of each GCM are included in Table A.1.

2.3 Methodology

Figure 2.2 shows the main steps of the approach. First, climate zones across Chile are delineated using cluster analysis (step 1) to examine possible relationships between climate types and the GCM-BCM-TS interplay. Step 2 considers different strategies for correcting biases in GCM outputs (i.e., seven bias-correction methods are applied using three different temporal stratifications). In step 3, several climate indices derived from precipitation and temperature are computed at different time scales (e.g., annual, seasonal, and monthly mean values) for a historical and future period. Finally, an Analysis

of Variance (ANOVA) is conducted to quantify the relative contribution of the different decisions to the spread of historical estimates. More details can be found in the following sections.

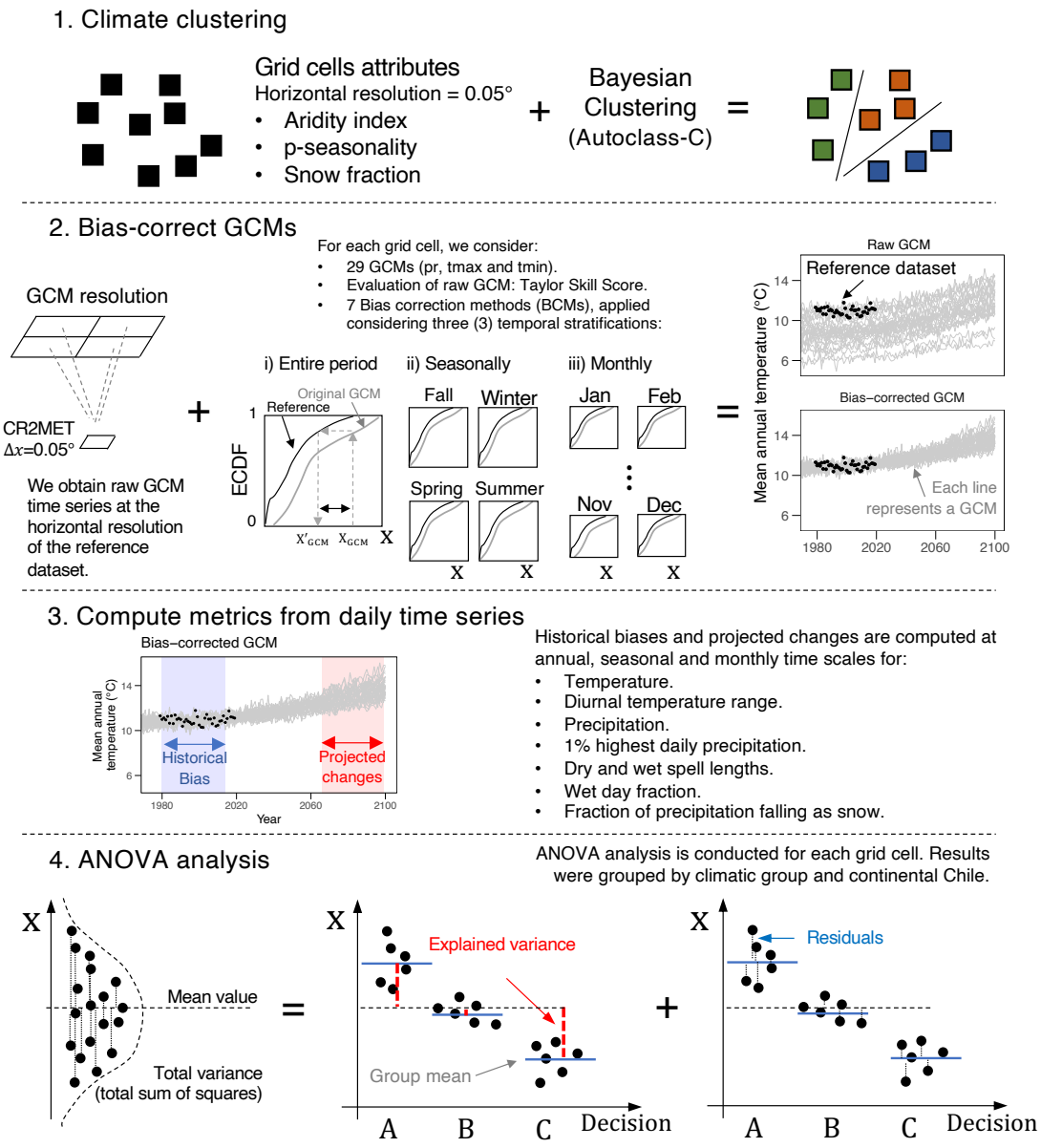

Figure 2.2: Diagram of the methodology used in this study

2.3.1 Climate clustering

A Bayesian clustering is performed to identify climate zones across Chile. To this end, the aridity index (PET/P), the p-seasonality, and the fraction of precipitation falling as snow are used as explanatory variables, since they reflect observed hydrological behaviors (W. J. Knoben et al., 2018). Prior climate groups are defined with the Autoclass-C software (Cheeseman et al., 1988, 1996), which has been previously used in hydrological applications (e.g., Sawicz et al., 2011). Clustering results are subsequently refined through visual inspection, grouping small clusters based on spatial proximity and climate similarity.

2.3.2 Raw GCM performance

The Taylor Skill Score (TSS; K. Taylor, 2001) is used to evaluate the role of the raw GCM performance and its interplay with BCM and TS to explain SDBC-biases and projections at different time scales. The TSS is computed at the grid cell level (0.05° x 0.05°) for the period 1980-2014, contrasting downscaled GCM outputs against the reference dataset, as is commonly done for local climate impact assessments (e.g., Lafon et al., 2013). In this study, TSS is computed for precipitation, as shown in Eq. 2.1.

$$TSS = \frac{4(1+R)}{\left(\hat{\sigma} + \frac{1}{\hat{\sigma}}\right)^2 (1+R_o)} \tag{2.1}$$

where R is the Pearson correlation coefficient between the raw GCM and the reference mean seasonality, and $\hat{\sigma} = \sigma_{GCM}/\sigma_{REF}$ is the ratio between the standard deviation of raw monthly values (σ_{GCM}) and the reference (σ_{REF}). R, and $\hat{\sigma}$ are computed using simulated and observed mean monthly values of each variable (i.e., 12 values of GCMs vs. 12 reference values). R_o is the maximum achievable Pearson correlation coefficient for a specific GCM, which is assumed to be $R_o \cong 1$ to simplify the analysis. When $R \to R_o$ and $\hat{\sigma} \to 1$, the $TSS \to 1$. Alternatively, $TSS \to 0$ when R decreases or $\hat{\sigma}$ approaches zero or infinity. Hence, TSS ranges between 0 and 1. Further, TSS is computed for each climate group, estimating the mean group climatology through spatial averages.

2.3.3 Bias correction of GCMs

Bias correction methods

Raw GCM outputs are spatially downscaled to the CR2MET grid using inverse distance weighting, considering the four closest GCM grid cells. Seven bias correction methods are used, including three univariate and four multivariate techniques, listed in Table 2.1 and briefly reviewed here. The quantile delta mapping (QDM) preserves the projected change for each quantile while correcting the bias. Empirical cumulative density functions are estimated for the historical reference ($F_{h,ref}$), the raw historical GCM ($F_{h,GCM}$), and the raw projected GCM ($F_{p,GCM}$) to relate (X) with the cumulative probability (τ). For a specific value during the historical period $X_{h,GCM}$, the correction (for pr) is given by $X'_{h,GCM} = F^{-1}_{h,ref}\left(F_{h,GCM}\left(X_{h,GCM}\right)\right)$, while for a projected raw GCM value $X_{p,GCM}$, the corrected value is $X'_{p,GCM} = \Delta \cdot F^{-1}_{h,GCM}\left(F_{p,GCM}\left(X_{p,GCM}\right)\right)$, where Δ is computed as $\Delta = X_{p,GCM}/F^{-1}_{h,GCM}\left(F_{p,GCM}\left(X_{p,GCM}\right)\right)$ for precipitation. In words, for each quantile, the bias-corrected future precipitation value is the observed historical value for that quantile, multiplied by the projected raw change from GCM data for the same quantile.
The asynchronous regression (AR) relies on a piecewise linear regression calibrated with sorted raw GCM and reference data during a historical period (i.e., $F_{h,ref}$ is a function of $F_{h,GCM}$). Although a simple linear regression could be used, the error in the tails of the regression can be large and, therefore, the data is split in groups (up to six) to conduct separate linear regressions for each group, aiming to reduce errors in low and high values. To bias-correct projected values, the calibrated piecewise linear regression is applied. The quantile regressions neural network (QRNN) uses neural networks to bias correct the sorted data (i.e., quantiles) from simulations and the reference. QRNN is a

flexible model since it does not assume a specific relationship between the raw GCM and the reference data.
The rank resampling for distributions and dependences (R^2D^2) corrects the covariance among sites and/or variables through four steps: (i) the univariate bias correction of each variable/site separately, (ii) the selection of one variable/site and the computation of the ranking for all variables/sites, (iii) for a specific date, select the same ranking in the reference period for the dimension selected, and (iv) the shuffling of the other variables and sites to maintain rank structure.
The 'multivariate bias correction' family (MBC) includes three different methods using the Pearson correlation coefficient (MBCp), the Spearman rank correlation coefficient (MBCr), and an N-dimensional probability density function (MBCn) to transform the raw correlated GCM data (i.e., the intervariable dependence structure) through consecutive iterations. For MBCp and MBCr, the transformation relies on the Cholesky matrix decomposition and the correction of the covariance matrix. Conversely, MBCn relies on an orthogonal rotation, the application of QDM to these orthogonal variables, and, finally, the application of an inverse matrix (the one used to compute the orthogonal variables) to obtain the resulting data. The reader is referred to the studies listed in Table 2.1 for more details on the methods. It should be stressed that the aim is to evaluate how different decisions may impact historical biases and future projections, rather than compare the ability of particular BCMs to remove biases.

Table 2.1: Methods considered in this study to bias-correct GCMs outputs (pr, tmax, and tmin). Univariate methods bias-correct each variable separately, while multivariate methods also adjust the raw GCM correlations to better replicate the observed inter-variable correlations.

Acronym	Name	Type	Reference
QDM	Quantile Delta Mapping	Univariate	Cannon et al. (2015)
AR	Asynchronous Regression		Dettinger et al. (2004); Stoner et al. (2013)
QRNN	Quantile Regression Neural Network		Cannon (2011)
R^2D^2	Rank Resampling for Distributions and Dependences	Multivariate	Vrac and Thao (2020)
MBCp	Multivariate Bias Correction method - Pearson		Cannon (2016)
MBCr	Multivariate Bias Correction method - Rank		
MBCn	Multivariate Bias Correction method – QDM		Cannon (2018)

All bias correction methods were applied using the statistical software "R" (http://www.r-project.org/). The QDM, MBCp, MBCr, MBCn, and R^2D^2 methods were applied using the library "MBC" (Cannon, 2018). QRNN was implemented using the "qrnn" library (also available in R), while the AR method was implemented following Stoner et al. (2013). To reduce the computational effort, 100 grid cells are randomly selected within each climate group, and all subsequent analyses are conducted at these grid cells ($100 \cdot N_{clusters}$).

Choice of the temporal stratification

Bias correction methods can be applied using different stratification strategies. For example, a BCM can be applied at daily time steps using all the data in the historical period (usually 30 years), which means that all ∼10,950 days (∼365 days · 30 years) are simultaneously bias-corrected. For a seasonal TS, BCMs are applied four times, each one considering ∼2730 days (∼91 days · 30 years), whereas for a monthly TS, the BCM is applied 12 times considering ∼900 days (∼30 days · 30 years). Note that other temporal stratifications could be considered. Here, BCMs are applied to daily time series of pr,

tmax, and tmin (e.g., Rastogi et al., 2022) using the entire time series in the historical period (1980-2014), and stratifying the data seasonally and monthly, since these TSs are typically considered for climate change impact assessments. For all combinations of BCM and TS, daily time series from 1980 to 2100 are obtained.

2.3.4 Climate indices

In this chapter, several climate indices are considered relevant to evaluate historically observed hydrological responses (e.g., Gutmann et al., 2014), including (i) mean annual, seasonal, and monthly total precipitation, (ii) highest 1% daily precipitation, (iii), wet-day fraction, (iv) wet and dry-spell lengths, (v) fraction of precipitation falling as snow, and (vi) annual, seasonal and monthly averages of mean daily temperature and diurnal temperature ranges. To estimate the mean annual snowfall, all precipitation amounts for days with a mean daily temperature below 2°C are added. Wet-spell and dry-spell lengths (mean consecutive rainy and non-rainy days, respectively), as well as the wet-day fraction (mean fraction of rainy days), are computed as in Gutmann et al. (2014), considering 0.1 mm/d as a threshold. To examine the capability of BCMs to replicate historically observed climate indices, the difference between SDBC-GCM outputs and the reference dataset during the historical period 1980-2014 are computed as a percent bias. Additionally, the effects of BCMs on climate projections are analyzed by computing the relative change for the period 2065-2099 with respect to the historical period (1980-2014). For consistency, the climate clustering (section 3.1) and application of BCMs (section 3.3) consider the period 1980-2014 as the reference.

2.3.5 Analysis of Variance

To evaluate the relative contribution of the BCM and TS decisions to the spread of SDBC-biases, an analysis of variance (ANOVA) is performed for each combination of GCM and grid cell. In this case, the ANOVA analysis is simplified as:

$$TV = BCM + TS + Residual \tag{2.2}$$

where TV stands for the total variance of SDBC-biases, and the residual term is the variance not explained by the BCM nor the TS for a specific GCM-grid cell combination. If the choice of TS has no impact on the biases in climate indices (after bias correction), the application of the BCM should be able to reduce biases at all temporal scales (e.g., annual, seasonal, or monthly), regardless of the GCM considered. To summarize the information at the grid cell level, the average of BCM/TV, TS/TV, and $Residual/TV$ fractions is computed across GCMs, whereas for the climate groups, the mean relative contribution (estimated by BCM/TV, TS/TV and $Residual/TV$) of TS and BCM to the spread is computed as the average of fractions across the grid cells within that group.

2.4 Results

This section shows the climate clustering results, the historical biases after applying the BCMs, and the relative contributions of different methodological choices to historical biases of climate indices at the annual and seasonal scales. Further, TSS performance is included to examine connections between the raw seasonality of the GCMs and the selection of BCM and TS. For simplicity, only the results for precipitation are shown here (see section A for more results).

2.4.1 Clustering

The Bayesian clustering and subsequent spatial aggregation through visual inspection provided ten climate groups for continental Chile (Figure 2.3). In general, the clusters follow two main climate patterns: (i) a latitudinal precipitation gradient, from very arid (north) to humid (south), and (ii) a west-east gradient from the coast to the Andes Cordillera. Although northern Chile encloses groups 1, 2, and 3, clusters 2 and 3 are located in the Altiplano region, where larger precipitation and lower temperatures are observed. Groups 5, 6, and 8 span the coast and valley, whereas groups 4 and 7 are located in the Andes. Finally, groups 9 (the rainiest group) and 10 are in southern Chile, characterized by large precipitation amounts in the Andes Cordillera and the coast, with decreasing precipitation and temperature towards the east (Patagonia).

2.4.2 Performance metrics after bias correction

Figure 2.4 shows precipitation biases (after bias correction) in three different climate groups. The results show that, regardless of the combination of GCM, BCM, TS, and grid cell, biases in annual amounts are close to zero (Figure 2.4a). When the BCM is applied using all the data in the historical period (Figure 2.4b, left), biases in monthly precipitation amounts can be large, although the magnitude varies among climate groups. In climate group 2 (Altiplano region), precipitation occurs mostly during the summer (DJF); in this season, the median bias associated with January precipitation is relatively lower - though still considerable (>20%) - compared to the remaining months. In group 6, most precipitation occurs during the winter (JJA), and biases can be found in any month. In group 10, precipitation falls uniformly throughout the year, with slightly larger amounts and larger biases during the summer (DJF). Monthly precipitation biases persist when the BCM is applied seasonally (Figure 2.4b, center). However, these biases are generally lower compared to the case when the bias correction is applied using the entire dataset, especially in climate group 10. As expected, biases are nearly removed at the monthly time scale with a monthly TS (Figure 2.4b, right), regardless of the GCM, bias correction method, grid cell, or climate group.

No considerable differences can be found in the ability of different BCMs to remove biases in P, T, and diurnal temperature range during the historical (training) period (Figures A.1-A.3). For P-1% (Figure A.4), the AR and QRNN methods yield larger biases for all temporal scales and TS. For DSL (Figure A.5), the QRNN provides larger biases when the entire period is used as TS. When BCMs are applied at seasonal and monthly TSs, larger biases are obtained with the R^2D^2 method. For WSL (Figure A.6), the QRNN yields

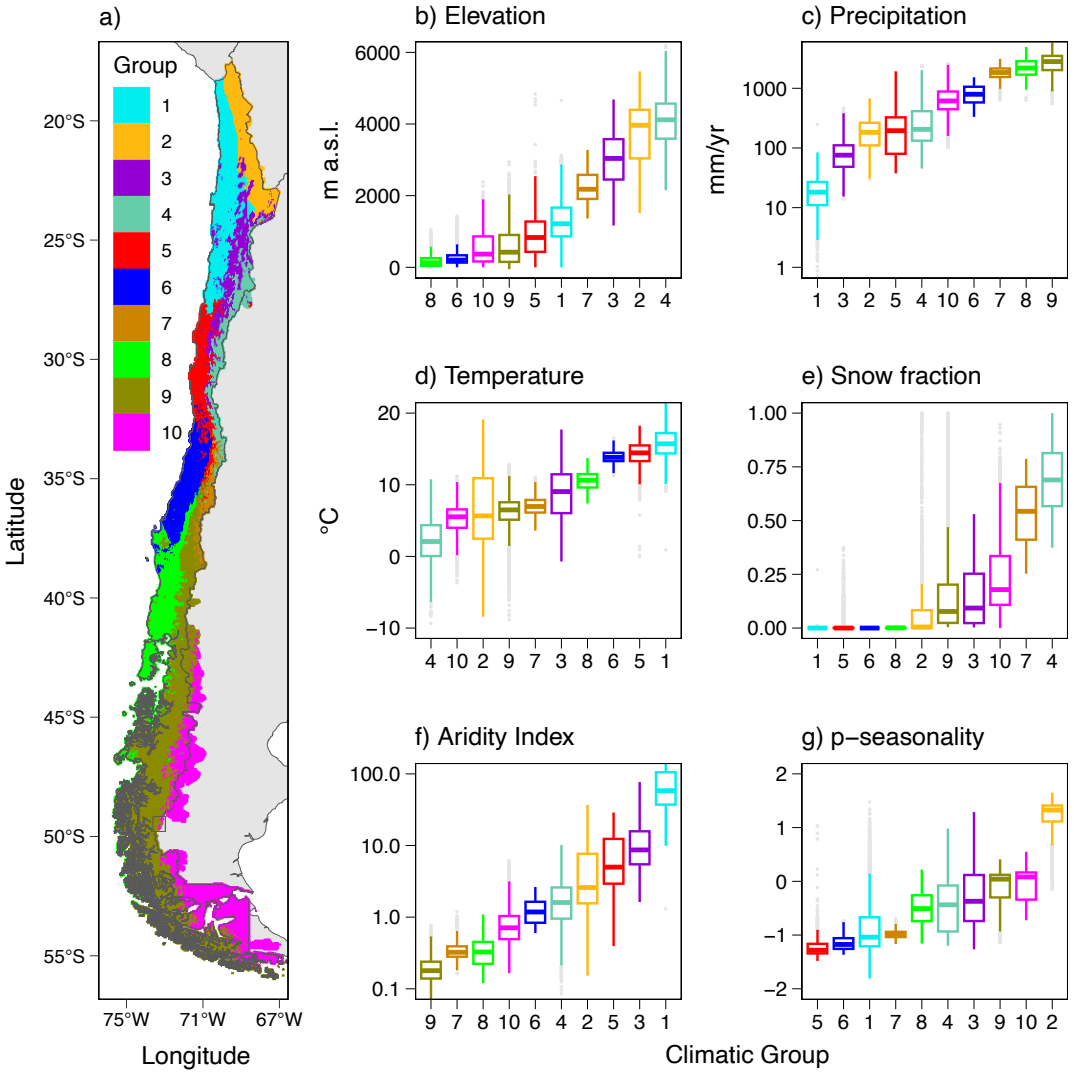

Figure 2.3: (a) Spatial distribution of climate clusters in continental Chile based on snowfall fraction, aridity index, and p-seasonality. The following attributes are ordered by the median of each group: (b) elevation, (c) precipitation, (d) temperature, (e) snowfall fraction, (f) aridity index, and (g) p-seasonality. All climate indices were computed for the period 1980-2014. Notice that the boxplots in panels b-g are sorted according to the median value (horizontal line within each box), and the group's order on the x-axis differs among variables.

the largest biases for the entire period TS, whereas, for seasonal and monthly TSs, the MBCn method produces the largest biases. For snowfall (Figure A.7), univariate (QDM, AR, and QRNN) BCMs yield larger biases compared to multivariate (MBCp, MBCr, MBCn, and R^2D^2) techniques, regardless of the TS. Such results are not surprising since multivariate methods aim to include the correlation between variables. However, among the multivariate BCMs, the MBCn and R^2D^2 methods show a better ability to reduce GCM biases for all TSs (especially for seasonal and monthly stratifications). It is worthy stressing that this analysis is only conducted for the training period and, therefore, these relative BCM performances are not expected to be preserved in independent evaluation periods (e.g., H. Li et al., 2010; Johnson & Sharma, 2011; Gutmann et al., 2014; Maity et al., 2019).

Figure 2.5 displays the relative contributions of the BCM, TS, and residuals for mean

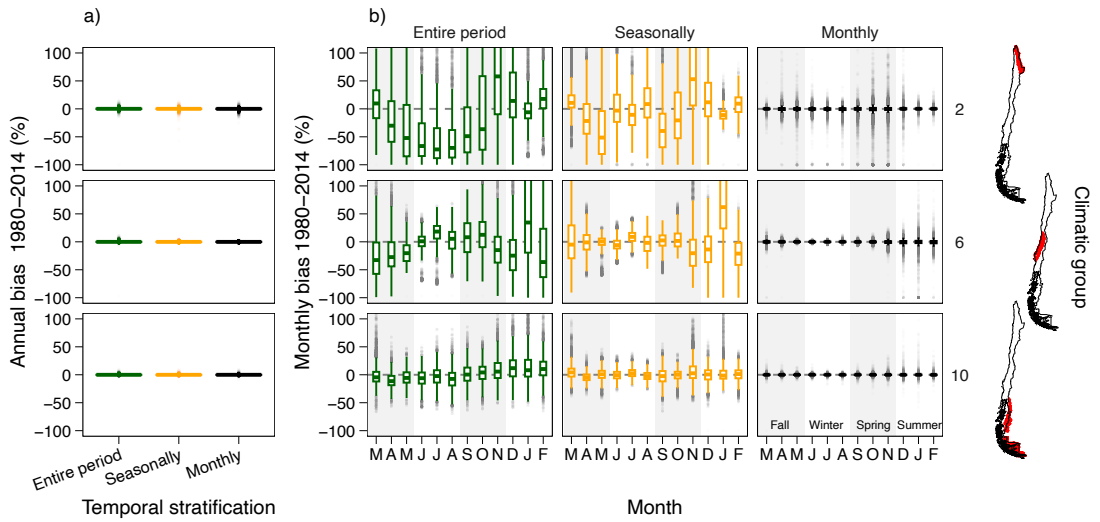

Figure 2.4: Historical biases (after bias correction) in precipitation at the (a) annual and (b) seasonal time scales in three climate groups (rows) after applying the BCMs. The columns in panel b) show results for the three TSs used to apply the BCMs. Each boxplot comprises results from the 100 grid cells within a specific climate group, 29 GCMs, and seven BCMs. The different seasons are highlighted through grey-white areas.

annual, seasonal (summer and winter), and monthly (January and July) precipitation biases averaged across 1,000 grid cells in continental Chile. Two seasons and months are shown to examine possible differences between the dry and wet seasons. Additionally, the results from different GCMs are stratified according to their historical raw performance, measured by the Taylor Skill Score. As in Figure 2.4, the ANOVA analysis for historical biases shows differences among temporal stratifications, especially when seasonal and monthly biases are compared to annual biases (Figure 2.5a). Because the relative contributions of BCM and TS to precipitation biases do not greatly differ among climate groups, results are shown at the national scale. The choice of BCM explains most of the variance for the mean annual precipitation bias, whereas the choice of TS explains almost all the variance for mean seasonal and monthly precipitation biases. It is worth noting that the biases at the annual scale are, in general, very low (Figure 2.4, <1%), and that the relative importance of the choice of TS for seasonal and monthly biases does not decrease for GCMs with high TSS values. The latter result is counterintuitive since one might expect GCMs with good raw precipitation seasonality to be effectively bias-corrected, regardless of the TS selected. For variables related to quantiles (highest 1% daily precipitation, dry and wet-spell lengths, and wet-day fraction), the relative importance of BCMs increases for GCMs with higher TSS, being BCM the most important decision, even at seasonally and monthly time scales (Figure A.10).

2.4.3 Projected changes

This section analyzes the interplay between the choice of TS, the raw GCM precipitation seasonality, and its effects on projected changes in precipitation for the period 2065-2099 (with respect to 1980-2014) at different time scales. Figure 2.6 displays projected changes in mean annual, seasonal, and monthly precipitation for one grid cell located in central Chile (red dot in map) and one GCM (INM-CM4-8) with a high R value. For this GCM and grid cell, $TSS = 0.76$ during the period 1980-2014, with a Pearson

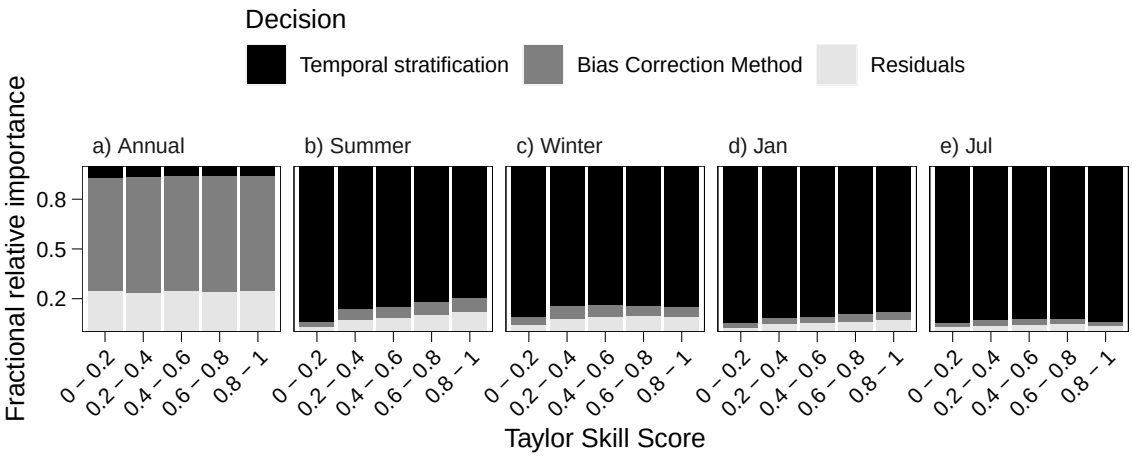

Figure 2.5: Relative importance (as a fraction averaged from all grid cells and GCMs for continental Chile) of the bias correction method and the temporal stratification to explain the precipitation biases at the annual, seasonal (DJF and JJA), and monthly (January and July) time scales during the historical period (1980-2014), for different levels of historical GCM performance (x-axis). Biases are computed after applying BCMs.

correlation coefficient between mean monthly raw GCM and reference amounts of 0.98, and a 41% underestimation of the standard deviation. The high value of R indicates a good seasonality of raw GCM outputs. Figure 2.6 shows that different BCMs yield a high dispersion in projected changes of mean annual precipitation (different lines), with little influence on the selected TS (x-axis of each subplot). Additionally, all BCMs alter the raw GCM projection. For example, if all BCMs are applied using the entire dataset, projected changes in summer precipitation range between -8% to 5%, whereas the raw projection is close to -30%. The application of MBCn using the entire period yields a positive projected change in the mean summer precipitation, while a seasonal and monthly application of the same BCM projects a decrease in summer precipitation. The results for individual months (January and July) reveal more dispersion and interaction among BCMs and the choice of TS. For example, applying the BCM with the entire time series results in positive and negative projections of mean July precipitation (the rainiest month for this grid cell). Similarly, different TSs can also provide different projected signals.

Figure 2.6 reveals that the choice of TS affects the signal of projected changes in summer precipitation (e.g., for the MBCn method), as well as in January and July precipitation amounts. The TS can be considered relevant for a specific grid cell if it is able to switch the projected signal of a variable for a particular GCM-BCM combination. This is, for example, the case of mean July precipitation (Figure 2.6e), for which the signal of projected changes is different among TSs for the MBCn, MBCr, and R^2D^2 methods.

Figure 2.7 shows, for all the grid cells analyzed, the fraction of "well-behaved" GCMs (i.e., with $TSS \geq 0.7$; e.g., Kwon et al., 2019) for which the selection of TS leads to different signs in projected precipitation changes. Note that the number of GCMs that meet the performance requirement - obtained by spatially averaging the number of GCMs with $TSS \geq 0.7$ at each latitudinal band - varies along the domain. In general, the choice of TS does not alter the signal of projected changes in mean annual precipitation, although a few GCMs are affected by this decision in some areas (e.g., northern Chile). Nevertheless, the effects of TS are more evident in seasonal projections (Figure 2.7b and 2.7c). During the summer, >50% of the number of GCMs are affected by the TS in central

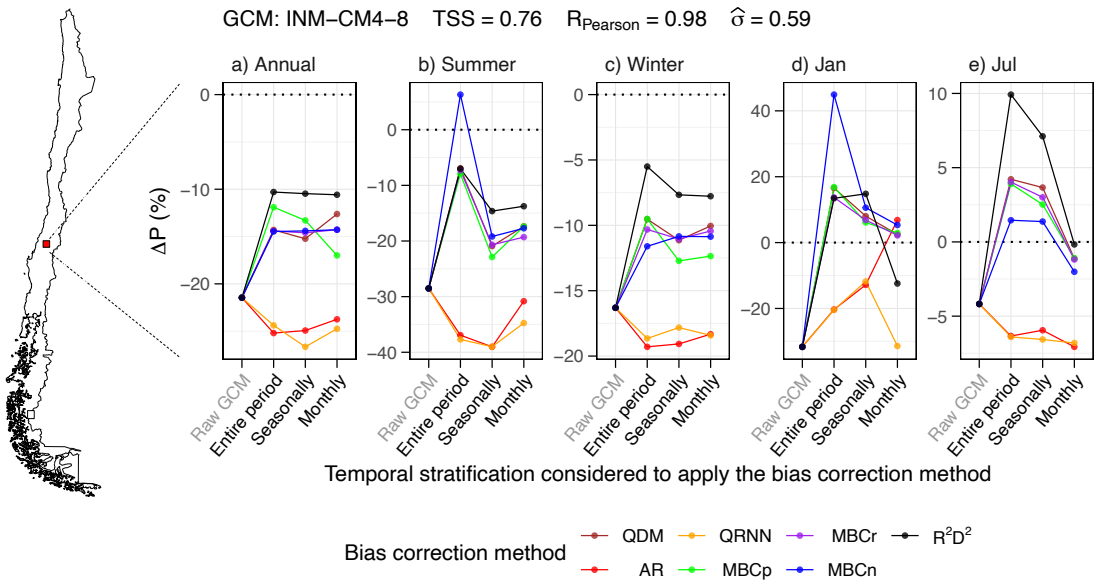

Figure 2.6: Projected change in annual, seasonal (summer and winter), and monthly (January and July) precipitation for different temporal stratifications (x-axis) and bias correction methods (lines). All combinations of TS and BCM decisions, along with projected changes from the raw (biased) GCMs, are displayed. The results are valid only for the grid cell shown and the GCM INM-CM4-8. The metrics (e.g., TSS) were computed using the raw (biased) GCM data for the period 1980-2014.

Chile (dry season). During winter, the Altiplano region and part of southern Chile are largely influenced by the choice of TS. It should be noted, however, that the summer season in central Chile and the winter season in the Altiplano region are dry seasons. Therefore, while the signal of projected changes may vary for different TSs, the precipitation amounts involved are small. For mean monthly January and July precipitation, the choice of TS is even more relevant. Indeed, nearly all GCMs are affected by the TS along the coast of northern Chile, while ∼50% of the GCMs yield different signals in projected changes for different TSs in central Chile. The case of July is more interesting since it is the rainiest month in most of continental Chile. In July, ∼50% of the GCMs are affected by the TS along the central Chilean Andes (western border), impacting the accumulation of snow and, therefore, meltwater volume and timing estimates for the spring and summer seasons. In southern Chile, one can find grid cells where GCMs are affected by the TS decision, though that fraction is lower compared to the central Chilean Andes.

Figure 2.8a compares the raw GCM output (obtained from the GCM ACCESS-CM2) and the reference precipitation seasonality over a historical period at one grid cell located in central-southern Chile (red dot on the map). For this GCM-grid cell combination, $TSS = 0.96$, $R = 0.94$ and $\hat{\sigma} = 1.08$. Note that the GCM simulates the maximum monthly precipitation in July instead of June (when the maximum occurs according to the reference). Figure 2.8b displays, for the same GCM-grid cell, the projected precipitation seasonality for each BCM-TS combination (thin lighter lines). The results show that applying a BCM using the entire period (green lines) provides the same seasonality as the raw GCM; however, seasonal and monthly TSs distort the raw projected seasonality. Further, when BCMs are applied using a monthly TS (black/gray lines), the projected month of maximum precipitation is June, whereas for seasonal and entire periods, such a month is July. Additionally, seasonal and monthly TSs yield higher precipitation fractions

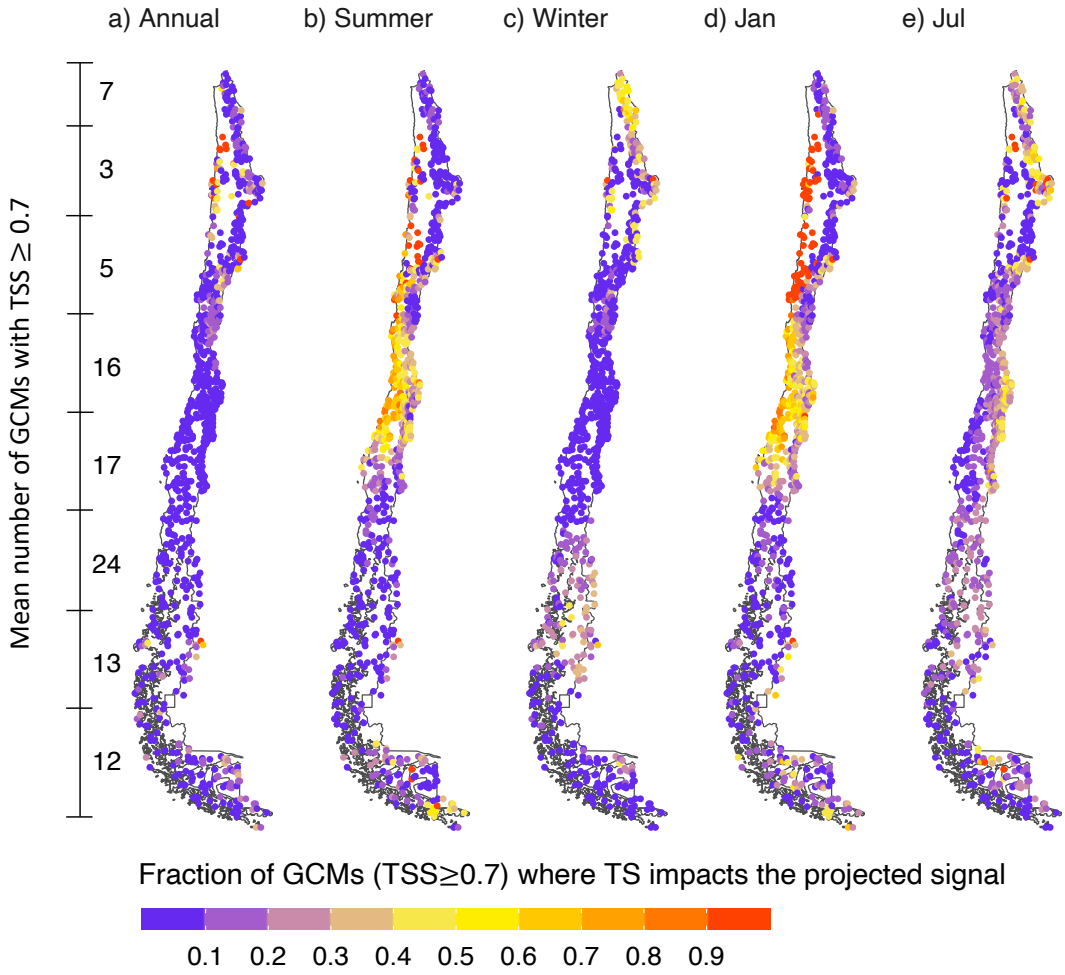

Figure 2.7: Fraction of GCMs with acceptable performance (i.e., with $TSS \geq 0.7$) for which the TS yields different projected precipitation signals. The number of GCMs that meet the threshold criteria at each 5° latitudinal non-overlapping band is computed as the average of GCMs with $TSS \geq 0.7$ from all grid cells within that band.

(compared to the raw GCM) during April and May, and smaller values during September and October. Such differences in projected precipitation seasonality may affect any subsequent analyses of simulated hydrological fluxes and states.

To examine the extent to which projected precipitation seasonality is affected by the temporal stratification, the projected maximum mean monthly precipitation is analyzed. Hence, for each GCM-grid cell combination, three curves obtained with the three temporal stratifications (each obtained by averaging the projections among BCMs for each GCM) are contrasted. It is considered that the TS affects the projected seasonality if the month where the maximum mean monthly precipitation amount occurs differs. Conversely, if such a month is the same for the three TSs, it is considered that this decision does not impact the seasonality. Figure 8c displays the fraction of the number of GCMs with $TSS \geq 0.7$ for which the TS impacts the projected precipitation seasonality. Interestingly, the number is relatively high (>40%) for most of continental Chile. The fraction of GCMs affected by the TS decision is even higher in northern Chile, the central Chilean Andes, and the southernmost part of Chile, where more than 60% of GCMs are affected.

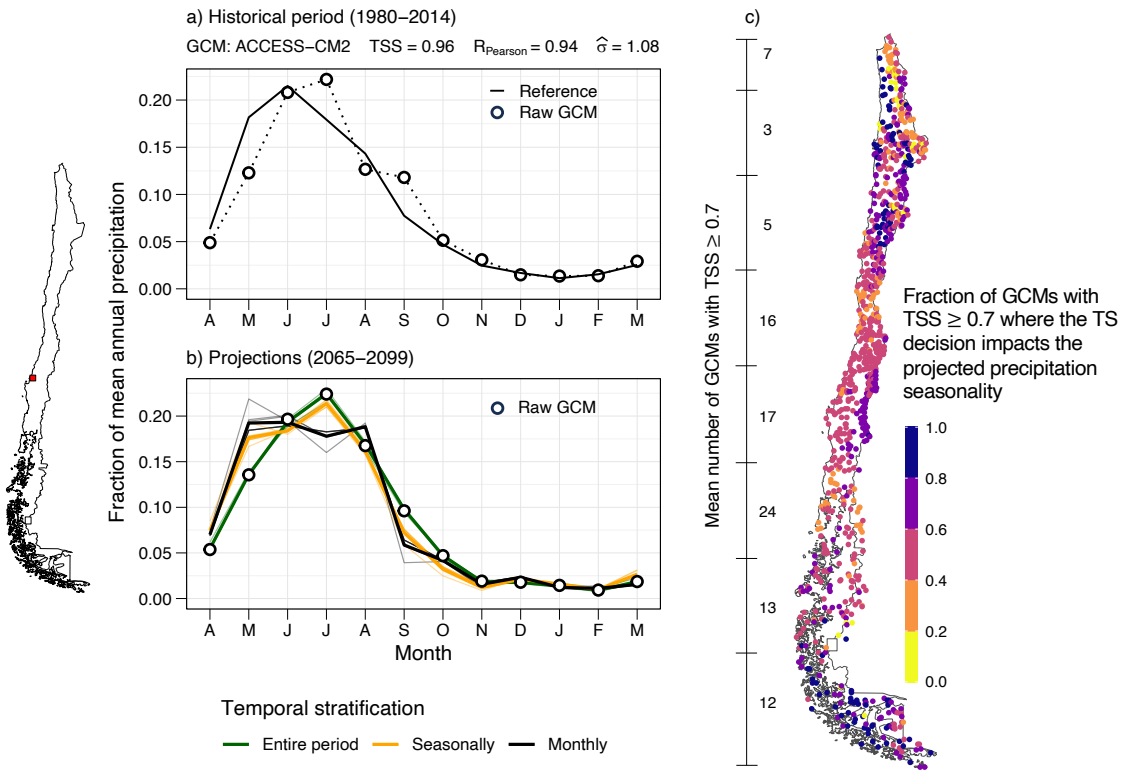

Figure 2.8: Influence of the temporal stratification used to apply bias correction methods on the projected precipitation seasonality. (a) Dimensionless historical seasonality for one grid cell (red dot on the map) and one GCM (ACCESS-CM2). Note that the sum of monthly fractions is equal to 1. (b) Projected raw (circles) and bias-corrected (colored lines) GCM precipitation seasonality. Lighter and thinner lines represent different BCMs, whereas thick lines represent the average across BCMs. (c) Fraction of the total number GCMs with $TSS \geq 0.7$, for which the temporal stratification yields different projected seasonality, measured as different months for maximum mean monthly precipitation for the 2065-2099 period. In c), the average number of GCMs meeting the TSS criterion is computed for latitudinal bands.

2.5 Discussion

The results presented here highlight the relevance of the temporal stratification used when applying bias correction techniques, which affects (i) SDBC-biases in seasonal and monthly precipitation amounts over a historical period, and (ii) the signal of projected changes and the seasonality of projections.

2.5.1 Temporal stratification and BCMs as sources of uncertainty

The results show that the choice of temporal stratification can largely affect precipitation biases during a historical period, as well as the signal and seasonality of projected changes. However, this methodological choice has rarely been explored in climate change impact assessments, and the lack of guidance has motivated the use of more than one TS in some studies (e.g., Wootten et al., 2021). Further, model errors may not necessarily be

removed after bias correction. For example, Hakala et al. (2018) obtained that biases in precipitation and streamflow seasonality remained after applying BCMs. The results suggest that only a monthly application of the BCM enables replicating the reference precipitation seasonality, even for GCMs with a good raw representation of annual cycles. On the other hand, the results differ from Aryal and Zhu (2017), who analyzed the remaining biases in drought frequency (after bias correction) obtained from four models (two RCMs and two GCMs), using three BCMs and two TSs (entire period and monthly stratification). Their results showed that the choice of TS made no difference in removing drought frequency biases or projections.

Despite some authors have explored the use of trend-preserving BCMs (e.g., Hempel et al., 2013; Pierce et al., 2015), maintaining the magnitude of mean projected changes after performing bias correction is not guaranteed (Hempel et al., 2013; Maraun, 2016; Guo et al., 2020). Such BCMs usually bias-correct the wet-day fraction (i.e., the number of rainy days) by imposing a precipitation threshold (τ) to subsequently correct the amount (e.g., Pierce et al., 2015) and/or the variability (e.g., Hempel et al., 2013) of precipitation. Thus, the trend of precipitation projections is usually preserved for amounts above τ. However, when adjusting the number of rainy days (to make it equal or comparable to the observational dataset), the raw GCM precipitation below τ is discarded. To explore the effects of selecting different τ values for each grid cell and GCM, three temporal stratifications (entire period, seasonal, and monthly) are computed. First, the observed and raw GCM precipitation amounts are re-ordered from highest to lowest and estimate the cumulative probability distribution using a Weibull probability density function; then τ is selected as the raw GCM pr value for the quantile where the first observed pr value is above 0.1 mm/d. Figure 2.9 shows projected precipitation changes before (i.e., using all the raw GCM data) and after (i.e., discarding raw GCM precipitation) applying the threshold τ. The projected precipitation change was also computed for the entire period, seasonal, and monthly stratifications. Note that the temporal scales to evaluate the projected change in Figure 2.9 are the same as the TSs. Differences in magnitude arise before applying any bias correction to precipitation amounts, showing a change in the projected change signal. For example, projections close to zero for April precipitation diminished, whereas the number of grid cells with positive and negative projections increases. Further, the discarded amount of precipitation (pr below τ) can reach up to 20-30% for April in central Chile (Figure A.9).

Since most of the BCMs used in this study bias-correct the wet-day fraction, particularly for a monthly TS (Figure A.9), the effect of τ is combined with the known effects of quantile-based BCMs. Hence, both decisions contribute to modifying the magnitude and signal of projected precipitation changes.

2.5.2 Projected seasonality

The study reveals that one of the main effects of selecting different TSs is the possibility of distorting the precipitation seasonality projected by raw GCM outputs. In hydrologic impact assessments, this artifact may propagate into the timing of simulated variables like snow accumulation and melting, energy fluxes, and streamflow (Meyer et al., 2019). The results show that when the raw GCM seasonality has timing errors (compared to the reference), a pronounced shift in the projected seasonality can be obtained after applying BCMs (compared to the case without bias correction). However, when the raw GCM

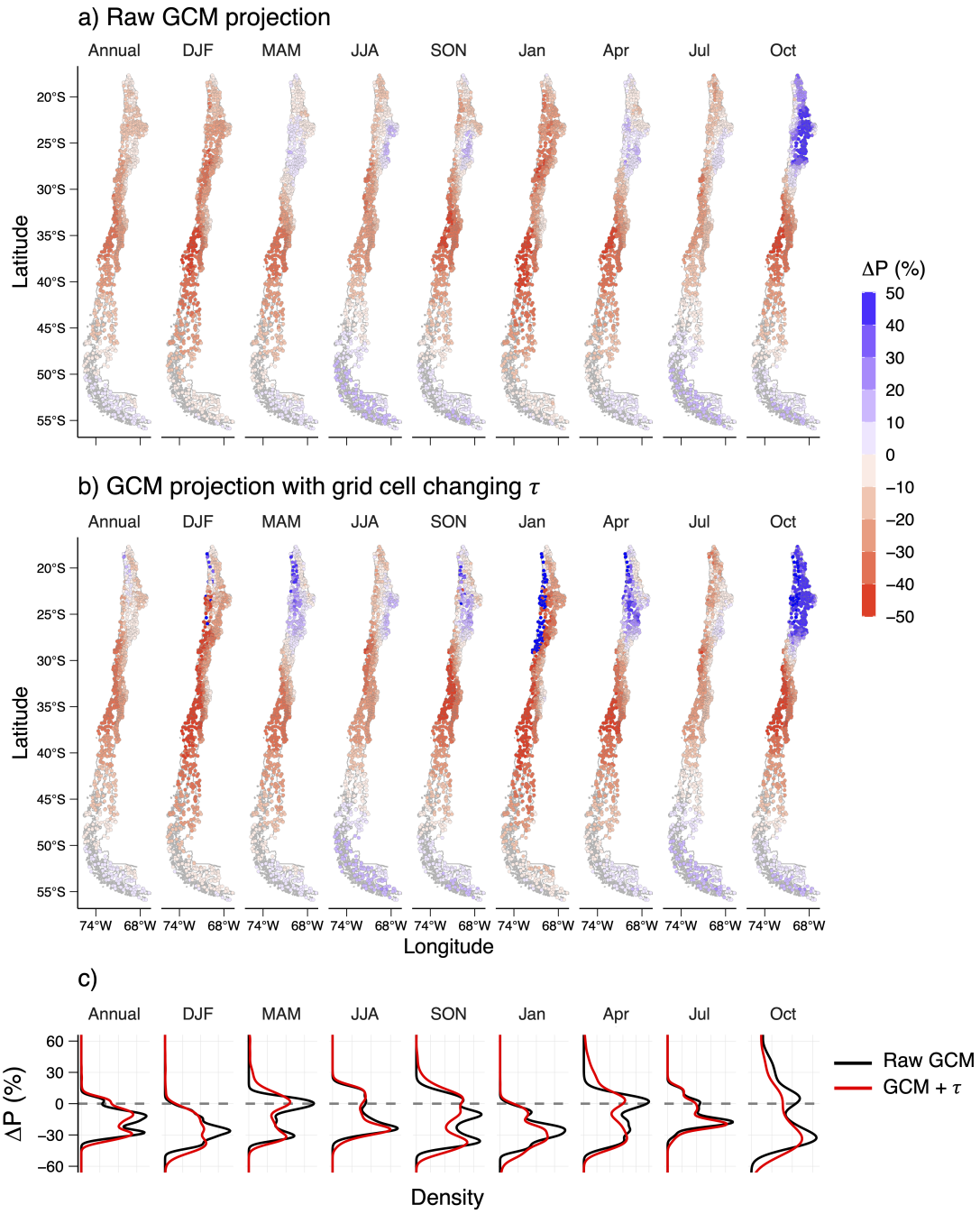

Figure 2.9: Projected precipitation change (median among models) for annual, seasonal, and monthly time scales (a) before and (b) after applying the threshold τ. c) displays the distribution of the projected precipitation change among the 1000 grid cells before (black) and after (red) applying the threshold τ. No bias correction of precipitation amounts is conducted here. Notice that the temporal scale used to compute the projected change (columns) corresponds to the temporal stratification used to obtain τ.

replicates the historically observed precipitation seasonality reasonably well, one might expect that different TSs yield the same projected seasonality. To test this hypothesis, the precipitation seasonality projected is compared for three TSs by two GCMs (CanESM5 and NorESM2-MM, Figure 2.10) that replicate annual cycles (i.e., high Pearson correlation coefficients, with GCM and reference maximum mean monthly precipitation being the same, top panels). For GCM CanESM5 (Figure 2.10a), the choice of TS has little effect

on the projected precipitation seasonality. Conversely, the temporal stratification affects the seasonality projected by NorESM2 (Figure 2.10b). For example, if the BCM is applied seasonally and monthly, the months of maximum mean monthly precipitation are May and August, respectively. Interestingly, TSS = 0.951 for this GCM, which is higher than the value obtained for CanESM5 (0.694), and both GCMs have similar Pearson correlation coefficients. These results emphasize that even GCMs with a good raw representation of historical seasonality can be affected by the temporal stratification used to apply BCMs.

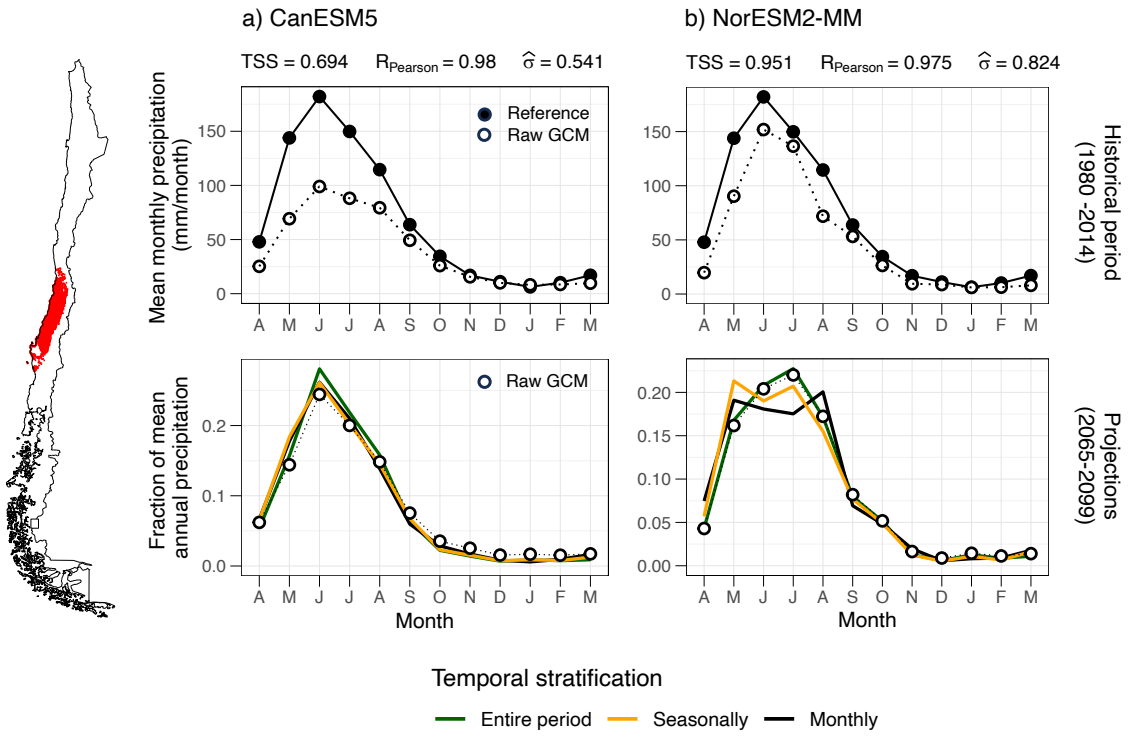

Figure 2.10: Impact of the temporal stratification used in bias correction for two GCMs. The results presented here are spatially averaged values of the grid cells contained in climate group 6 (highlighted in red on the map). Top row: comparison of the raw GCMs and the reference for the period 1980-2014. Bottom row: projected precipitation seasonality in terms of fraction of mean annual precipitation (average from the seven BCMs).

2.5.3 A priori evaluation of the TS impact on projected precipitation seasonality

Understanding the potential effects of the TS on the projected signal and seasonality of precipitation from a specific GCM could be helpful for a more detailed assessment of climate change and/or hydrological changes. Here, using the linear scaling method (LSM; Widmann et al., 2003; Maraun, 2016) is proposed - due to its low computational cost and simplicity (Lafon et al., 2013; Chaubey & Mall, 2023) -, as a quick diagnostics tool to inform if the TS may be an influential decision (an example of an LSM application is provided in Appendix A.4). The LSM removes the bias from the raw GCM time series through a multiplicative factor (f_{bias}) for the case of precipitation and an additive term for temperature, using an observational dataset as a reference. For example, if the reference and raw GCM mean annual precipitation amounts are 500 mm/year and 650 mm/year,

respectively, a factor $f_{bias} = 500/650 = 0.77$ is applied to the raw GCM time series to remove the bias. Accordingly, seasonal or monthly applications of LSM require more scaling factors (Maraun et al., 2010). Hence, the raw GCM projected change (f_Δ) is preserved (at the TS time scale), since the scaling factors are typically considered to be time-invariant. Additionally, the influence of the temporal stratification and the reference dataset (in case there is more than one available) can be isolated for a specific grid cell-GCM combination.

Figure 2.11a illustrates the application of the linear scaling method (dashed lines) to the GFDL-CM4 GCM in one grid cell (red dot in map), using the entire period and stratifying the data seasonally and monthly. For this GCM-grid cell combination, $TSS = 0.72$ and $R = 0.7$, and different TSs yield different projected precipitation seasonalities when applying the LSM. Figure 2.11a shows that the precipitation factors obtained with LSM agree with the averages obtained from all (seven) bias correction methods (solid lines).

Finally, the capability of the LSM to identify the precipitation seasonality projected with different TSs correctly is examined. To this end, for each grid cell-GCM-TS combination, the precipitation seasonalities is obtained from (i) the average between the seven BCMs and (ii) the application of the LSM. If the months of the projected maximum precipitation agree, then the LSM correctly identifies the seasonality, and if this occurs for the three TSs, then LSM successfully identifies the projected bias-corrected seasonality for that specific grid cell-GCM combination. Figure 2.11a illustrates a successful case since, for each TS, the month of maximum precipitation is the same for the average among seven BCMs and from the LSM. Then, for the 1,000 grid cells analyzed here, the fraction of GCMs for which the LSM successfully identifies the projected seasonality (accuracy, Figure 2.11b) is computed. The results show that, in almost all the grid cells, the LSM successfully identifies the projected seasonality of $\sim$70% of the GCMs, whereas, for most grid cells ($> 85\%$), the LSM successfully projects the seasonality for more than 85% of the GCMs.

2.5.4 Limitations and future work

In this study, the SSP5-8.5 scenario and 29 GCMs are selected, although other future scenarios and/or a subset of GCMs could be considered to assess the effects on historical biases (after bias correction) and/or future projections. This chapter did not focus on performance metrics for specific BCMs because evaluating the adequacy of particular bias correction methods is out of the scope of this work, and separate training and evaluation periods would be required; instead, it focuses on how these techniques are traditionally applied and how they impact model biases during the training period. Although univariate and multivariate BCMs were selected (e.g., Guo et al., 2020), quantile-based, neural networks, and linear regressions, different approaches could be considered, including trend-preserving BCMs.

Additionally, hydrological modeling was not included in this chapter. Instead, the focus is on the repercussions of some decisions on the historical biases and the projected seasonality of climate variables required to run hydrological and land surface models. However, previous work has shown that hydrological models tend to amplify biases in the forcings (Teng et al., 2015). It should be emphasized that any assessment of climate change impacts should ensure that the climatological annual cycles of hydrological simulations

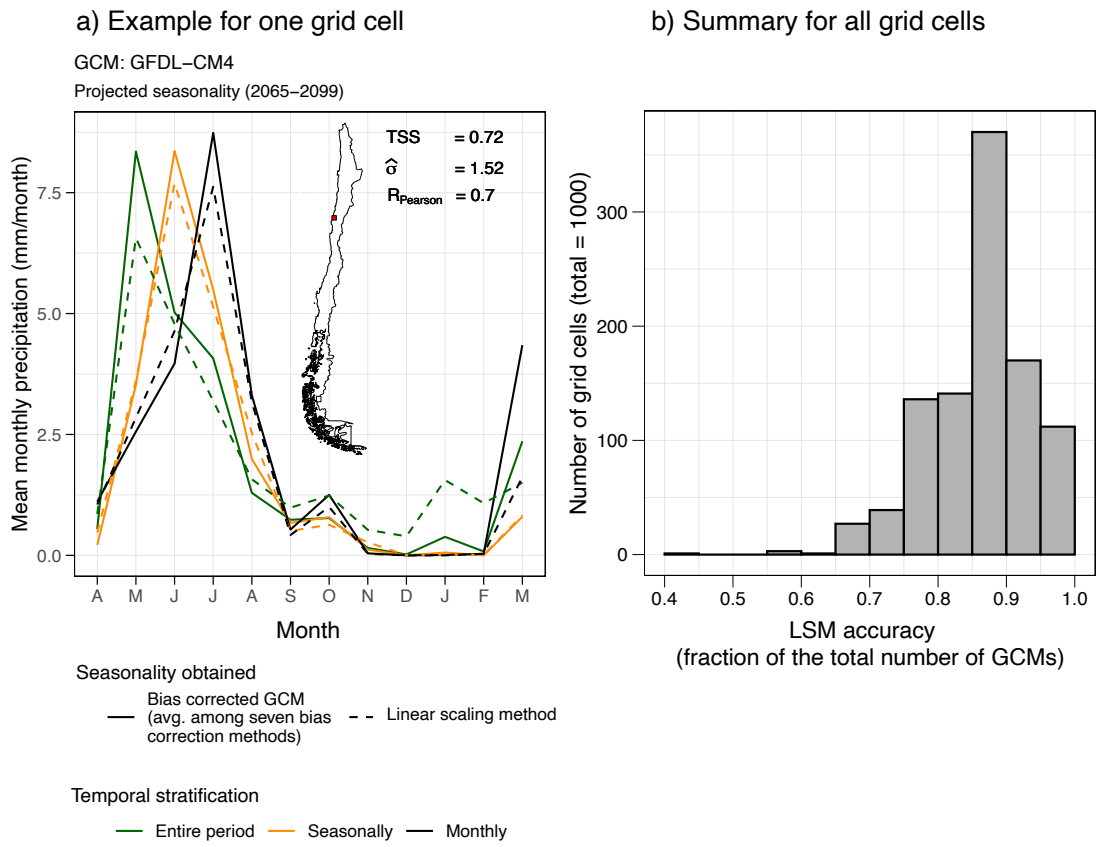

Figure 2.11: Linear scaling method used as a proxy to estimate the projected precipitation seasonality. (a) Example of projected precipitation seasonalities for one grid cell and one GCM, obtained by applying the LSM and the seven BCMs tested. The metrics summarize the raw (biased) GCM performance for the historical period (1980-2014). (b) LSM accuracy (as a fraction of the total number of GCMs) for all grid cells.

forced with (i) reference data sets and (ii) bias-corrected time series from GCMs/RCMs are similar (Hakala et al., 2018). Hence, verifying the reference and bias-corrected GCM forcing data during a historical period arises as a crucial step (Chen et al., 2013; Clark et al., 2016; Mendoza et al., 2016; L. A. Melsen et al., 2019).
Further, the effects of selecting different TSs, BCMs, and GCMs on simulated feedback mechanisms, extreme events, and spatial correlations were not evaluated.
Future work could consider the impacts of SDBC historical biases and differences in projected seasonality on different aspects of the hydrograph (e.g., mean values, extremes, timing, etc.) and signatures formulated from other variables than streamflow (e.g., SWE, soil moisture; McMillan et al., 2022; Araki et al., 2022).

2.6 Conclusions

This chapter contributes to the hydroclimate community by investigating how methodological choices in GCM bias correction affect portrayals of historical and future climates. To this end, seven bias correction methods, 29 CMIP6 GCMs, and three temporal stratifications were used. All the configurations were applied to daily time series of precipitation

and maximum and minimum daily temperature derived from the CR2MET gridded observational product, available for continental Chile. The main findings are as follows:

1. A monthly application of bias correction methods is required to replicate the reference precipitation seasonality, even for GCMs with good raw seasonality.
2. The temporal stratification is the most relevant decision to explain seasonal and monthly precipitation biases after bias correction.
3. Different temporal stratifications may yield different projected signals and seasonality, even for GCMs with good raw seasonality.
4. Since the application of different temporal stratifications in GCM bias correction affects the projected change signal and seasonality, the linear scaling method can be used to estimate such impacts beforehand and, therefore, identify the climate models for which the choice of temporal stratification may be critical. This procedure is particularly recommended if climate impact assessments involve more sophisticated and computationally expensive bias correction methods, making it infeasible to explore different TSs.
5. BCMs that bias-correct the number of rainy days using precipitation thresholds may dismiss a relevant fraction of historical and projected precipitation from raw GCMs in some areas and, therefore, modify the signal and magnitude of changes projected by raw GCMs (i.e., before bias-correcting precipitation amounts).

Chapter 3

Robust Spatial Changes in Climate Classes: Insights from Bias-Corrected CMIP6 Models across Chile

Summary

The climate in continental Chile is marked by strong latitudinal and elevation heterogeneities, exacerbated by diverse geographical features, such as the Andes Cordillera. Despite previous studies projecting warmer and dryer conditions for most of the territory, there is concern about the robustness (i.e., level of agreement among models) of changes projected for its magnitude, not only for the impact in climate indices across this domain but also to identify changes in the spatial distribution of climate classes. Hence, daily CMIP6 model outputs for continental Chile are statistically downscaled and bias-corrected using a multivariate bias correction method, to project climate changes under the SSP5-8.5 scenario. The results reveal that Global Climate Models (GCMs) project increased dryness across the study domain, especially in central Chile (-30%), with notable sensitivities of precipitation projections to the implementation of bias correction methods in the northern and austral macrozones. Temperature projections show less dispersion, with higher increments in northern Chile and the Andes (4-5°C). Notable changes in the spatial distribution of Köppen-Geiger climate classes are projected for the next decades, with the expansion of deserts in northern Chile and the prevalence of temperate climates with dry summers in central Chile. The Andes subdomain is expected to face the most dramatic changes in Köppen-Geiger classes. Surprisingly, despite the large spread in GCM projections, there is high agreement among models regarding spatial changes in climate classes. Additionally, the results suggest drastic reductions in snowfall across the Andes, with higher freezing level heights that may exacerbate flooding and landslide risk across the country[1].

[1]This chapter is currently under review in the Environmental Research Letters journal.

3.1 Introduction

Climate change threatens the environment and increases risks for human societies due to changes in the spatial and temporal patterns of climate seasonality and mean conditions, interannual variability, climate extremes, and cryosphere conditions, among other impacts on the biosphere and the hydrological cycle (IPCC, 2023). Chile, located in the south-western part of South America, will be particularly affected due to the dependence on snow and ice melt (Dussaillant et al., 2019; Barria et al., 2019), the projected reduction in precipitation (Ortega et al., 2021; Salazar et al., 2023) and increase droughts frequency and intensity (Ukkola et al., 2020). Moreover, continental Chile contains unique environments and a myriad of climatic regimes that are highly vulnerable to global warming (Heusser, 1974; Muñoz-Sáez et al., 2021; Eshel et al., 2021; Frêne et al., 2023; Marquet et al., 2023).

Over the past two decades, General Circulation Models (GCMs) have provided valuable information and insights about future climate change in Chile. Most studies have projected a general drying and warming for the country (CONAMA, 2006; MMA, 2016; Araya-Osses et al., 2020; Cortina & Madeira, 2023; Salazar et al., 2023; DGA, 2022), with considerable effects on glaciers (Carrasco et al., 2005; Vuille et al., 2015; Ayala et al., 2020; Mardones & Garreaud, 2020), snow processes (Vicuña et al., 2013; Demaria et al., 2013; Cordero et al., 2019; Aguayo et al., 2021), land cover (Bambach et al., 2013; Urrutia-Jalabert et al., 2018), socio-economic activities (e.g., Madeira, 2022) and water supply (e.g., Barría et al., 2021). In this regard, the last phase of the Coupled Model Intercomparison Project (CMIP6; O'Neill et al., 2016) offers a new scenario framework based on shared socio-economic pathways (SSP Riahi et al., 2017) and new capabilities to represent different climate features with respect to previous CMIP phases (Catalano et al., 2020; Gao et al., 2021; G. P. Taylor et al., 2023). However, such improvements are still insufficient to represent orographic effects associated with complex topography, such as the Andes Cordillera (Bozkurt et al., 2019). Moreover, assessing climate change impacts at the local scale requires downscaling GCM outputs to overcome horizontal resolution mismatches.

Since GCMs are among the main contributors to the spread of projected changes (e.g., Hagemann et al., 2011; Hattermann et al., 2018; Chegwidden et al., 2019; Vogel et al., 2023) and the small number of Regional Climate Models (RCMs) is restrictive to explore a wide range of future projections, several studies have relied on statistical downscaling (SD) and bias correction (BC) of GCMs due to their lower computational costs (H. J. Fowler et al., 2007; Wilby & Dessai, 2010; Vicuña et al., 2012; Gutmann et al., 2014; Pierce et al., 2015; Clark et al., 2016; Gutiérrez et al., 2019; Matiu & Hanzer, 2022; Alder & Hostetler, 2019; Vrac et al., 2022). Particularly, the latter approach has been used in continental Chile mostly for climate (e.g., Bambach et al., 2013; Boisier et al., 2018; Araya-Osses et al., 2020; Ortega et al., 2021; Salazar et al., 2023), hydrological (e.g., Vicuña et al., 2012; Demaria et al., 2013; Bozkurt et al., 2017; Ayala et al., 2020; Aguayo et al., 2019, 2021; Martínez-Retureta et al., 2021), land cover (e.g., Bambach et al., 2013) and water availability (e.g., Barría et al., 2021) projections.

Recently, Ortega et al. (2021) used precipitation and temperature data to evaluate the historical performance of 33 CMIP5 models for South America, while Salazar et al. (2023) also used precipitation (Pr) and temperature (T) data to evaluate the historical performance of 36 CMIP6 models and their projections under scenarios SSP1-2.6, SSP2-4.5,

SSP3-7.0 and SSP5-8.5 over continental Chile. Similarly, Gateño et al. (2024) proposed a framework for screening models, and applied it to 27 CMIP6 GCMs to produce climate projections for continental Chile. Furthermore, Beck et al. (2023) obtained 1-km worldwide maps with the spatial distribution of Köppen-Geiger climate classification from 1901 to 2099 (using 30-year intervals), showing that arid and temperate climates will dominate north-central (17°S-33°S) and central-southern (33°S-57°S) Chile, respectively, under scenario SSP5-8.5, being the Patagonian Ice Fields the exception. However, all previous studies used raw GCMs data and/or univariate bias correction methods (BCMs), which may be inappropriate for semi-arid snow-influenced environments such as northern and central Chile (Meyer et al., 2019; Guo et al., 2020) due to raw temperature biases, the coarse horizontal resolution or the lack of inter-variable dependencies when bias-correcting GCM outputs. Further, the choice of BCM can impact the signal and magnitude of projected changes (Hagemann et al., 2011; Maurer & Pierce, 2014; Wootten et al., 2021), affecting the level of agreement among models. Hence, the following questions are addressed:

- What is the level of (dis)agreement among CMIP6 climate models in terms of future projections before and after bias correction?
- Given the importance of snowpack for the hydroclimate of Chile, what are the projected changes for snowfall and freezing level heights?
- What is the level of agreement among GCMs in the changes in the projected spatial distribution of Köppen-Geiger climate classes?

3.2 Material & methods

3.2.1 Study area

The study domain is continental Chile (Figure 3.1). In the northern area (17°S-25°S), two main climate zones can be identified: (i) the arid coastal area, with low annual precipitation (<50 mm/year), and (ii) the Altiplano region, with lower temperatures due to increasing altitude and larger annual precipitation amounts (∼200 mm/year) mainly explained by the South American Monsoon during the austral summer (Aceituno et al., 2021; R. Garreaud et al., 2003). As the annual precipitation increases towards the south, the Andes Cordillera generates a west-east gradient that yields larger precipitation amounts and lower temperatures on the western slopes of the Andes Cordillera compared to the valleys and the coast (R. Garreaud et al., 2009; Viale et al., 2019). South from 45°S, a west-to-east precipitation gradient produces high precipitation amounts on the coast (>2,500 mm/year), whereas a dry climate develops in Patagonia a few kilometers to the east, with decreasing precipitation amounts. For a more comprehensive review of the climate and weather of Chile, readers are referred to Aceituno et al. (2021).

3.2.2 Datasets

The CR2MET (v2.5) database is considered as the observational product (Boisier et al., 2018; DGA, 2022; Boisier, 2023). CR2MET provides daily time series of Pr, maximum

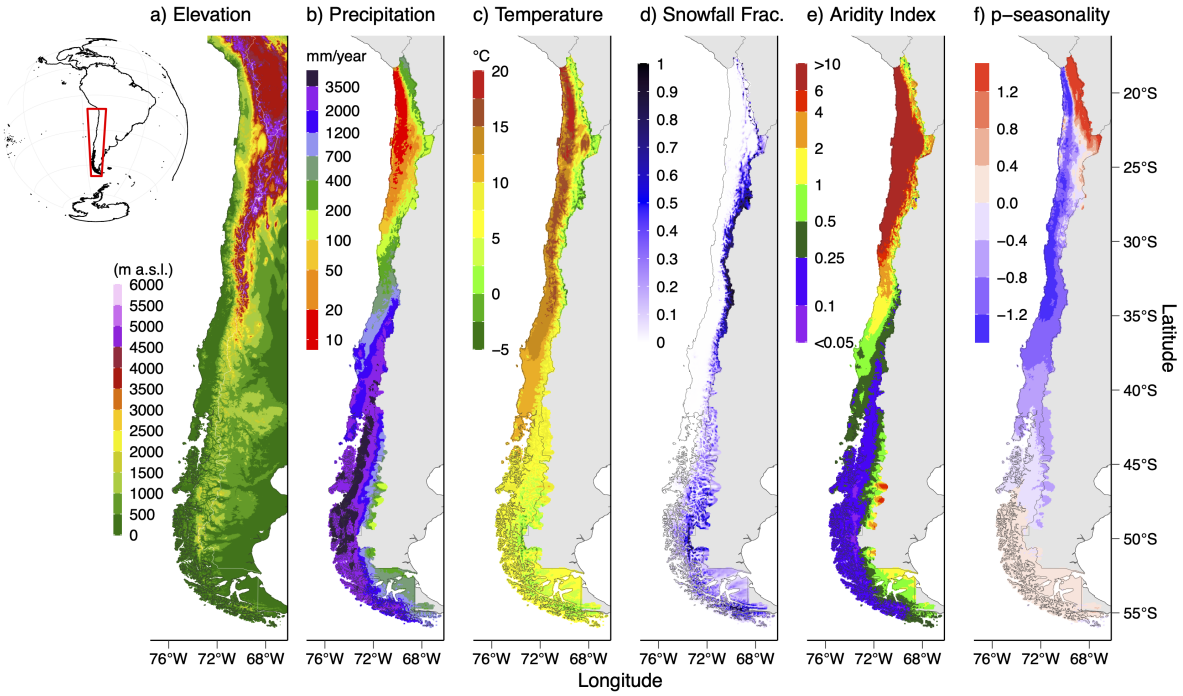

Figure 3.1: (a) Elevation and main climatic attributes for continental Chile (period 1980-2014): (b) precipitation, (c) temperature, (d) snowfall fraction, (e) aridity index, and (f) p-seasonality. The drainage contributing areas beyond the country's boundaries are also included.

(T_{max}) and minimum (T_{min}) temperatures, spanning continental Chile with a horizontal resolution of 0.05° for the period 1979-2020. Precipitation was estimated using geomorphologic attributes and ERA5 outputs (Hersbach et al., 2020) as predictors for logistic regression models that yield a probability of precipitation occurrence and multiple linear regression (MLR) models to compute precipitation amounts. Extreme daily temperatures were also obtained through MLR, with the same predictors of Pr but adding land surface temperature from MODIS (Wan et al., 1999) as a predictor.

For climate projections, 29 CMIP6 GCMs (Table A.1) are selected for the emission scenario SSP5-8.5. Additionally, one RCM (∼10-km horizontal resolution) forced with the CMIP5 model MPI-ESM-MR (CR2, 2018) is included for the emission scenario RCP8.5 (RegCM4-10k). The GCMs and the RCM are referred hereafter as models for simplicity.

3.2.3 Downscaling and bias correction

All the models were statistically downscaled to the CR2MET grid (0.05°) using the inverse distance weighting interpolation method (e.g., Demaria et al., 2013). Then, SD-GCM outputs were bias-corrected at each CR2MET grid cell using the Multivariate quantile-mapping bias correction method (MBCn; Cannon, 2018), available in the "MBC" library (Cannon, 2016, 2018), and implemented in the statistical software R (R Core Team, 2023). This method was selected because it better replicates the precipitation-temperature correlation in arid or temperate areas with considerable snowfall fractions (Meyer et al., 2019; Guo et al., 2020), such as northern and central Chilean Andes. The MBCn method relies on an orthogonal rotation followed by applying the Quantile Delta Mapping method

(Cannon et al., 2015) and, finally, applying an inverse matrix (the one used to compute the orthogonal variables). This process is repeated until convergence. For more details about MBCn, readers are referred to Cannon (2018). The MBCn is applied to the daily time series of Pr, T_{max}, and T_{min} using a monthly temporal stratification and considering 1980-2014 as the training period (35 years). As a result, we obtain daily bias-corrected time series of Pr, T_{max}, and T_{min} are obtained for the 1980-2100 period. Finally, mean daily temperatures (T_{mean}) are estimated as $(T_{max} + T_{min})/2$ (e.g.; Demaria et al., 2013).

3.2.4 Climate Indices

In this chapter the following climate indices are considered based on their capability to inform observed hydrological responses (e.g., Gutmann et al., 2014): (i) precipitation, (ii) wet-day fraction (WDF), (iii) wet and dry-spell lengths (WSL and DSL), (iv) fraction of precipitation falling as snow (S_f/P), and (vi) temperature. WSL and DSL are the mean consecutive rainy and non-rainy days, respectively, whereas WDF is the mean fraction of rainy days above 0.1 mm/d (e.g., Musselman et al., 2017). All indices are computed at the annual, seasonal, and monthly time scales. Annual snowfall (S_f) is calculated as the sum of precipitation for days with $T_{mean} \leq 2$°C (Jennings et al., 2018). Relative changes are obtained by comparing the future period 2065-2099 against the historical period 1980-2014. Biases are drastically reduced after applying the MBCn method (not shown).

3.2.5 Climate Classification

To evaluate projected changes in the spatial distribution of Köppen-Geiger climates across Chile, the climate classification criteria described by Peel et al. (2007) and Beck et al. (2018) is followed to produce, with each model, maps with climate classes for the historical (1980-2014) and future (2065-2099) periods. Then, the level of agreement among models is evaluated for the 2065-2099 period as the fraction of the number of models projecting the same climate classification.

3.2.6 Basin Scale Analysis

Precipitation and snowfall projections are spatially aggregated over 31 relevant basins in Chile, using the catchment delineation produced by the Chilean Water Directorate (DGA in Spanish). In these basins, projected changes in freezing level height (FLH) are computed during fall-winter storms covering at least 10% of the basin area with a spatial average precipitation rate $\geq$ 5 mm/d (Covián & Stowhas, 2015; Mardones & Garreaud, 2020). Here, FLH is considered as the altitude of isotherm 0°C, estimated by fitting a local linear regression between each grid cell elevation and the mean daily temperature for each storm.

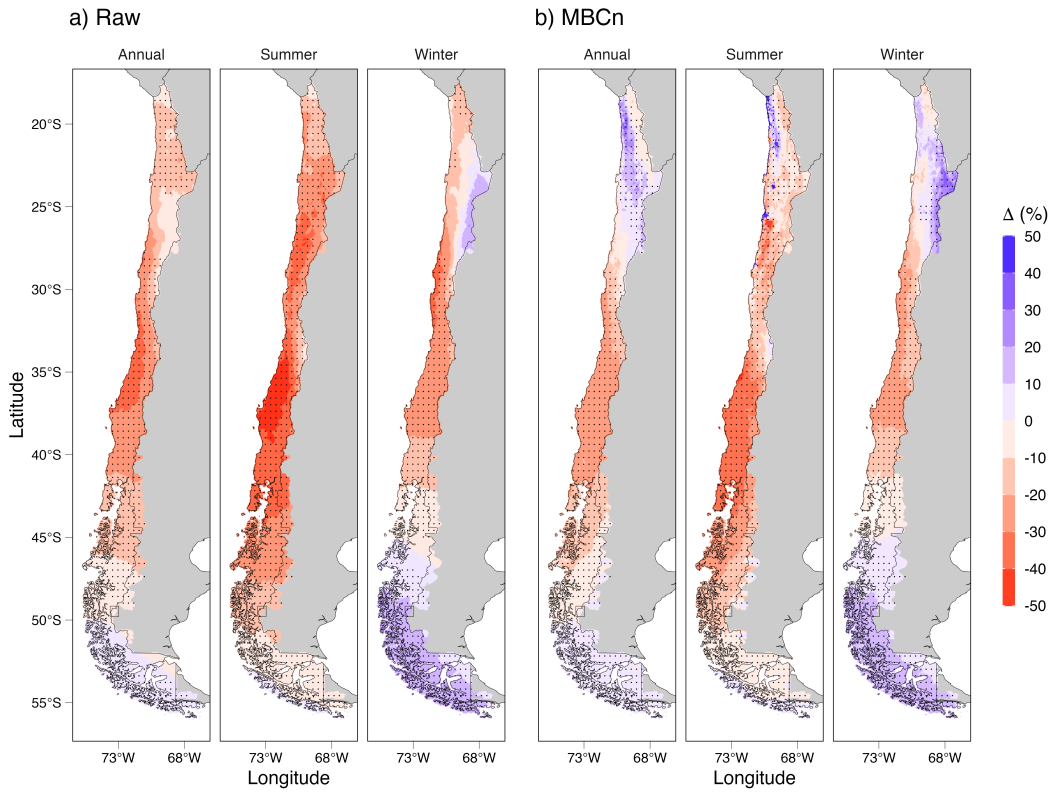

Figure 3.2: Changes in precipitation (median among models) at annual and seasonal (summer and winter) scales for the period 2065-2099 with respect to the period 1980-2014. Results are displayed (a) before (raw) and (b) after (MBCn) when the bias correction method is applied. Dotted areas represent inter-model agreement $\geq$75% in projected signal changes.

3.3 Results

3.3.1 Projected changes in precipitation and temperature

At the annual scale, MBCn modifies the signal of projected changes in the super-arid regions (17°S-30°S), where annual amounts are $\lesssim$40 mm/year (Figure 3.2). In the remaining areas, the signal of projected annual changes is preserved after bias correction, with spatial variations in magnitude, especially in central Chile (30°S-40°S). Similar spatial precipitation patterns of negative projected trends are obtained for the summer season. For winter precipitation, MBCn replicates the signal of raw projected changes for almost the entire study domain, with differences in the level of agreement (dotted areas in Figure 3.2) in northern Chile (20°S-27°S). Overall, a high inter-model agreement ($\geq$ 75%) is achieved for most of the domain before and after bias correction.

Regarding projected changes in temperature and snowfall, a north-south and east-west gradient in temperature increase (Figure 3.3a) is present, with the Andes Cordillera and northern Chile being the most affected areas. The general decrease in annual precipitation projected for most of continental Chile (Figure 3.2), combined with temperature increments (Figure 3.3a), yields a general decrease in snowfall amounts (Figure 3.3c). Moreover, snowfall is projected to decrease considerably ($< -50\%$) in low-elevation areas (see Figure 3.1), whereas smaller reductions are projected for higher altitudes of the

western slopes of the Andes Cordillera between 25°S and 35°S, where more than half of the Chilean population lives, and snowmelt is relevant for water supply during the spring-summer (dry) season. Relatively smaller snowfall changes are expected for the Ice Fields of Patagonia (south of 45°S) compared to lower latitudes, along with a future reduction in snowfall fraction. Snowfall is projected to remain the main form of precipitation only over the Andes between 30°S-35°S.

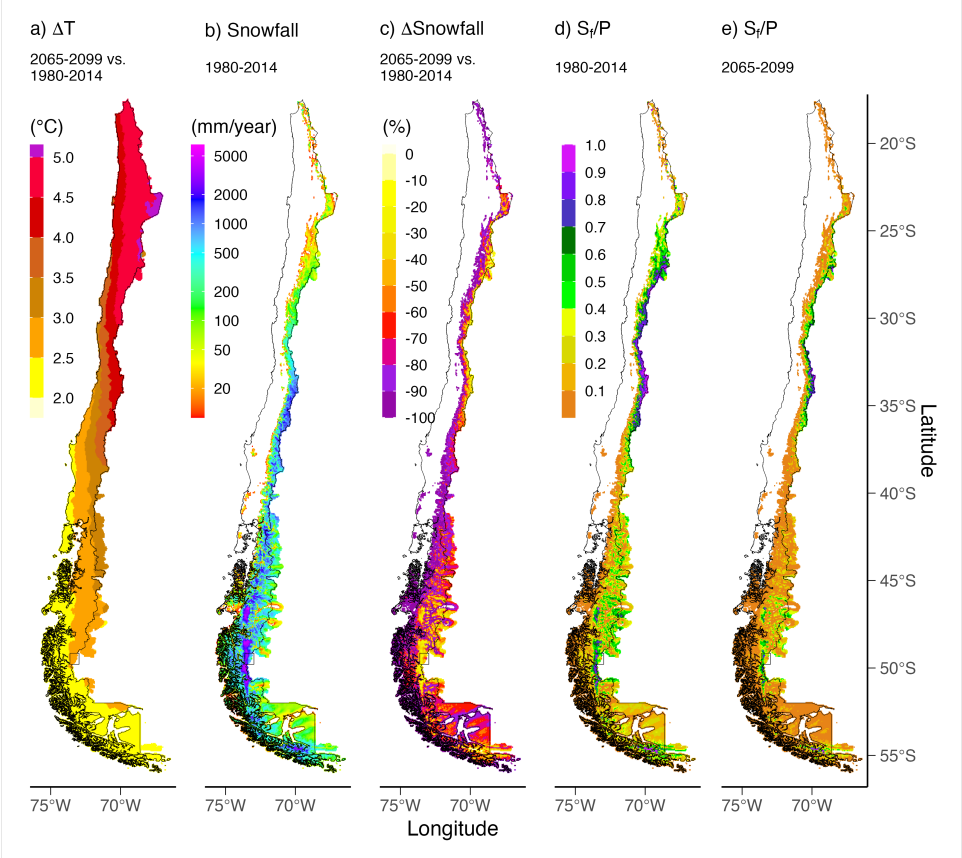

Figure 3.3: Projected changes for (a) mean annual temperature, (c) annual snowfall, and (e) snowfall fraction for the period 2065-2099 (relative to 1980-2014) after applying MBCn. The historical annual snowfall amount (b) and the historical snowfall fraction (d) are included for completeness. Values correspond to the median projected change among models.

3.3.2 Projected spatial distribution of climate classes

This section contrasts the spatial distribution of Köppen-Geiger climate classes during the historical reference period (Figure 3.4a) against the distribution of the future projection (Figure 3.4b). The climate for the 2065-2099 period in northern and central Chile would be dominated by classes BWh (Dry - Arid desert - Hot summer) and Csa (Temperate - Dry and hot summer), in agreement with reductions in annual precipitation and temperature increments projected for those regions (Figures 3.2 and 3.3). Despite the dispersion in the magnitude of P projections, Figure 3.4c reveals a high inter-model agreement in terms of projected climate classes. However, disagreement is present in the transition between climate zones and between 30°S and 35°S. Figure 3.4d shows that a change in the main climate classes (i.e., Tropical, Dry, Temperate, Cold, and Polar) is projected for the Altiplano region (north of 20°S), the western slopes of the Andes Cordillera between 25°S

and 40°S, and the islands and valleys of Patagonia (south of 42°S). The climate in the valleys between 30°S and 34°S is expected to evolve from a temperate to an arid climate, with the lowest inter-model agreement of future climate classes.

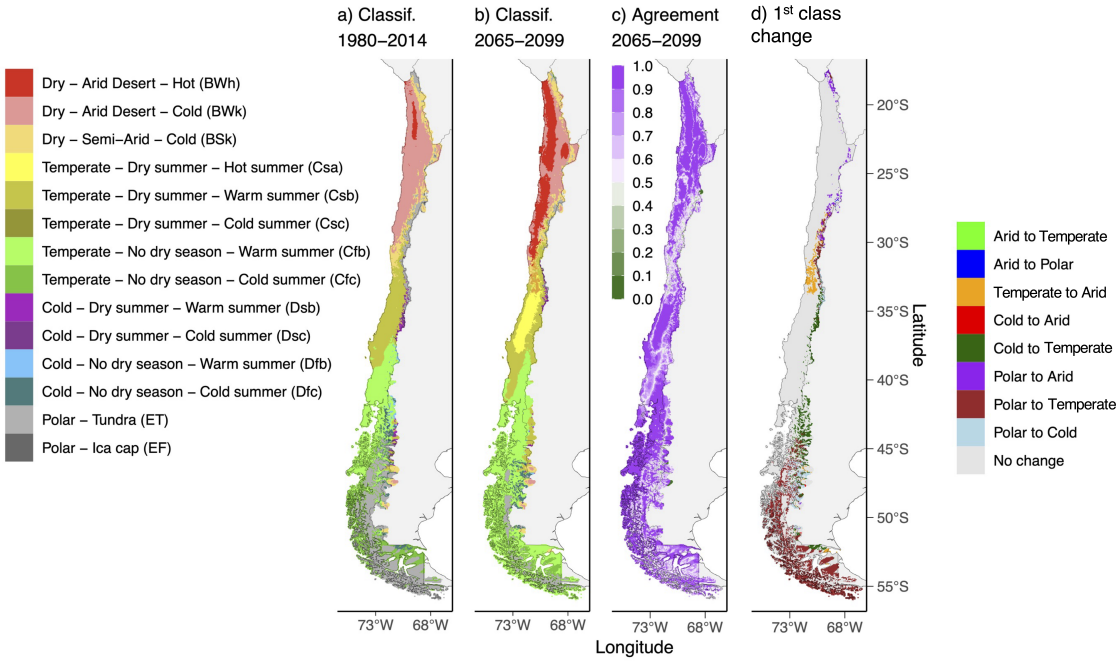

Figure 3.4: Climate classification (mode among models) during historical (a; 1980-2014) and future (b; 2065-2099) periods. c) Agreement (as a fraction) among the 30 models in terms of climate classes projected for the period 2065-2099. d) Projected changes in the main Köppen-Geiger climate classes. The color scale for climate classification is the same as in Beck et al. (2018).

3.3.3 Basin scale analysis

At the basin scale (Figure 3.5), there is a north-to-south gradient in projected precipitation changes, with a larger dispersion in the signal of changes in northern basins (Lluta-Huasco). The results show a high level of inter-model agreement ($> 75\%$) in the signal of precipitation changes between the Elqui (30°S) and the Baker River basins (47°S). Regarding snowfall, more drastic reductions are projected for basins with low snowfall values, with small or near zero future amounts, such as Lluta-Loa and the Altiplano region. In contrast, smaller reductions are projected in central (Elqui-Maule River basins) and austral Chile (Palena-Pascua River basins). Despite the reductions in S_f, snow would continue playing a key (although diminished) role in relevant basins. For example, from Copiapó to Limarí (northern Chile), Aconcagua to Mataquito (central Chile), and Aysén to Pascua River Basins, S_f/P (annual scale) is expected to be larger than 10%, which would be crucial for water resources management during the spring-summer (dry) season.

To explore possible reasons behind the drastic reductions in S_f (Figure 3.5c), the FLHs during fall and winter (April-September) are analyzed. Figure 3.6a and 3.6b display projected changes in FLHs and the change in rainfall contributing area (i.e., the catchments' area below the FLH), respectively. A projected rise of at least ~500 m is projected for all basins, with northern Chile being the most affected area, with a rise of FLH close

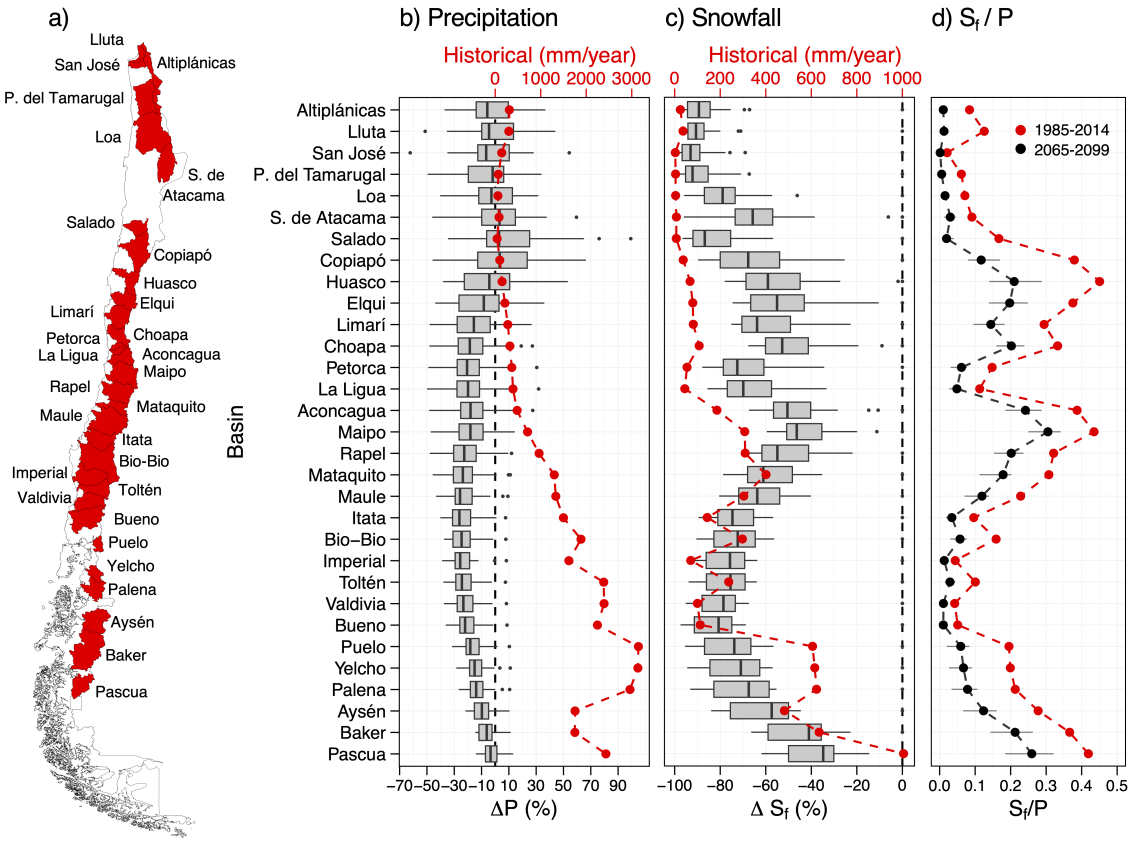

Figure 3.5: Basin-scale changes in P, S_f and S_f/P for the period 2065-2099 (with respect to 1980-2014) projected with MBCn bias-corrected models. Each boxplot comprises results from the 30 models; the boxes correspond to the interquartile range (IQR; i.e., 25th and 75th percentiles), the horizontal line in each box is the median, whiskers extend to the ±1.5IQR of the ensemble, and the red dots represent the historical reference (period 1980-2014).

to 1,000 m (Lluta-Elqui River basins). Figure 3.6b reveals that the rainfall contributing areas could increase by ≥1,000 km^2 in northern (Loa-Choapa), central (Aconcagua-Bio-Bio), and austral (Palena-Pascua) basins, likely impacting the hydrological regime of these basins while increasing flood risk. Such increases in rainfall-contributing areas can represent more than 20% of the total basin area in some cases (e.g., Copiapó, Huasco, and Elqui in Figure 3.6c).

3.4 Discussion

3.4.1 Projected changes in precipitation and temperature

The results for the Altiplano differ from CONAMA (2006), since they projected an increment of summer and spring precipitation for scenario A2, whereas this study projects a decline for the same seasons before and after applying bias correction, which aligns better with the dynamical downscaling conducted by CR2 (2018) for the scenario RCP8.5. Further CONAMA (2006) projected a general annual precipitation reduction around ∼ 40% for central Chile, while the results of this study project a decrease between 20-30% after

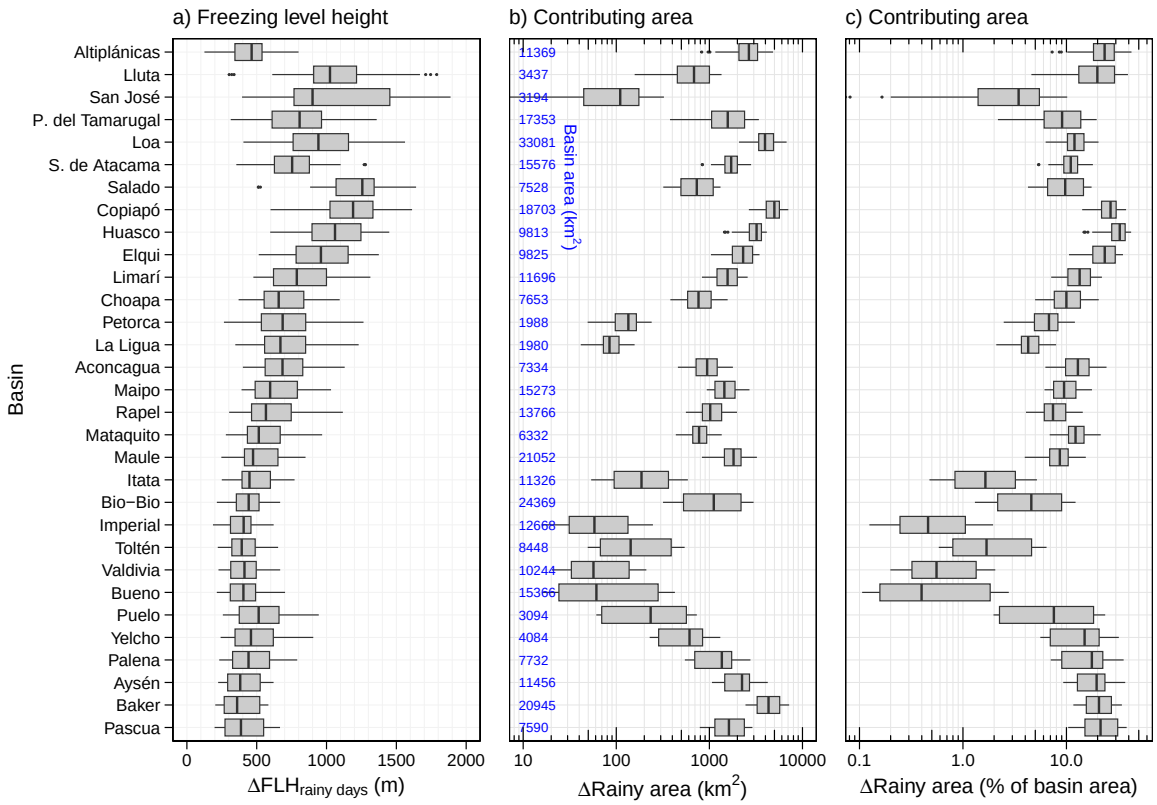

Figure 3.6: Projected changes in freezing level heights and rainfall contributing area during rainy days for the period 2065-2099 (with respect to 1980-2014), using MBCn bias-corrected models. Each boxplot comprises results from the 30 models; the boxes correspond to the interquartile range (IQR; i.e., 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles), the horizontal line in each box is the median, whiskers extend to the ±1.5IQR of the ensemble. Notice that the x-axis in panels b) and c) are on a logarithm scale.

applying bias correction. Such differences may be explained by the use of different GCMs or types of downscaling methods: here, a statistical downscaling is used, while CONAMA (2006) performed a dynamical downscaling of only one GCM and two CMIP3 scenarios (A2 and B2).

Salazar et al. (2023) projected, using raw GCM outputs under the SSP5-8.5 scenario, a general drying for the country, with precipitation declines up to 40% (contrasting the periods 2080-2099 and 1986-2014) in central Chile (30°S-40°S), and slight projected increments in the northern Andes (24°S-29°S) and Patagonia (south of 50°S). For the same scenario and future period, they projected a ∼5°C temperature increment in northern Chile (18°S-28°S) - especially in the Altiplano region -, temperature increments up to ∼2-3°C for Patagonia, and higher temperature increments in the Andes Cordillera compared to the coast and valleys. Such spatial patterns align with this work since a similar number of CMIP6 models (33 vs. 29) is used.

3.4.2 Spatial distribution of climate classes

Despite the large dispersion in future projections of Pr and T_{mean} (Figure 3.5), a high (≥0.7) inter-model agreement is found in the projected spatial distribution of Köppen-

Geiger climate classes. The historical and projected spatial distribution of these classes (computed as the mode among models) aligns well with Beck et al. (2018), though there are some differences in northern and central Chile during the historical period and in projected climate classes for Patagonia. This could be explained by differences in the periods used to obtain the spatial distribution of climate zones (1980-2016 vs. 1980-2014 as historical periods; 2065-2099 vs. 2070-2100 for projections), the choice of climate models, and the choice of observational datasets. In this regard, Beck et al. (2018) applied the projected changes in Pr and T_{mean} from models and applied them to their observational dataset. More recently, Bambach et al. (2022) obtained results similar to the results obtained in this work in northern Chile for the historical period using CESM model simulations. For projections, the results of this work align well for most of the domain, except for the projected climate of southern Chile (40°S-45°S).

3.4.3 Annual snowfall and freezing level heights

The results for projected snowfall changes align well with projections reported by Bambach et al. (2022) for continental Chile: most of the future snowfall would be concentrated in the Andes of central Chile (27°S-36°S) and the Patagonia Ice Fields (47°S-52°S). However, snowfall projections vary considerably among models since they are intrinsically related to changes in Pr and T_{mean} and, therefore, to the main large-scale patterns simulated by climate models (Cordero et al., 2019). At the catchment scale, the results of this work project a decline in annual precipitation for the Limarí River basin, ranging from 4-28%, compared to the 15-35% precipitation reduction and 3-4°C temperature increment projected by Vicuña et al. (2011) under scenarios B2 and A2. Vicuña et al. (2011) also projected a transition from a snowmelt-driven to a rainfall-driven hydrological regime due to declines in snowfall, while the results of this work project a ΔS_f ranging 51-70% in the same basin. Demaria et al. (2013) projected a precipitation change ranging between -19 and -21% (under higher emission scenarios A2 and RCP8.5) and an FLH rise of 700 m for the Mataquito River basin, whereas this study obtained, for the same catchment, a ΔFLH range (from the median values from each model) of 280-968 m, with an Interquartile Range (IQR) of 433-668 m. Mardones and Garreaud (2020) evaluated the tropospheric FLH change during rainy and non-rainy days between 30°S and 38°S (Huasco to Itata River basins). For this latitudinal range, they obtained an increment of ∼400 m under scenario RCP8.5, which is in the lower range of the projection estimates of this work. However, care should be taken when comparing both results since this work considered a topographic gradient, while Mardones and Garreaud (2020) considered a free tropospheric gradient.

The increasing FLHs during storms, along with projected reductions in precipitation, are expected to yield drastic reductions in snow accumulation over the study domain, likely affecting (i) hydrological regimes and water availability during the dry season (spring and summer) in snowmelt-driven catchments; and (ii) rainfall contributing area to runoff during rainy days, increasing flood risk (Mardones & Garreaud, 2020). For example, the results of this work project an increase of ∼1000-2000 km^2 in the contributing area of the Maipo River basin (Figure 3.6) - equivalent to ∼20-40% of the basin area contributing to the city of Santiago (∼7-million population)-. Despite these changes, snow is expected to continue playing a relevant role in the hydrological cycle in several basins in continental Chile.

3.4.4 Limitations

In this study, only one ensemble member (r1i1p1f1) per model is considered, assuming that, at the end of the 21st century, the spread arising from each model's internal variability will be smaller than the spread provided by inter-model variability (Hagemann et al., 2011; Hawkins & Sutton, 2012; Boisier et al., 2018; Hawkins et al., 2020; Jain et al., 2023). This study considers the worst emission scenario to align with the recommendations of Chilean authorities for climate change impact assessments in environmental studies (SEIA, 2023a, 2023b). Analyzing the sensitivity of the results to different emission levels is beyond the scope of this study and, therefore, proposed for future work. Moreover, this work applied statistical downscaling and bias correction, which do not necessarily maintain the physical consistency among variables in domains with complex topography and, therefore, may be missing physical processes that could impact future climate projections in the Andes. One way to address this issue would be through dynamical downscaling, which is beyond the scope of this work. Additionally, only one bias correction method was considered since bias-correcting the models using MBCn is computationally expensive, although other techniques could be explored (e.g., Kim et al., 2021; Vrac et al., 2022; Kusumastuti et al., 2022), and only one dataset (CR2MET) was considered as a reference. Since CR2MET estimates rely on meteorological gauges, higher uncertainties are expected in regions with a low gauge density, such as northern (17°S-25°S) and austral (40°S-50°S) Chile.

3.5 Conclusions

This work characterized the inter-model agreement in CMIP6 climate projections across continental Chile, with a focus on Pr, T_{mean}, and S_f, and assessed the robustness of the spatial distribution of Köppen-Geiger climate classes and the change in FLH projected for the end of the XXI century under a high-emission scenario. To this end, 30 models (29 CMIP6 GCMs and one RCM forced with a CMIP5 GCM) are used to produce climate projections considering the scenario SSP5-8.5. All the models were statistically downscaled to a 0.05° x 0.05° horizontal resolution grid and bias-corrected using the MBCn bias correction method.

Drier conditions (up to 40%) are expected for most of continental Chile, except Patagonia (south of 50°S). However, there is a high dispersion among models regarding the magnitude of the annual precipitation changes (Figure 3.5). SDBC models project drastic reductions in snowfall over most of the extratropical Andes (up to 80%), which can be explained by a reduction in annual precipitation and increasing freezing levels during rainy days (reducing snow accumulation). The basins most affected by snow accumulation reductions are in northern Chile (Salado-Limarí River basins). Further, larger rainfall-contributing areas could increase flood risk. Nevertheless, snowfall would still be relevant in central Chile and Patagonia ($S_f/P \geq 10\%$).

Regarding the projected spatial distribution of climate classes, most of northern Chile would classify as BWh (Dry - Arid desert - Hot summer) and central Chile as Csa (Temperate - dry and hot summer). The first Köppen-Geiger climate classes of coastal areas located on the coast between 18°S and 45°S are not expected to change. In Patagonia, the areas classified as Tundra would be constrained to the Ice Fields. The spatial dis-

tribution of Köppen-Geiger classes will change considerably over the Andes, particularly in the Altiplano (17°S-20°S), north and central Chile (27°S-37°S), and Patagonia (south of 43°S). Despite the large spread in precipitation and temperature projections, there is high agreement among CMIP6 models in the projected spatial distribution of climate Köppen-Geiger classes.

Open Research Section

Bias-corrected models using the MBCn method can be found at `https://doi.org/10.7910/DVN/O3YBOT`. To the best of the author's knowledge, this is the first publicly available dataset with SDBC CMIP6 model outputs with daily time series of Pr, T_{max} and T_{min} for continental Chile, which can be used for climate change impacts assessments and water resources planning.

Chapter 4

Pitfalls in Streamflow-only Calibration of Distributed Hydrological Models

Summary

Distributed hydrological models enable the characterization of spatial heterogeneities in states and fluxes, including streamflow at inner points of a basin. Despite the increasing availability of remotely sensed observations that could improve the estimation of model parameters, calibrating model parameters and their spatial distribution solely based on streamflow data collected at the catchment's outlet continues to be a common practice. This chapter examines how spatially distributing (i.e., regularizing) different parameters, each calibrated by optimizing different metrics, affects the average seasonality and spatial patterns of simulated evapotranspiration (ET), soil moisture (SM), land surface temperature (LST), and fractional snow-covered area (fSCA). To this end, calibration experiments with the Variable Infiltration Capacity (VIC) model are conducted in six basins located in continental Chile, using (i) different streamflow-based objective functions, and (ii) regularizing different parameters associated with different physical processes. For the latter step, a regularization strategy is tested based on principal component analysis of climatic and physiographic attributes of the modeling spatial units contained within each basin. The results suggest that these decisions may have large effects on the spatial representation of ET, SM, LST, and fSCA, without degrading the performance of streamflow simulations. Further, the average streamflow seasonality can be simulated reasonably well, with large biases in ET, fSCA, SM, and LST (in that order). In particular, different calibration configurations can yield the same annual streamflow cycle through very different ET seasonalities, affecting the catchment-scale seasonal water balance. Overall, the results presented here reinforce the benefits of including spatial patterns of hydrological variables in the calibration of distributed hydrological models and highlight the need to verify the seasonality of other simulated variables than streamflow.

4.1 Introduction

Distributed hydrological models are valuable tools to simulate the spatial heterogeneity of variables involved in the terrestrial water cycle (Reed et al., 2004), offering the potential to make runoff predictions in ungauged inner sub-catchments. Further, the increasing number of remotely sensed observational datasets (K. Wang & Dickinson, 2012; S.U. et al., 2014; Kinar & Pomeroy, 2015; McCabe et al., 2017) has facilitated the assessment of model fidelity through variables other than streamflow (Q), including evapotranspiration (ET; e.g., M. C. Demirel et al., 2018; Dembélé, Hrachowitz, et al., 2020), soil moisture (SM; e.g., Tong et al., 2021; Bajracharya et al., 2023), land surface temperature (LST; e.g., Zink et al., 2018), water storage variation (WS; e.g., Werth & Güntner, 2010), groundwater levels (GWL; e.g., Refsgaard & Knudsen, 1996), fractional snow-covered area (fSCA; Parajka & Blöschl, 2008; Duethmann et al., 2014; Bennett et al., 2019; Tong et al., 2021; Tang et al., 2023), and snow water equivalent (SWE; e.g., Avanzi et al., 2020). These variables can be used to evaluate the hydrological consistency of models calibrated using only streamflow data (e.g., Cuartas et al., 2012; Zhang et al., 2020; Wen et al., 2020; Odusanya et al., 2021; Shah et al., 2021), or be directly incorporated within the calibration process (e.g., Rakovec et al., 2016; Széles et al., 2020). In either case, summary performance metrics - similar to the case of streamflow - are typically formulated and evaluated (e.g., Koch et al., 2017; M. C. Demirel et al., 2018; Koch et al., 2018; Tong et al., 2021).

Despite the development of satellite remote sensing products and the progress in multivariate parameter estimation techniques (Lettenmaier et al., 2015; McCabe et al., 2017; Sheffield et al., 2018), the calibration of distributed hydrological models still relies strongly on streamflow data (e.g., Shafii & Tolson, 2015; L. Melsen et al., 2016; Mizukami et al., 2017, 2019; Beck et al., 2020; Clark et al., 2021; Aguayo et al., 2021; S. Wang et al., 2022; Cinkus et al., 2023). This is not only because accurate streamflow simulations are required for a myriad of water resources applications, but also for the adequate partition of annual precipitation into storage, evapotranspiration, and runoff (Kirchner, 2006; Troch et al., 2013; Mendoza et al., 2016; Akbar et al., 2020). In this regard, several multi-objective streamflow-based calibration approaches have been proposed and tested with distributed hydrological models for simultaneously matching different parts of the hydrograph, by incorporating, for example, hydrological signatures o streamflow transformations in the objective function (e.g., Pokhrel & Gupta, 2010; Westerberg et al., 2011; Garcia et al., 2017; McInerney et al., 2017; K. Fowler, Coxon, et al., 2018; K. Fowler, Peel, et al., 2018; Koppa et al., 2019; Todorović et al., 2022; Rakovec et al., 2019; Casper et al., 2023).

Another challenge in distributed hydrological modeling is the large number of model's spatial element such as grid cells and, therefore, the number of parameter values that need to be specified. This issue is typically addressed through spatial parameter regularization techniques (Pokhrel & Gupta, 2010; Samaniego et al., 2010; de Lavenne et al., 2019), which usually involve super-parameters or transfer functions applied under the assumption that the spatial distribution of climate and/or geomorphological attributes informs the spatial distribution of the model parameters (e.g., Mizukami et al., 2019; Beck et al., 2020). However, identifying the level of relationship between these parameters and the attributes is challenging, especially for "free" (i.e., non-physical) parameters. Moreover, tackling parameter equifinality (K. Beven & Binley, 1992; K. Beven, 2006; Khatami et al., 2019) is still an ongoing effort, which can be partially evaluated through multivariate model evaluation (e.g., M. C. Demirel et al., 2018; Dembélé, Ceperley, et al., 2020; Shah

et al., 2021; Alfieri et al., 2022; Yáñez-Morroni et al., 2023) and parameter uncertainty estimates (K. Beven, 2001; Koppa et al., 2019).

The intricate relationship between hydrologic modeling decisions and the capability to correctly simulate seasonal water balance introduces additional complexities. In this regard, the choice of soil parameter regularization technique plays a crucial role in shaping the simulated spatial distribution and overall model performance (Refsgaard & Knudsen, 1996; Reed et al., 2004; Samaniego et al., 2010; Rakovec et al., 2016; Dembélé, Ceperley, et al., 2020; M. C. Demirel et al., 2018). On the other hand, the choice of calibration metric is important since it influences the simulated hydrograph and the final selected model parameters (Kollat et al., 2012; Garcia et al., 2017; Pool et al., 2018; K. Fowler, Peel, et al., 2018; Cinkus et al., 2023). Further, the effects of these methodological decisions are typically evaluated using summary metrics (Clark et al., 2021), and limited attention has been paid to the assessment of annual cycles in water balance components (i.e., precipitation, evapotranspiration, and streamflow) during hydrologic model calibration (Khatami et al., 2019; Kirchner & Allen, 2020). Therefore, the research questions are as follows:

1. To what extent does the spatial regularization of different soil parameters and the choice of streamflow-based objective functions affect the model performance in terms of Q simulations and the spatial patterns of LST, ET, fSCA, and SM?
2. What are the tradeoffs between replicating Q annual cycles and the aim to adequately simulate LST, ET, fSCA, and SM seasonalities, and how can these be overcome?

Here, it is hypothesized that distributing soil parameters through a regularization strategy that is spatially coherent with grid cell attributes can improve the model's capability to replicate spatial patterns and seasonality of other variables besides streamflow. To test this, the Variable Infiltration Capacity model (VIC; Liang et al., 1994) is calibrated in six basins with different hydrological regimes in central Chile. This chapter uses (i) Principal Component Analysis (PCA) to derive the *a priori* spatial distribution of model parameters and (ii) different streamflow-based objective functions to obtain the model parameters. The effects of such decisions on simulated Q are evaluated, as well as ET, LST, SM, and fSCA, obtained from remote sensing products. To quantify the capability to replicate observed spatial patterns, this work uses biased and unbiased[1] metrics to summarize the overall model performance. However, instead of highlighting changes in overall metrics, the focus is on the average seasonalities of Q and ET to identify potential risks and challenges since they directly affect the water balance (Clark et al., 2021).

4.2 Study domain

This chapter considers six basins located in continental Chile that span three different hydrological regimes: snowmelt-driven, mixed, and rainfall-driven (Figure 4.1). The basins' boundaries and the streamflow gauge identification number (ID) are obtained from the

[1]Bias and unbiased metrics are performance indices that contrast observed and simulated variables with and without dimensions, respectively. Regarding the latter, unbiased metrics are useful when the reference data has biases.

CAMELS-CL dataset (Alvarez-Garreton et al., 2018). The basins, from north to south, are Cochiguaz River at El Peñón (675 km^2, Figure 4.1a, ID 4313001), Choapa River at Cuncumén (1132 km^2, Figure 4.1b, ID 4703002), Claro River at El Valle (349 km^2, Figure 4.1c, ID 6127001), Colorado River at the confluence with Palos River (878 km^2, Figure 4.1d, ID 7112001), Cautín River at Rari-Ruca (1306 km^2, Figure 4.1e, ID 9123001) and Futa River at Tres Chiflones (517 km^2, Figure 4.1f, ID 10142003). Figure 4.1 also shows the average seasonality for precipitation (P), temperature (T), and streamflow. The Cochiguaz and Choapa River basins have a snowmelt-driven regime, receiving precipitation mostly during the fall and winter seasons when water is mainly stored as snow and released during the spring and summer seasons. The Claro and Colorado River basins have a mixed regime, with higher streamflow values during the winter and spring/summer seasons, respectively. The two southernmost basins, Cautín and Futa, have maximum streamflow values during winter (rainfall-driven), though a slight influence of snowmelt is detected in the Cautín River during the spring season.
Five out of the six basins are located on the western slopes of the Andes Cordillera, with the Futa River basin being the exception. The Cochiguaz and Choapa basins have the highest altitudes, which explains the runoff dependence on snowmelt. The Claro and Colorado basins span a wide range of elevations from the Andes to the valleys, which explains their mixed hydrological regimes since, between -28°S and -42°S, snow only accumulates in the Andes Cordillera above ∼2,000 m a.s.l. The Cautín and Futa River basins have lower altitudes and low to no-snow influence. Finally, all the case study basins have a low degree of human intervention and were selected based on the following criteria: (a) a near-natural flow regime defined as a maximum threshold value of 5% for the relationship between annual volume of water assigned for permanent consumptive use and the mean annual flow, (b) absence of large reservoirs within each catchment, and (c) small (<2%) glacierized area (Alvarez-Garreton et al., 2018).

4.3 Datasets

4.3.1 Meteorological and streamflow data

Daily precipitation (P), and maximum (T_{max}) and minimum (T_{min}) air temperature are obtained from the CR2MET v2.0 dataset (Boisier et al., 2018; DGA, 2022; Boisier, 2023), which covers continental Chile with a horizontal resolution of 0.05° x 0.05° for the period 1979-2020. CR2MET P estimates are obtained through a two-step approach consisting of (i) computing the probability of precipitation at each grid cell through logistic regression models, and (ii) calculating daily precipitation amounts using multiple linear regression models. All the models use ERA5 reanalysis outputs (Hersbach et al., 2020) and geomorphological attributes as predictors, and daily precipitation from meteorological stations as predictands (note that, in the case of precipitation occurrence, the vector with predictands only contains zeroes and ones). For T_{max} and T_{min}, LST from MODIS is also included as a predictor.
To obtain sub-daily meteorological time series, CR2MET daily precipitation, and temperature are disaggregated into hourly time steps using the sub-daily distribution provided by ERA5-Land (Muñoz Sabater, 2019), which is bias corrected to match the CR2MET daily values. Relative humidity (RH), wind speed (W), atmospheric pressure (AP), and incoming shortwave radiation (K_{in}) are derived for the same horizontal resolution grid by spatially interpolating ERA5-Land outputs. ERA5-Land wind speed is bias-corrected us-

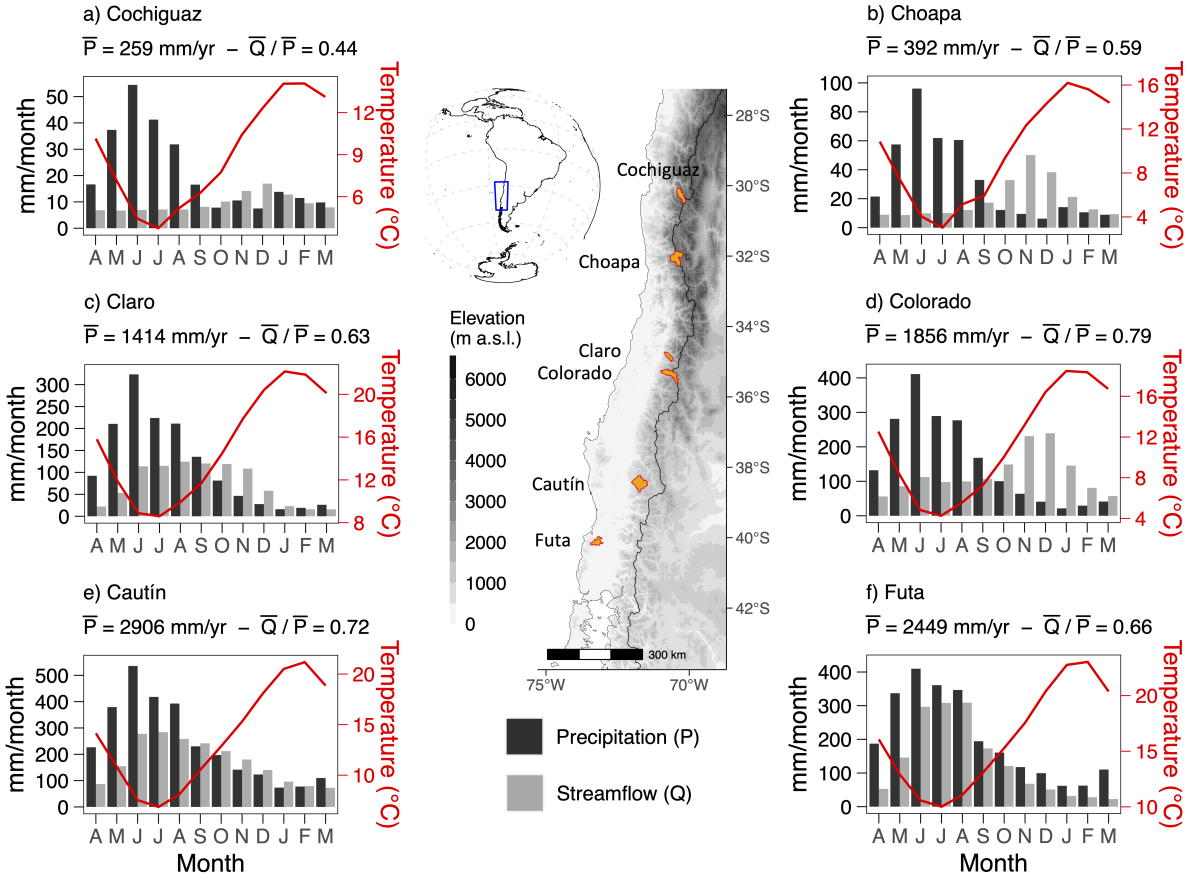

Figure 4.1: Location of the six case study basins considered in this study (center panel), along with the seasonal cycles of P, Q, and T (in red) for the climatological period 1980-2018. Overlines represent mean annual values. Notice that the winter season corresponds to JJA, while the summer season to DJF.

ing wind simulations from WRF at a 1-km horizontal resolution (Geophysics Department & Ministerio de Energía, 2018). Incoming longwave radiation (L_{in}) was computed with the parameterization proposed by Iziomon et al. (2003) using the bias-corrected hourly temperature. To conduct hydrological model simulations, all the meteorological variables are grouped to obtain 3-hourly time steps.
Finally, daily streamflow data is obtained from stations maintained by the Chilean Water Directorate (DGA, in Spanish), also available in the CAMELS-CL dataset (Alvarez-Garreton et al., 2018).

4.3.2 Remote sensing products

This work uses remotely sensed fractional snow-covered area, actual evapotranspiration, land surface temperature, and soil moisture for hydrologic model evaluation (Figure 4.2). The fractional snow-covered area is derived from MODIS products (MOD10/MYD10; Hall & Riggs, 2016). To obtain a unique time series, daily MOD10 and MYD10 estimates are averaged at each MODIS grid cell. If only one product is available for a specific day and grid cell, that value is used to estimate fSCA. All gaps (i.e., days and grid cells where MOD10 and MYD10 are not available) are filled at the original horizontal resolution (500 m) using the methodology proposed by Cornwell et al. (2016). Actual evapotranspiration is obtained from the MOD16 product (Mu et al., 2011) using 8-day estimates with a

1-km horizontal resolution. Land surface temperature for each day and each grid cell is obtained as the average between the MOD11 and MYD11 products (Wan, 2014) at a 1-km horizontal resolution. If one of the two LST estimates is unavailable, the day is considered without information. For soil moisture (0.25° horizontal resolution), the ESA-CCI product is used (Dorigo et al., 2017).

To resolve the mismatch between the horizontal grid of remote sensing products and the grid discretization of CR2MET (which is the same as in the hydrological model), ET, LST, and fSCA are upscaled using spatial averages, and (SM is downscaled to the closest CR2MET grid cell that is closest; e.g., dos Santos Araujo et al., 2024). Empty values were not filled out for ET, LST, and SM, as opposed to fSCA.

4.3.3 Ancillary data

Since this study considers a distributed hydrological model that requires *a priori* parameter fields, these are derived using information from different grid cell attributes. The SoilGrids dataset (Poggio et al., 2021) is used to derive mean clay and sand content and mean bulk density for the first 2 m soil depth at a 250-m horizontal resolution. The elevation is estimated from the Shuttle Radar Topography Mission (SRTM; Farr et al., 2007). All datasets are upscaled to match the CR2MET grid using spatial averages.

4.4 Approach

This work evaluates the impact of two methodological decisions - namely, the choice of the objective function (OF) and the choice of parameter regularized - on streamflow performance, the simulated spatial patterns of ET, SM, LST, and fSCA, and the simulated annual cycles of these variables. Figure 4.2 illustrates the main steps required to conduct the simulations. First, hydrometeorological datasets are obtained and processed to force and evaluate the VIC model (step 1 in Figure 4.2, sections 4.3 and 4.4.1); secondly, the parameter regularization strategy is designed and implemented (step 2 in Figure 4.2, section 4.4.2). Finally, a suite of model calibration experiments are conducted to assess the impact of the choice of calibration metric and parameter regularized on simulated spatial patterns and annual cycles of different hydrological variables (step 3, section 4.4.3).

4.4.1 Hydrological Modeling

The VIC model (Liang et al., 1994) is used to simulate state variables and fluxes at a 0.05°x 0.05° horizontal resolution. VIC is a semi-distributed physically based hydrological model that solves energy and mass balance equations. Precipitation can be partitioned into snowfall and rainfall that serve as inputs for canopy storage. The maximum amount of water that the canopy intercepts is estimated using the Leaf Area Index (LAI; Dickinson, 1984). The snowpack is represented by two layers, where the top layer is used for energy balance computations Andreadis and Lettenmaier (2006). The soil column is vertically discretized into three layers, with the top and bottom layers controlling the infiltration and baseflow generation, respectively. To calculate infiltration, VIC uses the Xinanjiang formulation (Zhao et al., 1980), assuming that infiltration capacity varies within an area

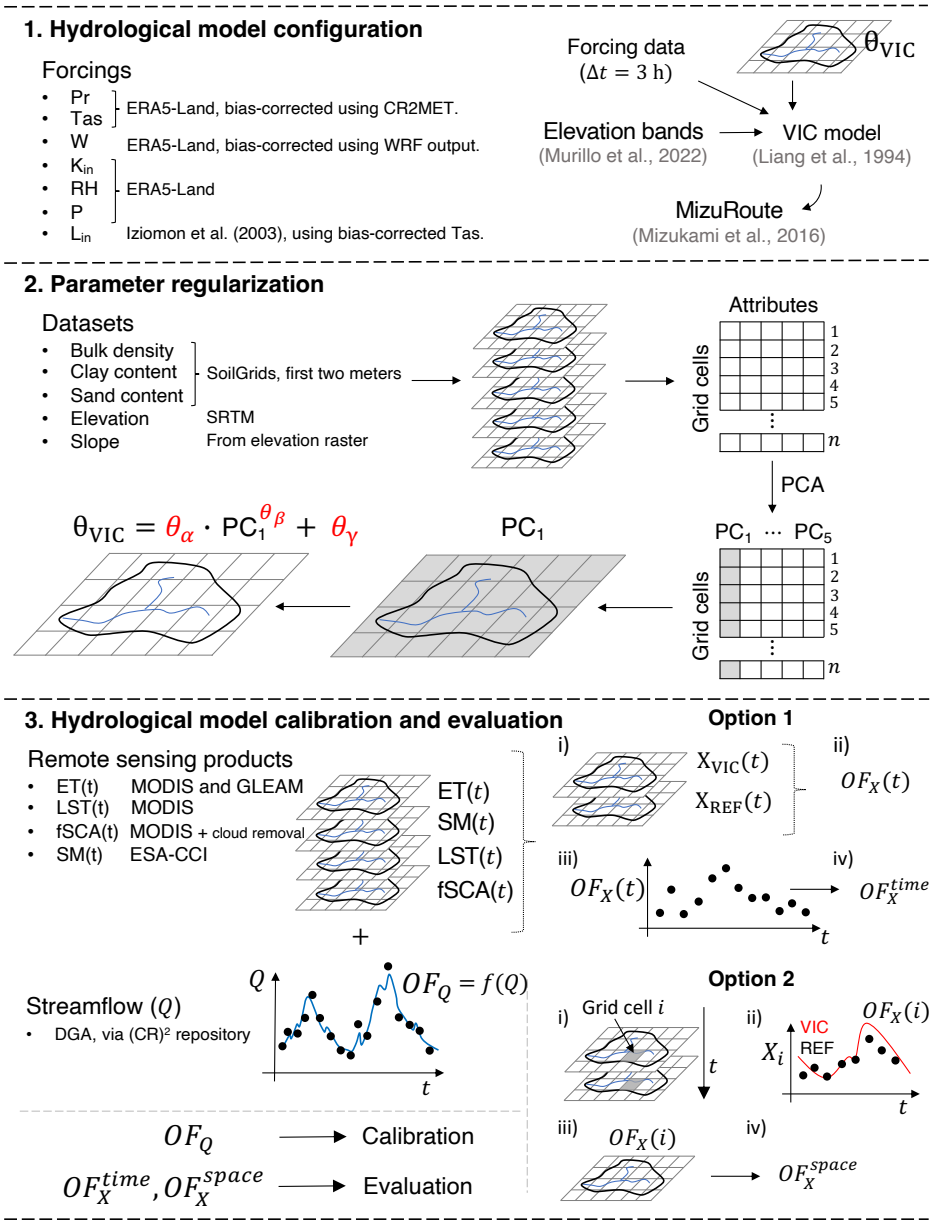

Figure 4.2: Schematics of the methodology used in this study. Panel 1) shows the sources of the different datasets considered to run the VIC model as well as the main model configuration features. 2) Approach used to obtain *a priori* soil parameter fields for each case study basin. 3) Evaluation of streamflow (Q) simulations and spatial patterns of other variables (X) by contrasting: (i) a simulated and reference map for the variable X (e.g., ET, SM, LST or fSCA) for each time step t, and (ii) simulated and reference time series at each grid cell.

(Wood et al., 1992). Excess runoff is generated in those areas where precipitation exceeds the available soil moisture storage of the first soil layer. VIC assumes that gravity drives drainage, using the formulation proposed by Brooks and Corey (1964). In this regard, water enters the cell only from the atmosphere, i.e., VIC does not consider lateral fluxes among grid cells. Baseflow is computed using a formulation proposed by Franchini and Pacciani (1991). The reader is referred to Liang et al. (1994) for more details about the VIC model.

Sub-grid horizontal heterogeneity in VIC can be considered by incorporating different

land cover types. Here, the International Geosphere-Biosphere Program (IGBP) classification for the year 2010 is used from the MCD12Q1 v006 land cover product (Sulla-Menashe & Friedl, 2018), in order to represent all land cover types spanning at least 2% of each grid cell area. Mean monthly LAI values for these land cover types are derived from the MOD15A2 product. Soil Bulk density is estimated as the vertical average of values retrieved from the SoilGrids product (Poggio et al., 2021) across the top 2 m soil depth. The vertical heterogeneity is considered by distributing precipitation and air temperature along 200-m elevation bands, following the recommendations by Murillo et al. (2022).

The vector-based mizuRoute model (Mizukami et al., 2016) is used to estimate the streamflow at the outlet of each basin. mizuRoute first performs a hillslope routing using a gamma-distribution-based unit-hydrograph to delay VIC runoff, and then routes the delayed runoff for each river reach defined by the river network topology. For river routing, the diffusive wave scheme is used, following the setup recommended by Cortés-Salazar et al. (2023). Manning's roughness coefficients and channel widths are spatially distributed within each basin through regression equations that use river reach attributes as predictors (Niño, 2002; Mendoza et al., 2012). Note that routing parameters were not included in subsequent calibration experiments (described in section 4.3).

4.4.2 Parameter regularization

The parameter regularization technique used here is based on a PCA over a suite of grid cell attributes within each basin, including bulk density, clay and sand content, elevation, and slope (e.g., Samaniego et al., 2010; Mizukami et al., 2017; Beck et al., 2020). The underlying motivation is to extract the main patterns explaining the spatial distribution of physiographic attributes and use the dominant signal to obtain VIC parameter fields. To this end, grid cell attributes within a matrix are sorted (step 2 in Figure 4.2), with grid cells as rows and attributes as columns, and perform PCA, obtaining a matrix with the same dimensions, where the first principal component (PC1; or first column of the new matrix) explains most of the variance among the attributes (See Figure B.1). Then, the PC1 map is used as a predictor to obtain *a priori* spatially distributed VIC parameters (θ_{VIC}) using three-super parameter (Pokhrel et al., 2008), highlighted in red in Figure 4.2. The super-parameters considered in this study are (i) an additive term (θ_γ), (ii) a factor (θ_α), and (iii) an exponent controlling the linearity of the relationship (θ_β).
It is important to note that none of the routing parameters are calibrated in any experiment since their spatial distribution is defined through regression equations that use river reach attributes as predictors (Niño, 2002; Mendoza et al., 2012).

4.4.3 Model calibration experiments

The VIC parameters considered for the calibration process were identified as the most sensitive by Sepúlveda et al. (2022), including (i) b_{infilt}, which controls the infiltration process, (ii) Ds and Ws, which control the linearity of the curve used to calculate the water leaving the third (deepest) soil layer, (iii) Ds_{max} as the maximum baseflow rate, (iv) hydraulic conductivity (K_{sat}), related to how water percolates between the soil layers, (iv) the soil layers' depths, (v) the threshold temperature to separate precipitation into rainfall and snowfall, (vi) maximum snow albedo and (vii) snow albedo temporal decay rate. The benchmark calibration considers spatially constant VIC parameters (e.g.,

L. Melsen et al., 2016), whereas additional calibration experiments consider the spatial regularization of individual soil parameters (b_{infilt}, Ds, Ds_{max}, K_{sat} and soil layers' depth), maintaining the remaining parameters spatially constant. With this, this chapter seeks to understand the added value of distributing in space the values of a parameter associated with specific processes. A final calibration experiment considers the vertical distribution of the hydraulic conductivity, for which the same equation of Figure 4.2 is applied (step 2; Pokhrel et al., 2008); however, instead of using the PC1, the vertical distribution of the bulk density obtained from the SOILGRIDS database is used to define an *a priori* vertical distribution. Hence, for this exercise, K_{sat} is horizontally constant within each soil layer, adopting different values in each layer.

Streamflow-based performance metrics

All the calibration experiments conducted here seek to maximize streamflow-based objective functions by running the Dynamically Dimensioned Search (DDS; Tolson & Shoemaker, 2007) algorithm, implemented within the OSTRICH software (Matott, 2017) with a maximum number of 2000 iterations. All the metrics consider daily time series of simulated and observed streamflow. First, the Kling-Gupta efficiency is used (KGE; Gupta et al., 2009; Kling et al., 2012), which seeks to minimize the Euclidean distance between performance metrics related to volume (β), variability (γ), and timing (r) and their optimal values (which is 1):

$$OF_1 = KGE = 1 - ED = 1 - \sqrt{(1-\beta)^2 + (1-\gamma)^2 + (1-r)^2} \tag{4.1}$$

Where $\beta = \mu_s/\mu_o$ is the ratio between simulated (s) and observed (o) average values (μ), $\gamma = (\sigma_s/\mu_s)/(\sigma_o/\mu_o)$ where σ represents the standard deviation and r is the temporal correlation between observed and simulated daily values. The second objective function (Eq. 4.2) is the Nash-Sutcliffe efficiency (NSE; Nash & Sutcliffe, 1970).

$$OF_2 = NSE = 1 - \frac{\sum_{t=1}^{T}(Q_s^t - Q_o^t)^2}{\sum_{t=1}^{T}(Q_o^t - \overline{Q}_o)^2} \tag{4.2}$$

Where Q_s and Q_o are simulated and observed daily streamflow, t indicates the time step, and $\overline{Q}$ is the mean daily streamflow. Since both KGE and NSE are influenced by high flows (Clark et al., 2021), the composite metric (Eq. 4.3) proposed by Garcia et al. (2017) that incorporates the KGE computed with $1/Q$ is used to give more weight to low flows:

$$OF_3 = \frac{KGE(Q) + KGE(1/Q)}{2} \tag{4.3}$$

The third calibration metric is the Nash-Sutcliffe efficiency computed from simulated and observed daily flow duration curves (FDCs) (e.g., Nijzink et al., 2016). Finally, the last OF (Eq. 4.4) is based on a Euclidean distance approach (e.g., Schoups et al., 2005) that combines the daily KGE with five hydrological signatures (HSs) proposed by Yilmaz et al. (2008), which focus on different aspects of the flow duration curve: (i) high flows, (ii) mean flows, and (iii) low flows (e.g., Westerberg et al., 2011; Shafii & Tolson, 2015).

$$OF = 1 - \sqrt{(1 - KGE(Q))^2 + \sum_{i=1}^{N} (HS_i^{opt} - HS_i)^2} \tag{4.4}$$

Where HS_i^{opt} is the optimal value for the i^{th} hydrological signature. The hydrological signatures in Eq. 4.4 are the bias in the runoff ratio (Eq. 4.5), bias for high flows (Eq. 4.6), bias in low flows (Eq. 4.7), bias in mid flows (Eq. 4.8) and median flows (Eq. 4.9).

$$HS_{biasRR} = \frac{\sum_{t=1}^{T} (Q_s^t - Q_o^t)}{\sum_{t=1}^{T} Q_o^t} \tag{4.5}$$

$$HS_{biasFHV} = \frac{\sum_{h=1}^{H} \left(Q_s^h - Q_o^h\right)}{\sum_{h=1}^{H} Q_o^h} \tag{4.6}$$

Where T is the number of time steps, and $h = 1, 2, ..., H$ is the high flow indices for flows with exceedance probabilities lower than 0.02.

$$HS_{biasFLV} = -1 \cdot \frac{\sum_{l=1}^{L} \left[log(Q_s^l) - log(Q_s^L)\right] - \sum_{l=1}^{L} \left[log(Q_o^l) - log(Q_o^L)\right]}{\sum_{l=1}^{L} \left[log(Q_o^l) - log(Q_o^L)\right]} \tag{4.7}$$

Where $l = 1, 2, ..., L$ is the index of flows with exceedance probabilities between 0.7 and 1, being L the index of the minimum flow.

$$HS_{biasFMS} = \frac{[log(Q_s^{m1}) - log(Q_s^{m2})] - [log(Q_o^{m1}) - log(Q_o^{m2})]}{[log(Q_o^{m1}) - log(Q_o^{m2})]} \tag{4.8}$$

Where $m1$ and $m2$ represent the lowest and highest exceedance probabilities (0.2 and 0.7, respectively) within the midsegment of the flow duration curve.

$$HS_{biasFMM} = \frac{log(Q_s^{med}) - log(Q_o^{med})}{log(Q_o^{med})} \tag{4.9}$$

Where *med* corresponds to the median value.

Spatial pattern performance metrics

The spatial efficiency metrics (Eq. 4.10) proposed by Dembélé, Ceperley, et al. (2020) is used to evaluate ET, SM, LST, and fSCA.

$$E_{SP} = 1 - \sqrt{(1 - r_s)^2 + (1 - \gamma)^2 + (1 - \alpha)^2} \tag{4.10}$$

Where r_s is the Spearman correlation coefficient between simulated and reference values, γ is the ratio of the coefficients of variation (as in KGE; Eq. 4.1), $\alpha = 1 -$

$RMSE(Z_{X_s}, Z_{X_o})$, being Z the time series with standardized values for the variable X (note that here it is used the same notation as in Dembélé, Hrachowitz, et al., 2020) and RMSE the Root Mean Square Error. The standardization aims to avoid a direct contrast between the model and reference values since remotely sensed products have biases that could affect the metric. However, bias-accounting metrics are used, such as KGE and RMSE, to assess the model's ability to reproduce the raw remotely sensed estimates.

Following Dembélé, Ceperley, et al. (2020), two approaches are used to contrast spatially distributed simulations against observations (options 1 and 2 in step 3, Figure 4.2). Option 1 considers the contrast between simulated and reference maps of a variable X at each time step, resulting in a time series of performance measures ($OF_X(t)$) that can be used to compute a summary metric OF_X^{time} by temporally averaging all values. Alternatively, option 2 involves the comparison between simulated and reference time series at each grid cell to obtain a map of performance metrics ($OF_X(i)$) that can be used to calculate a summary performance measure OF_X^{space} by spatially averaging all values.

For soil moisture, the correlation coefficient (Eq. 4.11) between simulated (θ_s) and reference soil moisture (θ_{ref}) is used as in Tong et al. (2021).

$$O_{SM} = \frac{\sum_{t=1}^{T} \left[(\theta_s^t - \overline{\theta}_s)(\theta_{ref}^t - \overline{\theta}_{ref})\right]}{\sqrt{\sum_{t=1}^{T} \left[(\theta_s^t - \overline{\theta}_s)^2(\theta_{ref}^t - \overline{\theta}_{ref})^2\right]}} \tag{4.11}$$

Where t represents the time step and overlines average values. For comparisons, only the simulated soil moisture from the first soil layer is included since the ESA-CCI soil moisture product considers the first 0.5–2 cm. To overcome mismatches and systematic biases between the reference and SM_1 due to different soil depths and properties, mean–standard deviation matching (Draper et al., 2009; López López et al., 2017) was used (Eq. 4.12).

$$\theta_s^* = \frac{\sigma_{\theta_{ref}}}{\sigma_{\theta_{sim}}} \cdot \left(\theta_s - \overline{\theta}_s\right) + \overline{\theta}_{ref} \tag{4.12}$$

Where $\sigma_{\theta_{ref}}$ and σ_{θ_s} are the standard deviation of the reference and simulated soil moisture. Finally, the quality of fSCA simulations is assessed by using the root mean squared error (RMSE) and the metric O_{SC} proposed by Tong et al. (2021):

$$O_{SC} = 1 - (S_O + S_U) \tag{4.13}$$

where S_O and S_U represent the fraction of the total number of simulation days with overestimation and underestimation of simulated fSCA, respectively. To avoid noise values, a threshold of 1% is applied; hence, simulated and reference grid cell fSCA values below 0.01 are considered no snow cover.

Finally, additional experiments are conducted to verify if the VIC model can simulate the Q and ET and SM seasonalities simultaneously using spatially constant parameters. To this end, the VIC model parameters are calibrated using an Euclidean-based objective function:

$$OF = \sqrt{(1 - KGE(Q)^2 + (1 - E_{SP}(X))^2} \tag{4.14}$$

Where X refers to ET or SM_1.

Only for graphical purposes, ET and SM annual values are normalized as $X' = (X - X_{min})/(X_{max} - X_{min})$ to avoid mismatches and biases in the reference product and focus on the simulated annual cycle. The notation SM_1, SM_2, and SM_3 is used to refer to the moisture content in soil layers 1, 2, and 3 of the VIC model, respectively.

4.5 Results

First, the impact of different calibration objective functions and regularized parameters on the performance of streamflow simulations is examined across basins (Figure 4.3). In general, for a given combination of calibration metric and basin the objective function values achieved are similar, regardless of the parameter regularized, suggesting a high degree of compensation among parameters. In the Choapa and Cautín River basins, lower OF values are achieved for K_{sat} parameter (similar to $Depth_2$ at the Futa River basin) when the OF is $0.5 \cdot KGE(Q) + 0.5 \cdot KGE(1/Q)$. Interestingly, the components of the objective functions can vary considerably, especially $KGE(1/Q)$, which gives more weight to the low flows. In particular, the values of this metric are generally negative or close to zero at the Claro (mixed regime) and Futa (rainfall-driven) River basins. On the other hand, the $KGE(1/Q)$ values are stable for all configurations in the snowmelt-driven basins (Cochiguaz and Choapa), regardless of the inclusion of $KGE(1/Q)$ in the calibration process.

Figure 4.4 illustrates the impacts of calibration OF and choice of regularized parameter on simulated ET at the Colorado River basin with respect to the benchmark calibration. Here, parameters controlling infiltration (b_{infilt}), soil depth ($Depth_2$), vertical moisture distribution (Vertical K_{sat}), and baseflow (Ds_{max}) are examined. Overall, the results show that some combinations of OF and parameter regularization strategy may enhance the performance of ET simulations compared to the benchmark calibration. In some grid cells, the changes in ET performance can be larger than 1 E_{SP}, with positive (e.g., for $KGE(Q)$ and Ds_{max}) or negative (e.g., $KGE(Q)$ & $KGE(1/Q)$ and b_{infilt}) effects. Additionally, the degree of improvement can vary depending on the choice of OF and parameter regularized. For example, when using $KGE(Q)$ and $KGE(Q)$ & $NSE(FDC(Q))$ as OFs, regularizing K_{sat} and Ds_{max} (b_{infilt} and $Depth_2$) yields improved (declined) performance in ET simulations. For $NSE(Q)$, regularizing any parameter yields a slight increase in E_{SP}(ET), whereas the opposite occurs with $KGE(Q)$ & $KGE(1/Q)$. For the OF combining $KGE(Q)$ and hydrological signatures, mixed results are obtained: $\Delta E_{SP} > 0$ (blue) for lower grid cells, and $\Delta E_{SP} < 0$ (red) for high altitude grid cells. Similar ET performance results are obtained for the remaining basins, although a general improvement in ET performance is achieved when regularizing parameters for the Claro and Futa River basins. Notably, parameter fields that yield dissimilar performance in ET provide similar streamflow performance metrics (Fig. 4.3).

Now, the focus is on how different combinations of calibration metrics and parameter regularization strategies affect simulated seasonal water balances. Figure 4.5 shows the average seasonality of streamflow, normalized catchment-scale ET and SM_1, land surface temperature, and fractional snow-covered area considering the parameter sets that maximize $OF = f(KGE(Q), KGE(1/Q))$ at the Choapa (snowmelt-driven), Colorado

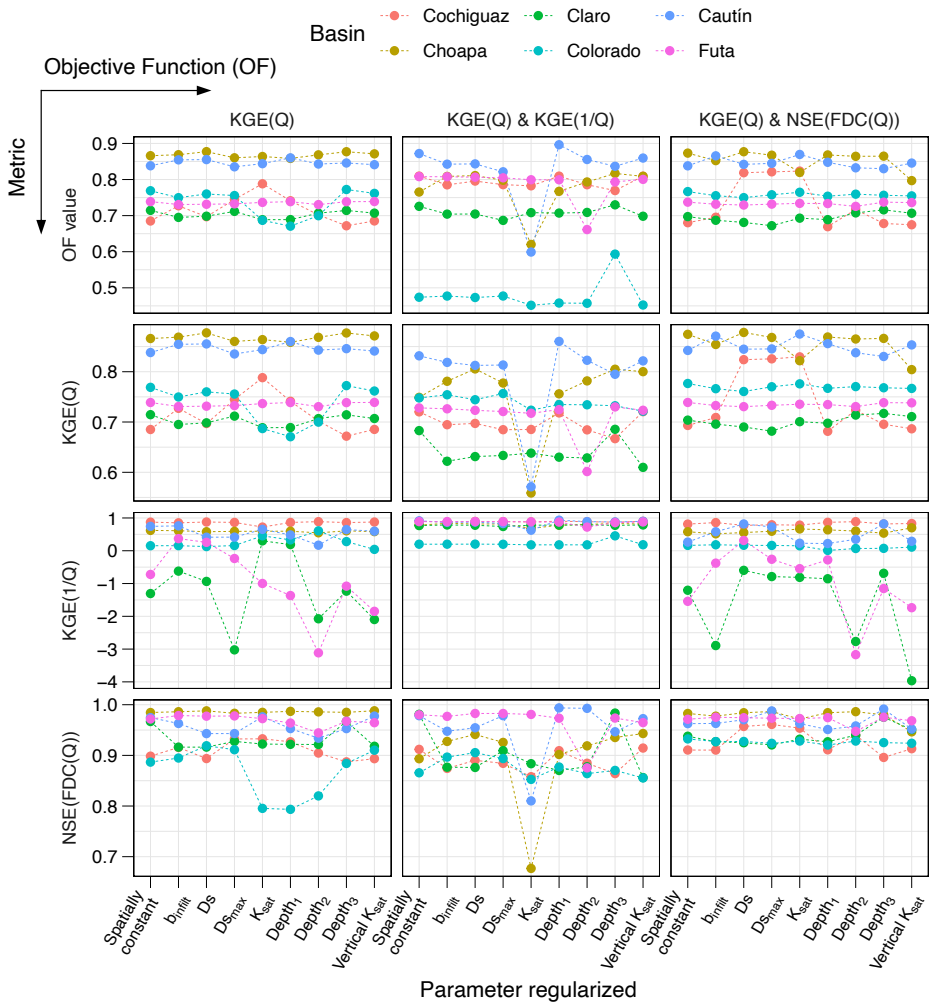

Figure 4.3: Best objective function value (top row) and streamflow performance metrics (remaining rows) associated with the best parameter set obtained with spatially constant parameters (i.e., benchmark calibration, $\theta_{VIC} = \theta_{\gamma}$ in Fig. 4.2) and different regularization strategies. The columns show results for three different calibration objective functions. In each panel, the results for individual basins are displayed with different colors.

(mixed regime), and Cautín (rainfall-driven) River basins. It can be noted that, although streamflow is underestimated during some spring and summer months in the Choapa River basin, all the calibration experiments yield a reasonable representation of the Q seasonality, with little differences among configurations. Similar results are obtained for LST, with an overall underestimation at the Choapa River basin.

The most striking result in Figure 4.5 is that ET seasonalities can change drastically depending on the model configuration, as opposed to Q, SM_1, LST, and fSCA. In the Choapa River basin, regularizing different parameters yields different biases, with the largest underestimations for the spatially constant and $Depth_2$ cases during spring. In the Colorado River basin, none of the model configurations provided a reasonable representation of ET seasonality, whereas mixed results are obtained at the Cautín River basin, with both good (e.g., b_{infilt}, K_{sat} and Ds_{max}) and poor (e.g., spatially constant and $Depth_2$) representations. In summary, regularizing different parameters may shift the seasonality and, consequently, reduce biases, without impacting the Q performance.

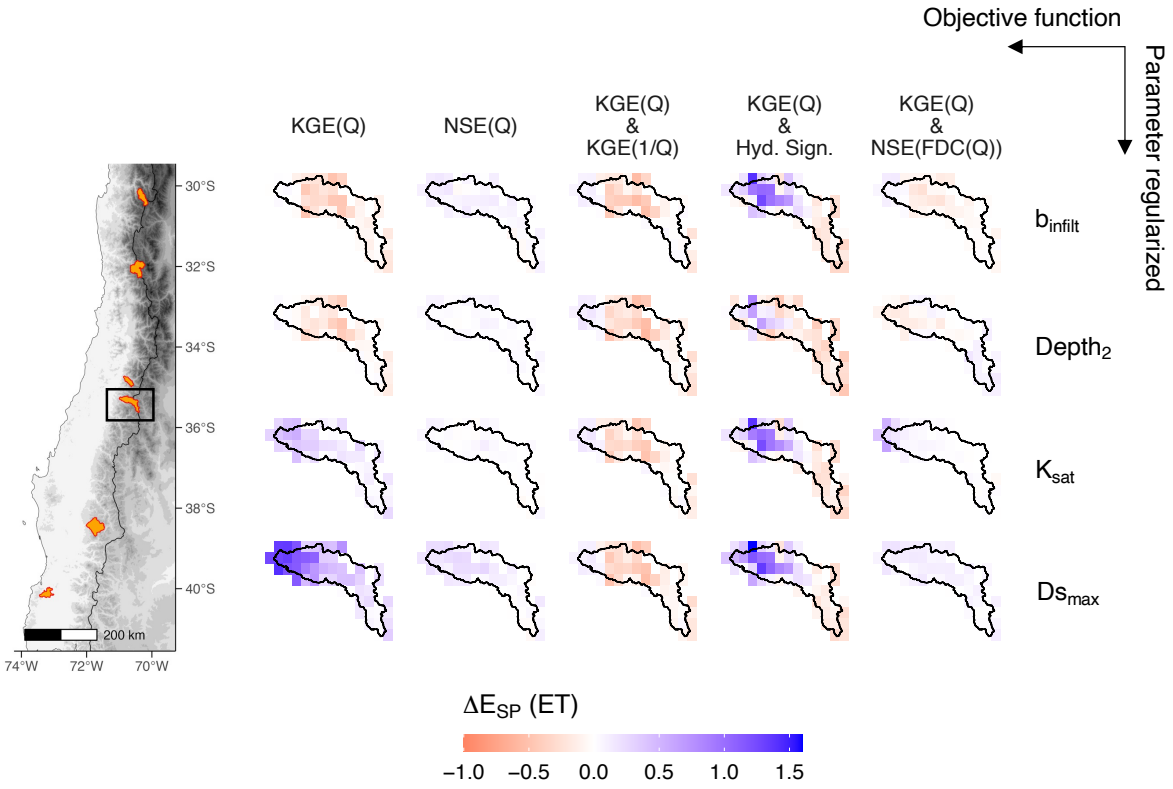

Figure 4.4: Change in ET performance (with respect to the benchmark calibration) for the Colorado River basin during the calibration period (2005-2018). Blue (red) colors represent an improvement (decline) in the performance of simulations. Note that the benchmark changes across objective functions.

Figure 4.6 illustrates how the choice of streamflow-based OF affects the seasonality of simulated Q, ET, SM_1, LST, and fSCA at the Cautín River basin. In general, different OFs yield similar streamflow responses, except when (i) the model parameters are spatially constant and calibrated using $KGE(Q)$ (black line), which yields an overestimation in Q, and (ii) $Depth_2$ is regularized and the OF is $KGE(Q)$ and $NSE(FDC(Q))$, which yields underestimation of Q.
The results for LST show little disagreement among OFs and model configurations, with a reasonable representation of seasonality. For SM_1, the results reveal a slight dispersion among configurations, with larger biases with respect to the reference (white dots) when K_{sat} is regularized. The timing of seasonality of fSCA is well simulated with all the combinations of OF and regularization strategy, though the latter decision introduces large discrepancies in biases. For example, when b_{infilt} is regularized and the calibration metric is $KGE(Q)(KGE(Q)$ & $NSE(FDC(Q)))$, fSCA is underestimated(overestimated).
Notably, the effects of the regularization strategy on simulated ET annual cycles can be very different depending on the calibration metric. For example, when $KGE(Q)$ or $KGE(Q)$ & $KGE(1/Q)$ are used as OFs, shifted ET seasonalities are obtained if b_{infilt} and $Depth_2$ are regularized, or if model parameters are spatially constant. Conversely, when $KGE(Q)$ & $NSE(FDC(Q))$ is maximized in the calibration process, all ET seasonalities collapse into the reference values, regardless of the regularization strategy. A similar behavior is obtained for the rest of the basins.

Does the improvement in streamflow-based OFs relate to improvements in other Q

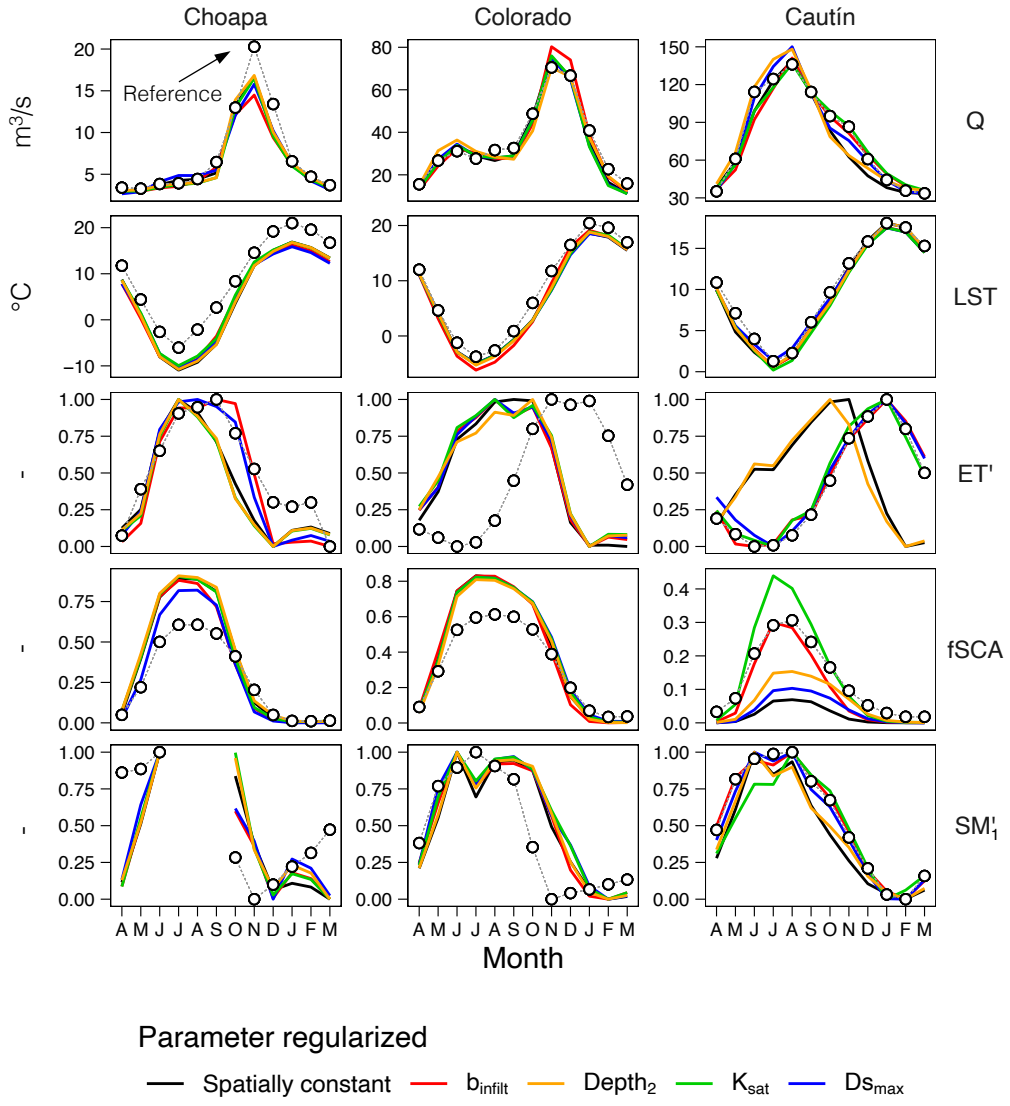

Figure 4.5: Catchment-scale annual cycles (calibration period 2005-2018) of streamflow (Q), normalized evapotranspiration (ET'), normalized soil moisture (SM_1'), land surface temperature (LST), and fractional snow-covered area (fSCA) in the Choapa (snowmelt-driven), Colorado (mixed regime) and Cautín (rainfall-driven) River basins. Mean monthly values are computed only if at least 50 days with information are available, and the calculation considers only days with information. Note that there is not enough data to compute monthly SM_1 averages at the Choapa River basin during winter. All the results are associated with the parameter sets that maximize KGE(Q) & KGE(1/Q). Notice that winter and summer correspond to JJA and DJF, respectively. Streamflow observations and remotely sensed variables are referred to as "references" and are represented with white dots. Notice that ET and SM_1 annual values are normalized as $X' = (X - X_{min})/(X_{max} - X_{min})$.

metrics and simulated variables? To seek answers, the Spearman's rank correlation coefficient is computed between the OF values obtained in N = 2,000 iterations of the optimization process, and other performance metrics for Q, ET, SM_1, LST, and fSCA. Figure 4.7 displays results for calibrations conducted at the Choapa, Colorado, and Cautín River basins using $OF = KGE(Q)$ & $KGE(1/Q)$ as OF. As expected, high correlations are obtained between streamflow-based metrics and OFs. However, negative correlations

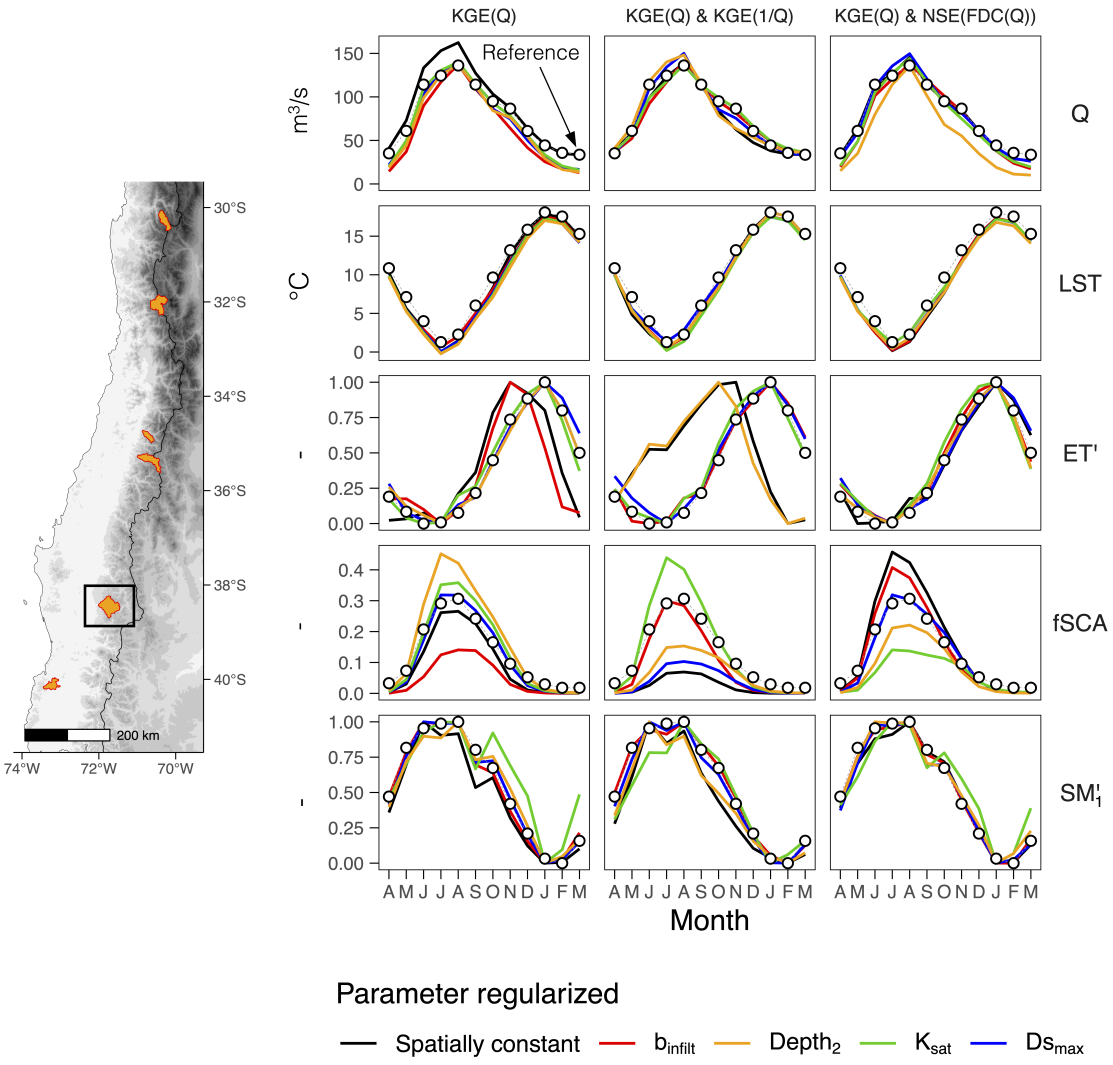

Figure 4.6: Impacts of the spatial regularization strategy on simulated annual cycles for a given calibration metric (columns). The variables analyzed are streamflow (Q), normalized evapotranspiration (ET'), normalized soil moisture (SM_1'), land surface temperature (LST), and fractional snow-covered area (fSCA) for the Cautín River basin (highlighted by a square on the map) during the calibration period (2005-2018). The reference datasets are shown as white dots. The normalization of ET and SM_1 is computed as $X' = (X - X_{min})/(X_{max} - X_{min})$.

are obtained for ET, especially at the Choapa and Colorado River basins. For the case of SM_1, a larger number of parameter configurations yield positive correlations with OF values. For LST and fSCA, poor correlations are obtained with OF, which can be explained by the relatively good performance obtained for these variables, regardless of the value of OF.

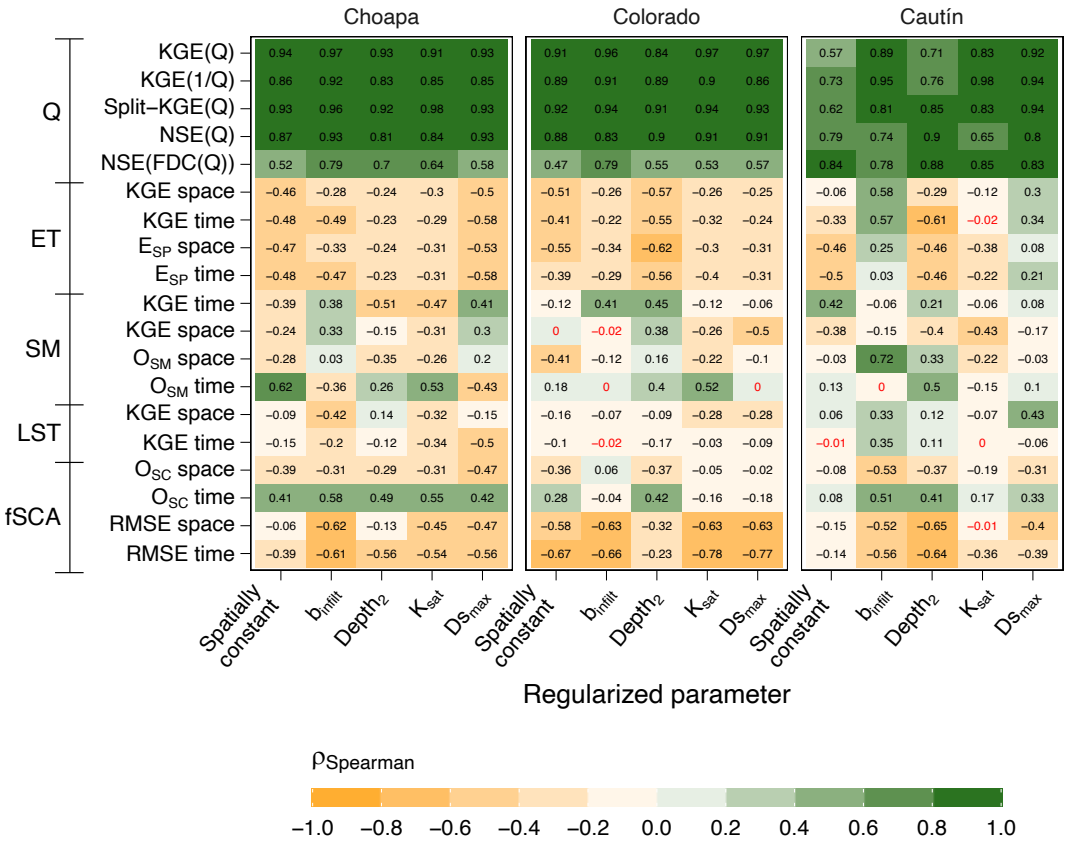

Figure 4.7: Spearman's rank correlation coefficient between the values of calibration metric ($OF = KGE(Q)$ & $KGE(1/Q)$) and performance measures of Q and other simulated variables during the calibration process (N = 2,000 parameter sets). Black(red) numbers represent correlations with $p_{value} < 0.05$ ($p_{value} > 0.05$). Note that lower values are better for $RSME(fSCA)$, while the Q-based metrics are positively oriented (higher, better).

4.6 Discussion

4.6.1 Effects on spatial patterns

The effects of the choice of OF on spatial patterns reported here are somewhat expected, since the impacts on behavioral parameters and, therefore, the simulation of different parts of the hydrograph, have been vastly discussed in the literature (e.g., Gupta et al., 1998; Merz et al., 2011; Garcia et al., 2017; Westerberg et al., 2011; K. Fowler, Peel, et al., 2018; Khatami et al., 2019). Additionally, different well-behaved parameter sets in terms of streamflow may impact other hydrological variables. However, the impacts of different parameter regularization techniques or the choice of parameters regularized on simulated spatial patterns are less documented in the literature. The results of this work show that regularizing different parameters impact simulated spatial patterns differently, regardless of the calibration metric used and without compromising streamflow performance, which aligns well with previous work (Samaniego et al., 2010; M. Demirel et al., 2018). Moreover, the simulation of spatial patterns can be improved without losing streamflow performance, in agreement with (M. C. Demirel et al., 2024). Nevertheless, this work did not find a unique combination of OF and parameter regularization strategy that improves simulated spatial patterns for all variables and basins compared to the benchmark calibration case.

In this regard, it would be useful to characterize parameter sensitivities in replicating observed spatial patterns (M. Demirel et al., 2018; Saavedra et al., 2022).

4.6.2 Biases in simulated annual cycles

In addition to the evident effects of different parameter configurations on simulated spatial patterns of ET, SM_1, LST, and fSCA, large biases in modeled annual cycles may be obtained for these variables, even if realistic streamflow seasonalities are achieved. In this regard, since ET affects the water balance, this work also considered GLEAM ET simulations (horizontal resolution of 0.25°; Martens et al., 2017) without finding discrepancies with ET MOD16 estimates. Smaller biases in the average seasonality of LST and SM_1 are obtained compared to fSCA and ET. For fSCA, similar annual cycles are obtained with all the spatial regularizations tested, with large discrepancies with respect to the reference. On the other hand, despite a reasonable representation of streamflow seasonality and improvements in spatial pattern efficiency metrics, different parameter configurations can shift simulated annual cycles of ET, affecting the seasonal water balance. The results of this chapter differ from (Rakovec et al., 2016), who calibrated the mHM model for Europe using only streamflow data, finding that ET was better simulated than total water storage and SM.

To explore possible reasons that could explain different ET seasonalities, simulated annual cycles in catchment-scale normalized soil moisture are examined. Figure 4.8 shows normalized SM seasonalities for the second ($SM_2^{'}$) and third ($SM_3^{'}$) soil layers, along with the combinations $SM_{2+3}^{'}$ and $SM_{1+2+3}^{'}$ (i.e., total water content in the soil column). As obtained for ET annual cycles, SM seasonality can also be shifted depending on the parameter regularized. Interestingly, little differences in SM seasonality arise from the model configurations in the Colorado River basin, which produce the same annual cycles for ET (Figure 4.5). For Choapa and Cautín, simulated ET annual cycles for b_{infilt} and Ds_{max} depart from the remaining configurations (Figure 4.5), which can be explained by differences in simulated SM in layers 2 and 3 (Figure 4.8). Further, discrepancies in simulated annual cycles in ET with spatially constant parameters and $Depth_2$ at the Cautín basin are also explained by the simulated behavior in layers 2 and 3. Such behavior can be explained by the VIC model configuration, which enables vegetation roots to access different soil layers' moisture content; hence, a compensation among simulated fluxes yielding similar Q response but with discrepancies in ET could be explained by an extreme case of flux equifinality (Khatami et al., 2019). Since ET directly affects the water balance, it is recommended an assessment of simulated annual cycles of ET when streamflow-only calibrations are conducted.

4.6.3 Overall model performance

The results presented here reveal trade-offs between the performance of streamflow simulations through the parameter estimation process and the performance of other hydrological variables. Specifically, this work obtains that different parameter configurations and objective functions may yield similar streamflow performance, but for very different reasons in terms of simulated spatial patterns, which aligns well with previous work incorporating spatial pattern efficiency metrics in the calibration of distributed hydrological models

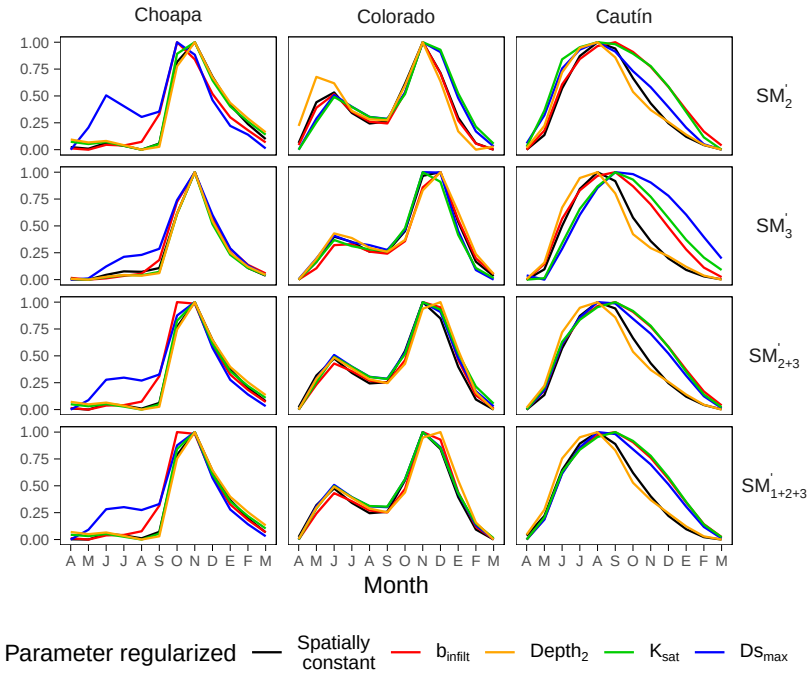

Figure 4.8: Catchment-scale annual cycles of normalized soil moisture (SM') at the Choapa (snow-driven), Colorado (mixed regime), and Cautín (rainfall-driven) River basins over the calibration period (April/2005-March/2018). The results are associated with the best parameter set for objective function KGE(Q) & KGE(1/Q). Notice that winter corresponds to JJA, while summer to DJF. The normalization of SM is computed as $X' = (X - X_{min})/(X_{max} - X_{min})$.

(e.g., López López et al., 2017; M. C. Demirel et al., 2018; Zink et al., 2018; Dembélé, Hrachowitz, et al., 2020; Tong et al., 2021; M. C. Demirel et al., 2024). However, the results of this work also show that, despite possible improvements in spatial efficiencies, large biases may remain in basin-scale annual cycles. In this regard, the examination of simulated seasonalities was crucial to detect deficiencies, especially for ET and fSCA (Rakovec et al., 2016).

The results for the Colorado River basin highlight the need to verify (or calibrate) additional variables since different parameter configurations affect the spatial patterns' performance, and none of them was able to replicate the annual cycle of ET. Figure 4.9 displays the average seasonalities obtained after calibrating model parameters with objective functions that consider (i) only streamflow, (ii) Q and ET, and (iii) Q and SM_1. The results show that streamflow performance decreases considerably in all basins when ET is added to the OF (Jiménez-Navarro et al., 2024), whereas the annual cycle of ET improves drastically. It is worth noting that identifying this issue (ET seasonality poorly simulated while Q seasonality is well represented) was only possible when ET seasonality was visualized, although calibrating with Q + SM_1 improves the seasonality of ET at Choapa and to a lesser degree at the Cautin River basin.

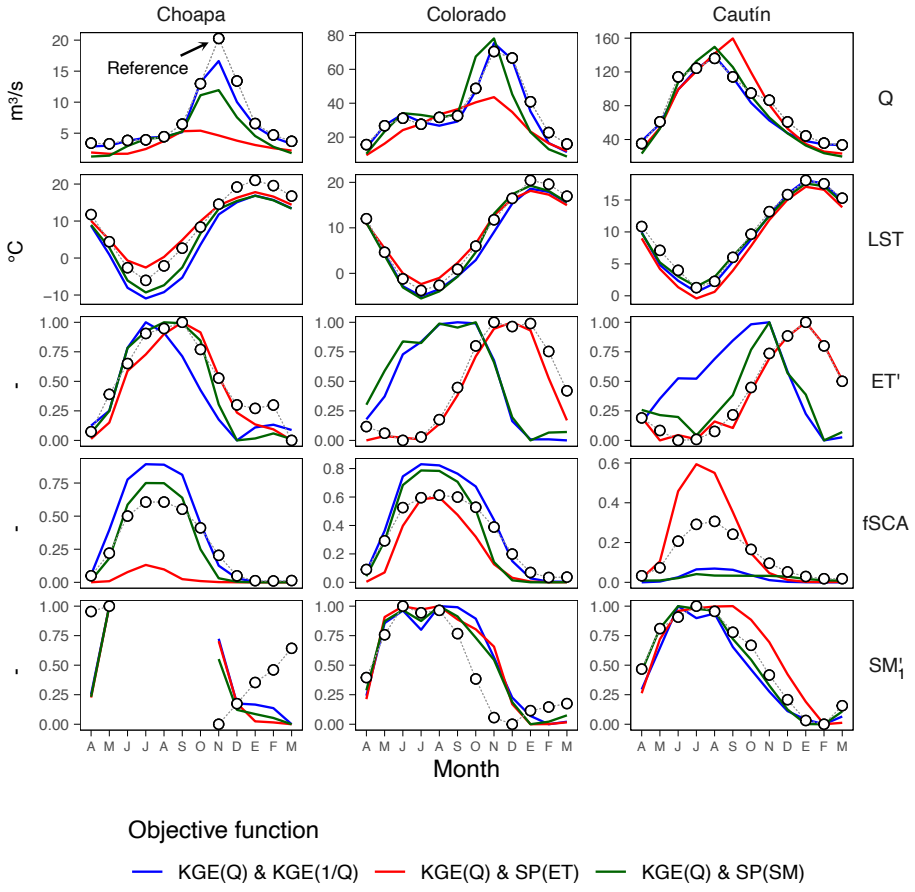

Figure 4.9: Catchment-scale annual cycles of streamflow (Q), normalized evapotranspiration (ET'), normalized soil moisture (SM_1'), land surface temperature (LST), and fractional snow-covered area (fSCA) obtained from a multivariate calibration where parameters are spatially constant. Mean monthly values are computed only if there are at least 50 days with information and consider the same days during the calibration period (2005-2018). Results correspond to the best parameter set for each objective function (colors). Notice that winter corresponds to JJA, while summer to DJF. Streamflow observations and remotely sensed variables are referred to as "reference" and symbolized with white dots. The normalization of ET and SM is computed as $X' = (X - X_{min})/(X_{max} - X_{min})$.

4.6.4 Limitations

In this study, a single satellite remote sensing dataset is used for SM_1, LST, and fSCA (i.e., uncertainties arising from the choice of an observational reference product were not analyzed). Further, the proposed methodology regularized one parameter at a time, while maintaining the rest spatially constant. However, two or more parameters could be simultaneously regularized (e.g., Mendoza et al., 2012, 2016), although that would increase the number of (super)parameters to be calibrated (see equation in Figure 4.2; Pokhrel et al., 2008). In this regard, a sensitivity analysis of VIC model parameters focused on spatial patterns would be useful to reduce the number of (super)parameters to calibrate, though such analysis is beyond the scope of this work.

This study examined the effects of streamflow-only calibration of the VIC model on the spatial patterns and annual cycle of other variables than Q. Although this work conducted additional calibration experiments adding more variables to the OF, no systematic assessment that includes all variables across basins was conducted since the associated benefits

have been widely discussed in the literature (e.g., López López et al., 2017; M. C. Demirel et al., 2018; Koppa et al., 2019; Dembélé, Ceperley, et al., 2020; Tong et al., 2021; Shah et al., 2021; Pool et al., 2024).
Although a single model structure is used, the analyses presented here could be extended to other lumped or semi-distributed models with varying degrees of complexity, taking advantage of recently developed modular modeling platforms (e.g., Clark et al., 2015; Coxon et al., 2019; W. J. M. Knoben et al., 2019; Craig et al., 2020), to address well-known streamflow equifinality issues arising from the compensation of parameters, state variables, and fluxes (Khatami et al., 2019).
The assessment of models in which the water available for vegetation transpiration comes from different soil layers may benefit from including satellite products that provide estimates of total water storage, such as GRACE (Güntner, 2008; Werth et al., 2009; Soltani et al., 2021) to complement other sources of information regarding water storages. However, the incorporation of such observational products in the study domain considered in this work remains challenging due to their relatively coarser horizontal resolution. Future work could explore the potential of statistically downscaled datasets (e.g., Rakovec et al., 2016; Yin et al., 2018; Vishwakarma et al., 2021; Fatolazadeh et al., 2022; Khorrami et al., 2023; Mei et al., 2023) for the calibration/evaluation of hydrological models across mountainous catchments.

4.7 Conclusions

Despite the tremendous advances in the development of satellite remote sensing products that provide information of hydrological variables in space and time, streamflow-only calibration of distributed hydrological models remains a popular practice. To identify drawbacks for simulating spatial patterns and annual cycles of other variables than streamflow, this work performed several streamflow-only calibration experiments with the VIC model in six basins located in continental Chile, testing the impacts of regularizing different soil parameters and the choice of calibration objective function. The main conclusions are as follows:

- For a given streamflow-based OF, most of the spatial regularization strategies tested in this study provide similar OF values, with varying (either positive or negative) effects on the realism of simulated spatial patterns of other variables. None of the spatial model configurations tested here was able to provide simultaneous improvements in the spatial patterns and annual cycles of the variables evaluated across all catchments.
- The improvements in spatial patterns via spatial regularization techniques do not guarantee a correct simulation of ET and fSCA annual cycles.
- For the VIC model, ET, SM_1, LST, and fSCA simulations do not necessarily improve when Q performance improves.
- For water balance characterizations, evaluating simulated annual cycles of ET (besides Q) is required to detect model or data deficiencies and, therefore, decide on the need to incorporate these variables within the calibration process.
- In model structures where several soil layers contribute to transpiration, differences in simulated annual cycles of soil moisture - obtained from different parameter con-

figurations - may explain discrepancies in ET annual cycles, which, in turn, may provide very similar streamflow seasonalities (an extreme case of flux equifinality).

Chapter 5

When Streamflow isn't Enough: Misrepresenting Evapotranspiration Simulations alters Hydrological Projections

Summary

Hydrological model calibration and evaluation are crucial steps to understand and test model robustness under different hydroclimatic conditions and to produce robust hydrological projections. Despite the increasing number of datasets for different hydrological variables, the current modeling practice heavily relies on streamflow data for model calibration, assuming that accurate streamflow (Q) performance ensures reliable simulations of ET and/or other variables. This study explores the impact of including evapotranspiration (ET) in the calibration of a distributed hydrological model on (i) seasonal water balances, and (ii) hydrological projections under the SSP5-8.5 climate scenario in continental Chile. The Variable Infiltration Capacity (VIC) model is calibrated in 120 basins with near-natural regimes in Chile, using a Q-only objective function and a calibration metric that combines Q and ET. Hydrological projections are obtained using statistically downscaled and bias-corrected outputs from six Global Climate Models (GCMs) spanning from 1970 to 2100. Other key findings are: (i) the inclusion of ET in calibration considerably influences hydrological projections; (ii) different parameter sets achieving similar streamflow responses can provide very different ET estimates, affecting historical and projected seasonal water balances; and (iii) the basins sensitive to the inclusion of ET are primarily located in humid regions, with high annual precipitation amounts and low aridity indices. This study highlights the need to integrate both Q and ET in hydrologic model evaluation to better capture water balance seasonality and improve the reliability of climate change impact assessments.

5.1 Introduction

Hydrological models are commonly used to generate projections under land use/cover and climate change scenarios. In this context, estimating hydrological change estimates requires several decisions, such as projected conditions (e.g., greenhouse emissions scenarios, global climate models, land cover, among others), selecting (one or several) hydrological models and, usually, calibrating the model parameters by using an optimization algorithm and an objective function (Wilby & Dessai, 2010; Clark et al., 2016; Duethmann et al., 2020; Zaitchik et al., 2023). In this regard, the objective function summarizes the model performance of key hydrological variables, and its choice is critical for hydrological projections. Moreover, the evaluation of simulated hydrological processes, such as evapotranspiration or snow, can provide insights on the suitability of hydrological model structure and parameters for changing hydroclimatic conditions (e.g., Merz et al., 2011; K. Fowler, Coxon, et al., 2018; K. Fowler, Peel, et al., 2018; Duethmann et al., 2020), although identifying the effects of choosing different hydrological models and its parameters on hydrological projections remains challenging (K. Beven & Binley, 1992; K. Beven, 2012; Her & Chaubey, 2015; Khatami et al., 2019; K. J. Beven & Chappell, 2021; Muñoz-Castro et al., 2023). Streamflow-based calibration relies on the assumption that a reasonable streamflow (Q) performance implies a reasonable performance in other variables. Nevertheless, the evaluation process is essential if the aim is to ensure model fidelity (Refsgaard, 1997; Kirchner, 2006; M. C. Demirel et al., 2018; Dembélé, Ceperley, et al., 2020; Jiménez-Navarro et al., 2024). Calibration and/or evaluation of simulated evapotranspiration (ET; e.g., M. C. Demirel et al., 2018; Dembélé, Hrachowitz, et al., 2020), soil moisture (SM; e.g., Tong et al., 2021; Bajracharya et al., 2023), total water storage (TWS; e.g., Werth & Güntner, 2010), snow cover (Parajka & Blöschl, 2008; Duethmann et al., 2014; Bennett et al., 2019; Tong et al., 2021; Tang et al., 2023), snow water equivalent (SWE; e.g., Avanzi et al., 2020), land surface temperature (LST; e.g., Zink et al., 2018), and groundwater levels (GWL; e.g., Refsgaard & Knudsen, 1996; Yáñez-Morroni et al., 2023), among other variables, have become more frequent given the increasing number of available datasets (e.g., remotely-sensed images and products; McCabe et al., 2017). Further, several metrics have been proposed to assess the simulation of these variables (M. C. Demirel et al., 2018; Tong et al., 2021; Dembélé, Ceperley, et al., 2020; Dembélé, Hrachowitz, et al., 2020). However, overconfidence in different metric (including streamflow-based metrics) values could be misleading and hide pitfalls in hydrological simulations (Clark et al., 2021; Cinkus et al., 2023).

The evaluation of hydrological models is a critical step for understanding the predictive capabilities of the model (KlemeŠ, 1986; K. Beven & Binley, 1992; Refsgaard & Knudsen, 1996; Refsgaard, 1997; Cinkus et al., 2023; Keller et al., 2023), advocating for "good" simulations compared to reference datasets and the conceptual model for the right reasons (Kirchner, 2006). Further, the dependence of model performance on the calibration/evaluation periods (e.g., Merz et al., 2011; Duethmann et al., 2020) may preclude the reliability of hydrological projections under climate change scenarios, since future climate may largely differ from the period used for estimating the model parameter (e.g., Saavedra et al., 2022). The evaluation process usually involves the contrast against different parts of the hydrograph (Yilmaz et al., 2008; Gupta et al., 2009; Cinkus et al., 2023) and/or the model's capability to "successfully" simulate hydrological conditions under different climate conditions (K. Fowler, Coxon, et al., 2018; Duethmann et al., 2020). Such analyses are, in general, focused on streamflow only (Moriasi et al., 2007; Gupta et al., 2009; Westerberg et al., 2011; Garcia et al., 2017; K. Fowler, Coxon, et al., 2018;

K. Fowler, Peel, et al., 2018; Schwemmle et al., 2020), under the assumption that verifying streamflow-only suffices the evaluation of (at least) the simulated water balance.

Despite recent progress in the assessment of the models' capacity to simultaneously replicate several hydrological variables (model fidelity), the calibration of hydrological models to estimate climate change impacts on water resources strongly relies on only streamflow-based calibration and evaluation (e.g., Shafii & Tolson, 2015; L. Melsen et al., 2016; Mizukami et al., 2017, 2019; Beck et al., 2020; Clark et al., 2021; Aguayo et al., 2021; S. Wang et al., 2022), with unfavorable consequences on streamflow projections (e.g., McInerney et al., 2024). Moreover, determining the possible effects of not verifying relevant hydrological processes on hydrological projections under climate change remains challenging (Mendoza, Clark, Mizukami, et al., 2015; Mendoza et al., 2016). Hence, the the potential effects that streamflow-only calibration may have on projected changes in hydrological states and fluxes remain unknown.
Although several variables are simulated by hydrological models, this work focuses most of the analysis on simulated ET and Q since both control the overall water balance at the climatological temporal scale. Therefore, the following questions are addressed:

1. How does the calibration/representation of average annual cycles of Q and ET affect the projections of these variables under climate change scenarios?
2. What types of catchments are more sensitive (i.e., projections with greater variability) to the inclusion of ET in the calibration process?

To address the above questions, the Variable Infiltration Capacity model (VIC; Liang et al., 1994) is calibrated using (i) a Q-only objective function and (ii) a combination of Q and ET. Although the focus is on the role of ET in model fidelity, the quality of simulated LST, fractional snow-covered area (fSCA), and SM for the upper soil layer are also assessed. The analyses are conducted in basins with near-natural regimes located along continental Chile. To the best of the author's knowledge, hydrological projections over this domain have relied on streamflow-based calibrations using a variety of hydrological models in specific basins (Vicuña et al., 2011; Demaria et al., 2013; Vicuña et al., 2013; Bozkurt et al., 2017; DGA, 2017; Aguayo et al., 2019, 2021; DGA, 2022). For hydrological projections, six Global Climate Models are statistically downscaled and bias-corrected, leading to hydrological projections from 1970 to 2100. In this chapter the relative importance of selecting between two objective functions (Q-only and Q+ET) is assessed. Moreover, the choice of GCMs is also incorporated since previous work suggests that GCMs are among the main drivers explaining the spread in hydrological projections. This study complements previous work aimed to examine modeling decisions to improve process representations (Sepúlveda et al., 2022; Murillo et al., 2022; Cortés-Salazar et al., 2023) and climate change projections (Gateño et al., 2024; Vásquez et al., 2024) in continental Chile.

5.2 Study Area

The study domain encompasses 120 basins located in continental Chile (Figure 5.1), which meet the following requirements: (i) a low human intervention degree index (i.e. $< 5\%$), which is defined as the ratio between the annual flow of surface water rights (consumptive

permanent continuous), and the mean annual runoff measured at the catchment outlet (Alvarez-Garreton et al., 2018), (ii) absence of lakes or reservoirs, and non-consumptive water withdrawals, unless their restitution is located upstream of the streamflow gauge and (iii) glacier coverage smaller or equal than 2% of the total basin area. Almost all the basins drain from the Andes Cordillera (east) to the Pacific Ocean (west). The basin areas range from 20.6 km^2 to 5,152 km^2, and most of them are located in central (25-35°S) and southern (35-42°S) Chile, where the majority of the population lives.

Overall, the selected basins span a wide range of hydroclimates with snow-dominated, mixed, and rainfall-driven hydrological regimes. Catchment-scale mean annual precipitation (P) ranges between 117-3,246 mm/year, with northern Chile being the most arid region, while central and southern Chile are the wettest areas. Snowfall plays a key role in the hydrology of the study domain, particularly between 30°S and 40°S due to the elevation of the Andes Cordillera, and in southern Chile (<40°S) due to the cold winter months. The snowfall fraction, estimated as the ratio between mean annual snowfall (S_f) and P, reaches its highest value between 30°S and 40°S. In basins located on the coastal mountain ranges, the relevance of snowfall is negligible. The precipitation seasonality index (δ_p^*; Woods, 2009) indicates the season that concentrates most of the annual precipitation, with $\delta_p^* \approx -1$ ($\delta_p^* \approx 1$) when precipitation amounts are concentrated in winter (summer). In the study domain, winter and summer correspond to JJA and DJF months, respectively. Figure 5.1d displays the basin-scale δ_p^*, showing a clear north-to-south gradient. In the Altiplano Region (20°S), the rainy season occurs during the austral summer, whereas the wet season occurs during the austral winter south of 28°S. Precipitation becomes progressively more uniform throughout the year as one moves southward (<45°S), with δ_p^* reaching values close to zero south of 50°S. Figure 5.1e shows the aridity index, which quantifies the ratio between the atmospheric energy demand (estimated through the potential evapotranspiration; PET) and the annual precipitation (i.e., PET/P). AI values close to 1 in continental Chile are located around 35°S (excepting some basins south of 50°S), separating the arid northern Chile, where AI can be ≥4, from the (more humid) southern half of the country, where AI can be as low as ≈0.2).

5.3 Datasets

5.3.1 Historical Hydrometeorological datase

Reference daily precipitation and maximum (T_{max}) and minimum (T_{min}) air temperature are obtained from the CR2MET v2.5 dataset (Boisier et al., 2018; DGA, 2022; Boisier, 2023), which covers continental Chile with a horizontal resolution of 0.05° x 0.05° for the period 1979-2020. CR2MET P estimates are obtained through a two-step approach consisting of (i) computing the probability of precipitation at each grid cell through logistic regression models, and (ii) calculating daily precipitation amounts using multiple linear regression equations. All the models use ERA5 reanalysis outputs (Hersbach et al., 2020) and geomorphological attributes as predictors, and daily precipitation from meteorological stations as predictands. For T_{max} and T_{min}, land surface temperature from MODIS is also included as a predictor.
To obtain sub-daily meteorological time series, CR2MET daily precipitation, and temperature are disaggregated into hourly time steps using the sub-daily distribution provided by ERA5-Land (Muñoz Sabater, 2019), which is bias corrected to match the CR2MET

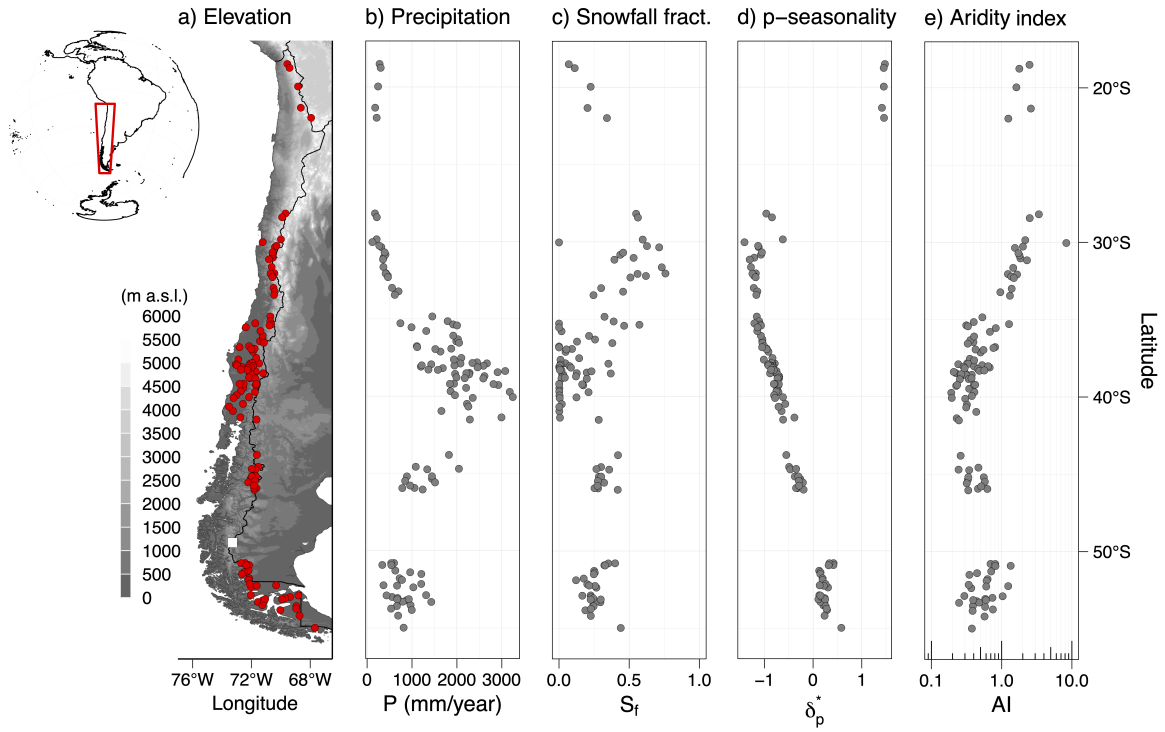

Figure 5.1: a) Location of the catchments in near-natural regime considered in this study. b) Mean annual precipitation, c) snowfall fraction, d) p-seasonality index, and e) aridity index. The values are computed for the period April/1985-March/2015 (30 WYs). Note that the x-axis is in logarithmic scale for the aridity index (panel e).

daily values. Relative humidity (RH), wind speed (W), atmospheric pressure (AP), and incoming shortwave radiation (K_{in}) are derived for the same horizontal resolution grid by spatially interpolating ERA5-Land outputs. ERA5-Land wind speed is bias-corrected using wind simulations from WRF at a 1-km horizontal resolution (Geophysics Department & Ministerio de Energía, 2018). Incoming longwave radiation (L_{in}) is computed with the parameterization proposed by Iziomon et al. (2003) using the bias-corrected hourly temperature. To conduct hydrological model simulations, all the meteorological variables are grouped to obtain 3-hourly time series.
Daily streamflow data is obtained from stations maintained by the Chilean Water Directorate (DGA, in Spanish), also available in the CAMELS-CL dataset (Alvarez-Garreton et al., 2018).

5.3.2 GCM data

For climate projections, six CMIP6 GCMs are considered (Table 5.1) under the emission scenario SSP5-8.5. The GCMs were selected based on the results of Gateño et al. (2024), who evaluated the models' capability to replicate climate indices derived from precipitation and temperature for the period April/1985-Mar/2014, 30 water years (WY), including: (i) mean annual bias, (ii) mean annual cycle (iii) interannual variability, (iv) spatial correlation of mean annual values, (v) distribution of monthly values and (vi) model genealogy. The historical performance for each model, variable and macrozone was summarized using the Past Performance Index (PPI; see Gateño et al., 2024 for details). In this chapter, only GCMs holding the top 60% PPI values in at least four of the five

macrozones are selected.

Table 5.1: CMIP6 GCMs considered in this study. Δlon and Δlat correspond to the GCM's horizontal resolution.

GCM	Δlat	Δlon	Institution
CanESM5	2.77	2.81	Canadian Centre for Climate Modelling and Analysis, Canada.
CMCC-ESM2	0.94	1.25	Euro-Mediterranean Centre on Climate Change Coupled Climate Model, Italy.
EC-Earth3-CC	0.70	0.70	EC-Earth Consortium, Europe.
EC-Earth3-Veg-LR	1.12	1.13	
GFDL-ESM4	1.00	1.25	Geophysical Fluid Dynamics Laboratory, USA.
KIOST-ESM	1.88	1.88	Korea Institute of Ocean Science and Technology Earth System Model and Its Simulation Characteristics, South Korea.

5.3.3 Remote sensing products

In this chapter, remotely sensed fractional snow-covered area, actual evapotranspiration, land surface temperature, and soil moisture are used for hydrologic model evaluation. The fractional snow-covered area is derived from MODIS products (MOD10/MYD10; Hall & Riggs, 2016). To obtain a unique time series, daily MOD10 and MYD10 estimates are averaged at each MODIS grid cell. If only one product is available for a specific day and grid cell, that value is used to estimate fSCA. All the gaps (i.e., days and grid cells where MOD10 and MYD10 are not available) are filled at the original horizontal resolution (500 m) using the methodology proposed by Cornwell et al. (2016). Actual evapotranspiration is obtained from the MOD16 product (Mu et al., 2011) using 8-day estimates with a 1-km horizontal resolution. Land surface temperature for each day and each grid cell is obtained as the average between the MOD11 and MYD11 products (Wan, 2014) at a 1-km horizontal resolution. If one of the two LST estimates is unavailable, the day is considered missing. For soil moisture (0.25° horizontal resolution), the ESA-CCI product is used (Dorigo et al., 2017).

To resolve the mismatch between the horizontal grid of remote sensing products and the grid discretization of CR2MET (which is the same as in the hydrological model), ET, LST, and fSCA are upscaled using spatial averages, and (SM is downscaled to the CR2MET grid cell that is closest; e.g., dos Santos Araujo et al., 2024). Grid cells without data are not filled out for ET, LST, and SM, as opposed to fSCA.

5.3.4 Ancillary data

Since this study considers a distributed hydrological model that requires *a priori* parameter fields, the SoilGrids dataset (Poggio et al., 2021) is used to derive mean clay and sand content, and mean bulk density for the first 2 m soil depth at a 250-m horizontal resolution. The elevation is estimated from the Shuttle Radar Topography Mission (SRTM; Farr et al., 2007), while the aspect is derived from the SRTM elevation raster at the original horizontal resolution (30 m). All datasets are upscaled to match the CR2MET grid using spatial averages.

5.4 Modeling approach

5.4.1 Hydrological model

In this chapter, the physically-based and semi-distributed VIC model (Liang et al., 1994) is used to simulate state variables and fluxes at a 0.05°x 0.05° horizontal resolution. In VIC, precipitation is partitioned into snowfall or rainfall, and both can be stored in the canopy. The maximum amount of water that the canopy intercepts is estimated using the Leaf Area Index (LAI; Dickinson, 1984). The snowpack is represented by two layers, where the top and bottom layers are used for atmosphere and ground energy balance computations, respectively (Andreadis et al., 2009). The soil column has three layers, with the first one controlling infiltration, and the third layer simulating baseflow processes. To compute infiltration, VIC uses the Xinanjiang formulation (Zhao et al., 1980), assuming that the infiltration capacity varies within the simulation unit area (Wood et al., 1992). Excess runoff (R) is generated in those areas where precipitation exceeds the moisture storage of the first soil layer. VIC assumes that gravity drives drainage, using the formulation proposed by Brooks and Corey (1964). In this regard, water enters the cell only from the atmosphere, i.e., VIC does not consider lateral fluxes among grid cells. Baseflow (B) is generated in the third (deepest) soil layer using a formulation proposed by Franchini and Pacciani (1991). It is assumed that Q = R + B. The reader is referred to Liang et al. (1994) for more details. Horizontal heterogeneity in each grid cell can be considered by incorporating different land cover types. Here, the International Geosphere-Biosphere Program (IGBP) classification for the year 2010 from the MCD12Q1 v006 land cover product (Sulla-Menashe & Friedl, 2018) is used to represent all land cover types spanning at least 2% of each grid cell area. Mean monthly LAI values for these land cover types are derived from the MOD15A2 product. Soil Bulk density is estimated as the vertical average from the first 2 m soil depth, obtained from the SoilGrids product (Poggio et al., 2021). VIC is coupled with the mizuRoute model (Mizukami et al., 2016) to estimate the streamflow at the outlet of each basin. mizuRoute first performs a hillslope routing using a gamma-distribution-based unit-hydrograph to delay VIC runoff, and then routes the delayed runoff for each river reach defined by the river network topology. For river routing, the diffusive wave scheme is used, following the setup recommended by Cortés-Salazar et al. (2023). Manning's roughness coefficient and riverbed width are derived for all basins using relationships derived from Mendoza et al. (2012) and Niño (2002). Each grid cell considered elevation bands delineated with a 200 m vertical discretization (Murillo et al., 2022).

5.4.2 Calibrated parameters

In this chapter, the VIC parameters calibrated are the ones identified as the most sensitive by Sepúlveda et al. (2022), including: (i) b_{infilt}, which controls the infiltration process, (ii) Ds and Ws, which control the linearity of the curve used to calculate the water leaving the third (deepest) soil layer, (iii) Ds_{max} as the maximum baseflow rate, (iv) hydraulic conductivity (K_{sat}), related to how water percolates between the soil layers, (iv) the soil layers' depths, (v) the threshold temperature to separate precipitation into rainfall and snowfall, (vi) maximum snow albedo and (vii) its temporal decay rate.
The parameters regularized are b_{infilt} and the soil layers' depths. To define *a priori* spatial distribution of these parameters, a Principal Component Analysis (PCA) is conducted

considering a suite of grid cell attributes within each basin, including bulk density, clay and sand content, elevation, and slope (e.g., Samaniego et al., 2010; Mizukami et al., 2017; Beck et al., 2020). The *a priori* map is based on the first component (PC1) of the PCA analysis as well as super-parameters (as in step 2 in Figure 4.2; Pokhrel et al., 2008)

5.4.3 Performance metrics

Streamflow-based performance metrics

The Kling-Gupta efficiency (KGE; Gupta et al., 2009; Kling et al., 2012) is used to inform the performance of streamflow simulations. KGE seeks to minimize the Euclidean distance between performance metrics related to volume (β), variability (γ) and their optimal values (i.e., 1):

$$KGE = 1 - ED = 1 - \sqrt{(1-\beta)^2 + (1-\gamma)^2 + (1-r)^2} \tag{5.1}$$

where $\beta = \mu_s/\mu_o$ is the ratio between simulated (s) and observed (o) average values (μ); $\gamma = (\sigma_s/\mu_s)/(\sigma_o/\mu_o)$ with σ being the standard deviation and r is the temporal correlation between observed and simulated daily values. KGE ranges from $-\infty$ to 1, where 1 represents a "perfect" model simulation (i.e., ED = 0).

Spatial pattern performance metrics

This chapter uses the spatial efficiency metric (Eq. 5.2) proposed by Dembélé, Ceperley, et al. (2020) to evaluate simulated ET, SM, LST, and fSCA.

$$E_{SP} = 1 - \sqrt{(1-r_s)^2 + (1-\gamma)^2 + (1-\alpha)^2} \tag{5.2}$$

Where r_s is the Spearman correlation coefficient between simulated and reference values, γ is the ratio between the coefficients of variation (as in KGE; Eq. 5.1), $\alpha = 1 - RMSE(Z_{X_s}, Z_{X_o})$, being Z the time series with standardized values for the variable X (notice that this chapter follows the same notation as in Dembélé, Hrachowitz, et al. (2020)) and RMSE the Root Mean Square Error. The standardization aims to avoid a direct contrast between the model and reference values since remotely sensed products have biases that could affect the metric. However, bias-accounting metrics, such as KGE and RMSE, are also considered to assess the model's ability to reproduce the raw remotely sensed estimates.

Following Dembélé, Ceperley, et al. (2020), this chapter uses two approaches to contrast spatially distributed simulations against observations. Option 1 considers the contrast between simulated and reference maps of a variable X at each time step, resulting in a time series of performance measures ($OF_X(t)$) that can be used to compute a summary metric OF_X^{time} by temporally averaging $OF_X(t)$ values. Option 2 involves the comparison between simulated and reference time series at each grid cell to obtain a map of performance metrics ($OF_X(i)$) that can be used to calculate a summary performance measure OF_X^{space} by spatially averaging $OF_X(i)$ values.

For soil moisture, the correlation coefficient (Eq. 5.3) between simulated (θ_s) and reference soil moisture (θ_{ref}) is included as in Tong et al. (2021).

$$O_{SM} = \frac{\sum_{t=1}^{T} \left[(\theta_s^t - \overline{\theta}_s)(\theta_{ref}^t - \overline{\theta}_{ref})\right]}{\sqrt{\sum_{t=1}^{T} \left[(\theta_s^t - \overline{\theta}_s)^2(\theta_{ref}^t - \overline{\theta}_{ref})^2\right]}} \tag{5.3}$$

Where t represents the time step and the overbars represent average values. This chapter only considers the simulated soil moisture from the first soil layer since the ESA-CCI soil moisture product considers the top 0.5–2 cm soil depth. To overcome mismatches and systematic biases between the reference and SM_1 due to different soil depths and properties, mean–standard deviation matching (Draper et al., 2009; López López et al., 2017) was used (Eq. 5.4).

$$\theta_s^* = \frac{\sigma_{\theta_{ref}}}{\sigma_{\theta_{sim}}} \cdot \left(\theta_s - \overline{\theta}_s\right) + \overline{\theta}_{ref} \tag{5.4}$$

Where $\sigma_{\theta_{ref}}$ and σ_{θ_s} are the standard deviation of the reference and simulated soil moisture. Finally, the quality of fSCA simulations are assessed by using the root mean squared error (RMSE) and the metric O_{SC} proposed by Tong et al. (2021):

$$O_{SC} = 1 - (S_O + S_U) \tag{5.5}$$

where S_O and S_U represent the fraction of the total number of simulated days with overestimation and underestimation of simulated fSCA, respectively. To avoid noisy values, a threshold of 1% is applied; hence, a grid cell with simulated and reference fSCA values below 0.01 is considered snow-free.

Only for graphical analyses, ET and SM annual values are normalized as $X^* = (X - X_{min})/(X_{max} - X_{min})$ to avoid mismatches and biases in the reference product and focus on the simulated annual cycle. This chapter uses the notation SM_j, to refer to the moisture content in the j-th soil layerof the VIC model.

Additionally, the model's capability to replicate the average seasonality of Q, LST, fSCA, SM_1, and ET during the calibration period is evaluated. To this end, for each basin and variable, the Pearson correlation coefficient between simulated and reference average monthly values (12 vs. 12 values) is computed.

Objective functions to obtain VIC parameters

Because previous analyses show that incorporating ET in the calibration could be crucial to adequately simulate the average water balance seasonality (e.g, Jiménez-Navarro et al., 2024), this chapter considers an objective function (OF), combining Q and ET, as shown in Eq. 5.6:

$$OF = 1 - \sqrt{s_Q \cdot [1 - E(Q)]^2 + s_{ET} \cdot [1 - E_{SP}(ET)]^2} \tag{5.6}$$

where s_Q and s_{ET} allow weighing differently the model performance for Q and ET, respectively. Further, to give more weight to low flows, this chapter considers the streamflow performance metric, E(Q), proposed by Garcia et al. (2017) and shown in Eq. 5.7:

$$E(Q) = \frac{KGE(Q) + KGE(1/Q)}{2} \tag{5.7}$$

The VIC model parameters are obtained by optimizing two OFs: (i) a traditional calibration approach considering only streamflow (i.e., $s_{ET} = 0$ in Eq. 5.6) and (ii) by including Q and ET simultaneously. This chapter refers to both OFs as OF_{Q} and $OF_{Q\&ET}$ when only Q and Q and ET are calibrated, respectively. Hence, two "best parameter" sets are obtained (one for each OF) to verify the impacts of including ET in the calibration.

The individual basin calibrations were conducted by maximizing the OFs using the Dynamically Dimensioned Search (DDS; Tolson & Shoemaker, 2007) algorithm, implemented within the OSTRICH software (Matott, 2017). For each OF, DDS is initialized from six different and randomly selected parameter sets, with a maximum search number of 500 iterations. Hence, the model is run with 3,000 different parameter sets, selecting the one that yields the best OF value. All the performance metrics considered are computed at a daily time step. The calibration period is Apr/2005-Mar/2018, while the evaluation period is Apr/1995-Mar/2005. The calibration period was selected based on the diversity of normal, wet and, in particular, dry WYs, including seven WYs of the central Chile mega-drought (R. D. Garreaud et al., 2017; Boisier et al., 2018; R. D. Garreaud et al., 2020).

5.4.4 Downscaling and bias correction of climate models

To obtain hydrological projections under climate change scenarios, all the models were statistically downscaled to the CR2MET horizontal resolution using an inverse distance weighting interpolation method (e.g., Demaria et al., 2013). Then, SD GCM outputs were bias-corrected at each CR2MET grid cell using a multivariate BCM to better replicate the precipitation-temperature correlation in arid or temperature areas with considerable snowfall fractions (Meyer et al., 2019; Guo et al., 2020), such as northern and central Chile. The Multivariate quantile-mapping bias correction method (MBCn, Cannon, 2018) is used, and can be found in the "MBC" library (Cannon, 2016, 2018), implemented in the statistical software R (R Core Team, 2023). The MBCn method relies on three steps that are repeated until convergence: (i) an orthogonal rotation, (ii) the application of the Quantile Delta Mapping method (Cannon et al., 2015) and (iii) the application of an inverse matrix (the one used to compute the orthogonal variables). The MBCn method is applied to daily time series of P, T_{max}, and T_{min} using a monthly temporal stratification and considering 1980-2014 as the training period (35 years). As a result, daily bias-corrected time series of P, T_{max}, and T_{min} are obtained for the 1980-2100 period.
The remaining variables required to run the VIC model (AP, w, RH, K_{in} and L_{in}) are obtained by searching, for each day t, the most similar day in terms of P, T_{max} and T_{min}. Such day is found by using a Euclidean distance approach between the day to fill (from SDBC GCM; t) and the CR2MET dataset ($t_{closest}$). This process is conducted for each day t at each grid cell and GCM. When the "closest" day is identified, the remaining variables are completed by copying the values from day $t_{closest}$ to t.

In order to compare how the choice of OF can affect annual and seasonal projected changes in Q and ET for a specific GCM, the "best" two GCMs according to the criteria detailed in section 5.3.1, are selected. The number of GCMs selected is referential and selected just for graphical purposes.

5.4.5 Analysis framework

Hydrological projections are obtained by forcing the VIC model with the downscaled and bias-corrected GCM outputs at each grid cell, considering the two parameter sets obtained from the optimization of OF_Q and $OF_{Q\&ET}$. Hence, for each basin, 12 hydrological projections are obtained (two parameter sets and six GCMs).

In this study, the total dispersion (TD) in hydrological projections is derived from two methodological decisions: (i) the number of GCMs, and (ii) the choice of the objective function. To evaluate the relative importance of each decision in the spread of hydrological projections, an Analysis of Variance (ANOVA) is performed;

$$TD = OF + GCMs + Residuals \tag{5.8}$$

The ANOVA analysis is applied to changes in Q and ET at the annual and seasonal time scales. Since VIC provides spatially distributed hydrological projections, this chapter evaluates, for each basin and GCM, if the spatial distribution of projected changes in Q and ET are statistically different between the OFs by applying a Kolmogorov-Smirnov test. Additionally, temporal changes are statistically verified by using the Wilcoxon-Mann-Whitney test.

5.5 Results

First, this section presents results for historical simulations, focusing on average seasonalities. Then, the analysis continues with projected hydrological changes under a future climate scenario.

5.5.1 Historical simulations

Figures 5.2a and 5.2b display the KGE(Q) and KGE(1/Q) values for the calibration period (Apr/2005-Mar/2018), obtained with parameters sets that maximize OF_Q and $OF_{Q\&ET}$. KGE(Q) values are more stable between OF and periods, while KGE(1/Q) yields larger differences between periods. For KGE(Q), most differences arise between the calibration and evaluation periods, although for higher KGE(Q) values, a calibration based on Q only leads to slightly better results compared to $OF_{Q\&ET}$ simulations. In this regard, 70% of the basins have KGE(Q) ≥ 0.3 during the calibration period. For KGE(1/Q), despite larger differences between calibration and evaluation periods, differences arising from the choice of objective function are smaller. The results show that higher (better) metrics are obtained, mostly in central and southern Chile (30°S-45°S; Figures 5.2a and 5.2b), where the climate is humid (P>1,000 mm/year and AI $\lesssim 0.5$; see Figure 5.1), and most of the annual precipitation occurs in fall and winter ($\delta_P^* \lesssim 0.5$). Figure 5.2c shows the empirical cumulative distribution of KGE(Q) and KGE(1/Q) values for the calibration and evaluation periods, and the best parameter sets from both objective functions.

Figure 5.3 displays the correlation for each variable and objective function. The best results are obtained for LST, followed by fSCA and Q, which is expected since the latter variable is included in both OFs. SM_1 and ET are the two more poorly simulated variables

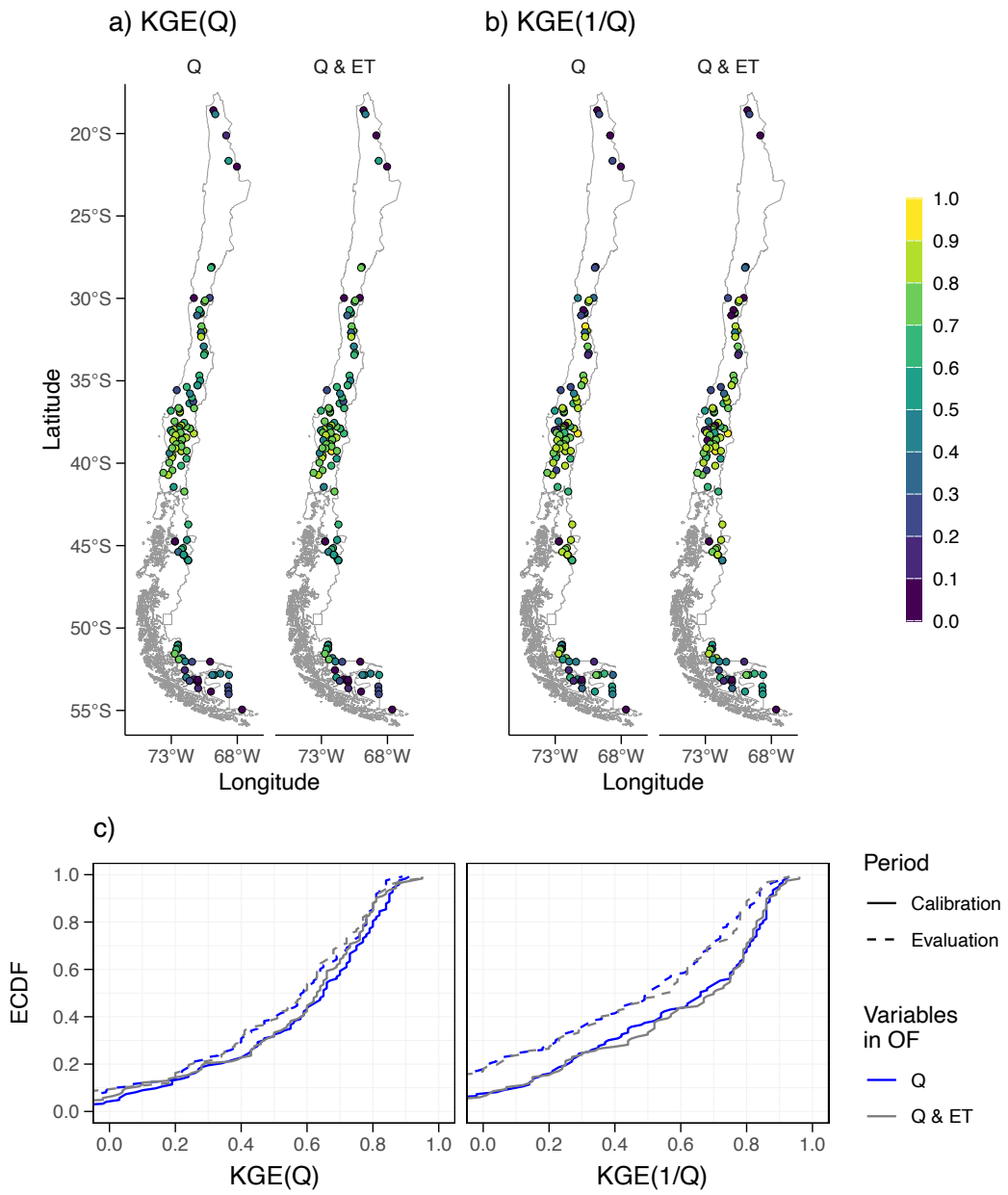

Figure 5.2: Spatial distribution of (a)KGE(Q) and (b) KGE(1/Q) for the best parameter set obtained by optimizing OF_Q (or Q) and $OF_{Q\&ET}$ (Q & ET). Panel c) displays the empirical cumulative distribution function (ECDF, y-axis) of the metrics obtained for the 120 calibrated basins.

(particularly ET), which can be explained by the dependence between SM_1 and ET from the parameters related to soil processes (calibrated to match streamflow observations).

Since the annual cycles of SM_1 and ET are poorly simulated, the subsequent analyses focus on these variables and Q. Figure 5.4 shows the simulated and reference average seasonalities during the calibration period for five basins where the annual cycle of Q is reasonably well simulated using parameter sets obtained from OF_Q and $OF_{Q\&ET}$, though simulated performance differs for SM_1 and ET. The selected basins are, from north to south, (i) Codpa River at Cala-Cala, (ii) Colorado River at Palos, (iii) Mulchén River at Mulchén, (iv) Allipén River at Los Laureles and (v) Prat River at Desembocadura. For simplicity, these basins are referred to using the number identifiers (i)-(v) in Figure

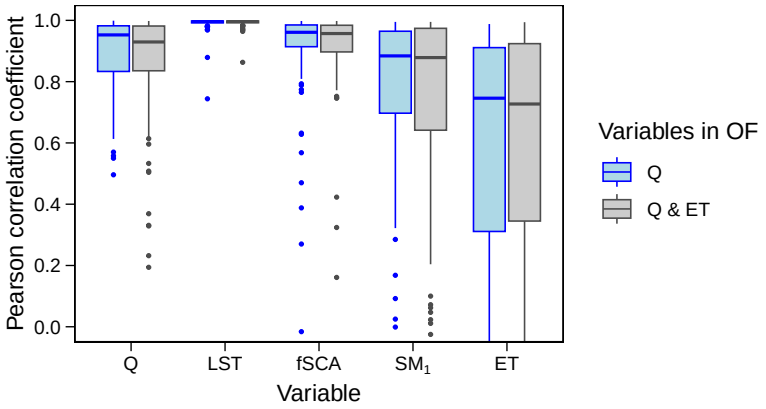

Figure 5.3: Pearson correlation coefficients between the simulated and reference average seasonality for the period Apr/2005-Mar/2018. Each boxplot comprises results for 120 basins and five variables (Q, LST, fSCA, SM, and ET), using parameter sets obtained by maximizing Q and ET+Q.

5.4. Basin (i) has a rainfall-driven regime, with higher streamflow values in the summer and fall seasons (JFM). In this basin, the VIC model overestimated Q in January and underestimated it in March. Moreover, no considerable differences arise from the choice of OF, and SM_1 and ET seasonalities are reasonably well simulated. In basin (ii), with a mixed-hydrological regime and a higher snow-driven influence, the two OFs provide similar annual cycles of streamflow, though $OF_{Q\&ET}$ yields a Q underestimation in December. In this basin, no relevant differences between OFs are obtained in simulated SM_1, although both overestimate the reference values during the Spring season (SON). Nevertheless, there are important differences in ET, with OF_Q producing higher values during winter and early spring (JJAS), though reference data suggest that seasonalities from both OFs are wrong, and the model can not capture the higher ET values observed in summer (DJFM). In basin (iii), with a rainfall-driven hydrological regime, reasonable simulations for Q, SM_1 and ET are obtained, regardless of the OF. The VIC model is able to replicate the Q annual cycle in basin (iv) - with a mixed hydrological regime and predominancy of rainfall-driven events - though winter and spring (JJASO) streamflow is underestimated for OF_Q. The parameter sets obtained with $OF_{Q\&ET}$ (OF_Q) overestimate (underestimate) SM_1 during the spring season (SON). However, $OF_{Q\&ET}$ yields reasonable simulations of ET, and inadequate ET performance for OF_Q. Finally, reasonable (poor) simulated annual cycles for Q and ET (SM_1) are obtained in basin (v), with a mixed hydrological regime.

5.5.2 Hydrological projections

This section explores the extent to which the choice of OF (i.e., OF_Q or $OF_{Q\&ET}$) may affect - for each combination of GCM and basin - the projections of Q and ET at the annual and seasonal scales, as well as the contribution of baseflow to Q. Then, the number of basins where the spatial distribution of projected Q and ET are statistically different due to the choice of OF is evaluated.

Figure 5.5 displays projected changes in catchment-scale annual P, Q, and ET. The location of these basins is shown in Figure 5.5a. In northern Chile (17°S-23°S), the choice of GCM introduces a large spread in projected precipitation changes, without a clear pro-

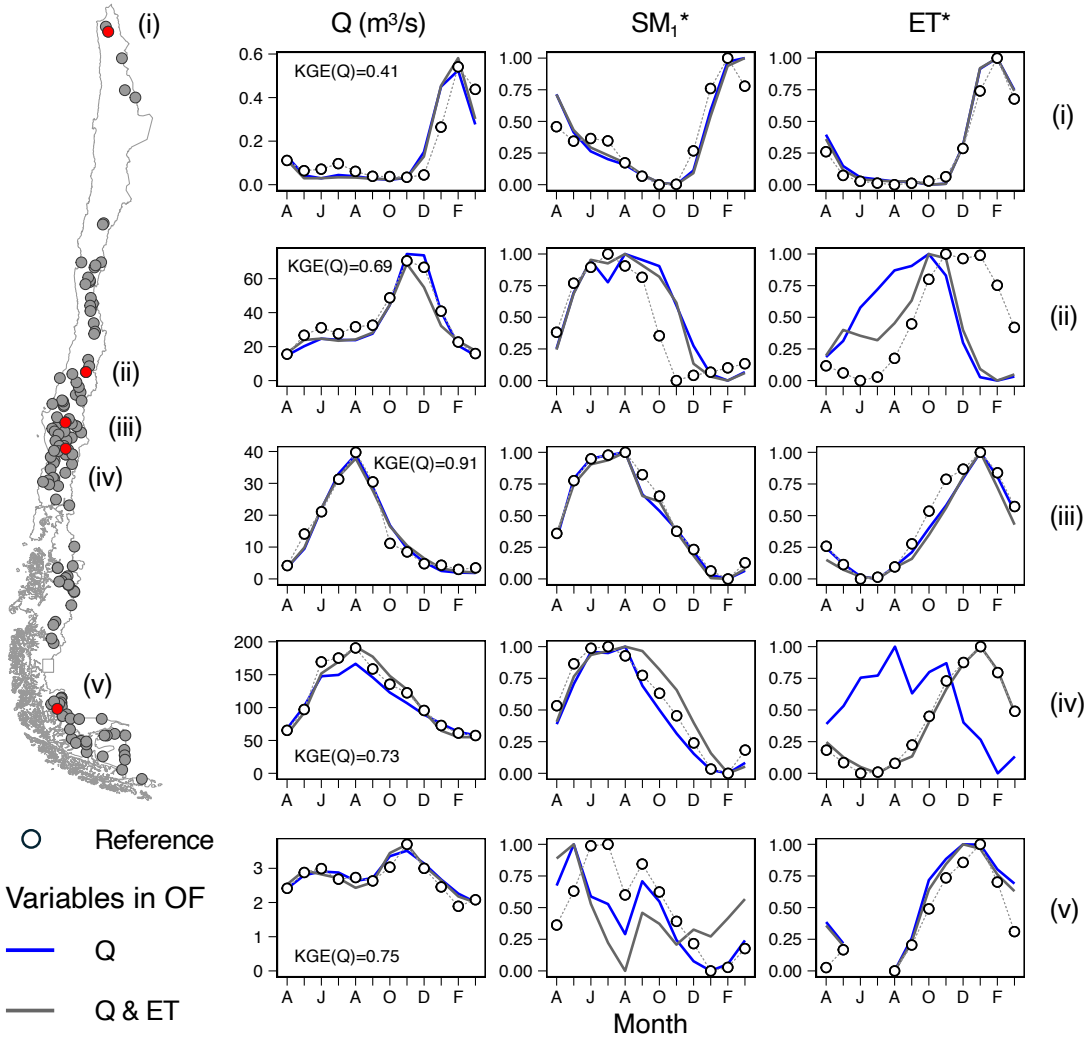

Figure 5.4: Average seasonality of Q, $SM_1{}^*$ and ET^* during the calibration period (Apr/2005-Mar/2018). White dots represent the reference (observational) datasets. The selected basins (location shown in the map) are (i) Codpa River at Cala-Cala (ID 1410004), (ii) Colorado River at Palos (ID 7112001), (iii) Mulchén River at Mulchén (ID 8330001), (iv) Allipén River at Los Laureles (ID 9404001), and (v) Prat River at Desembocadura (ID 12291001). Monthly values are computed for the exact days with both, simulated and reference data and only if more than 50 days with reference information are available.

jected change signal. In central and southern Chile (33°S-46°S), catchment-scale projected P changes suggest a decrease in annual amounts, whereas projected changes are relatively smaller (with respect to the historical period) in austral Chile (south of 50°S). The median of projected changes in Q (Figure 5.5d) follows a similar latitudinal pattern compared to ΔP (Figure 5.5b), although the projected changes are more pronounced than ΔP and the spread among GCMs is larger. For example, the median $\Delta P \approx -30\%$ between 35°S and 45°S, while $\Delta Q \approx -45\%$. From 20°S to 30°S, the results suggest an increase in Q by the end of the century (2070-2100 w.r.t. 1985-2015). From 20°S to 35°S, a general annual decrease in the median among GCMs is projected. Between 35°S and 43°S, an increase in annual ET is projected. South from 50°S, median projected changes in ET are mostly positive, though the choice of GCM introduces some spread that can switch the negative or positive signal of projections in some catchments. Nevertheless, the spread in the ET projected change signal suggests that the increase or decrease in annual ET values at the

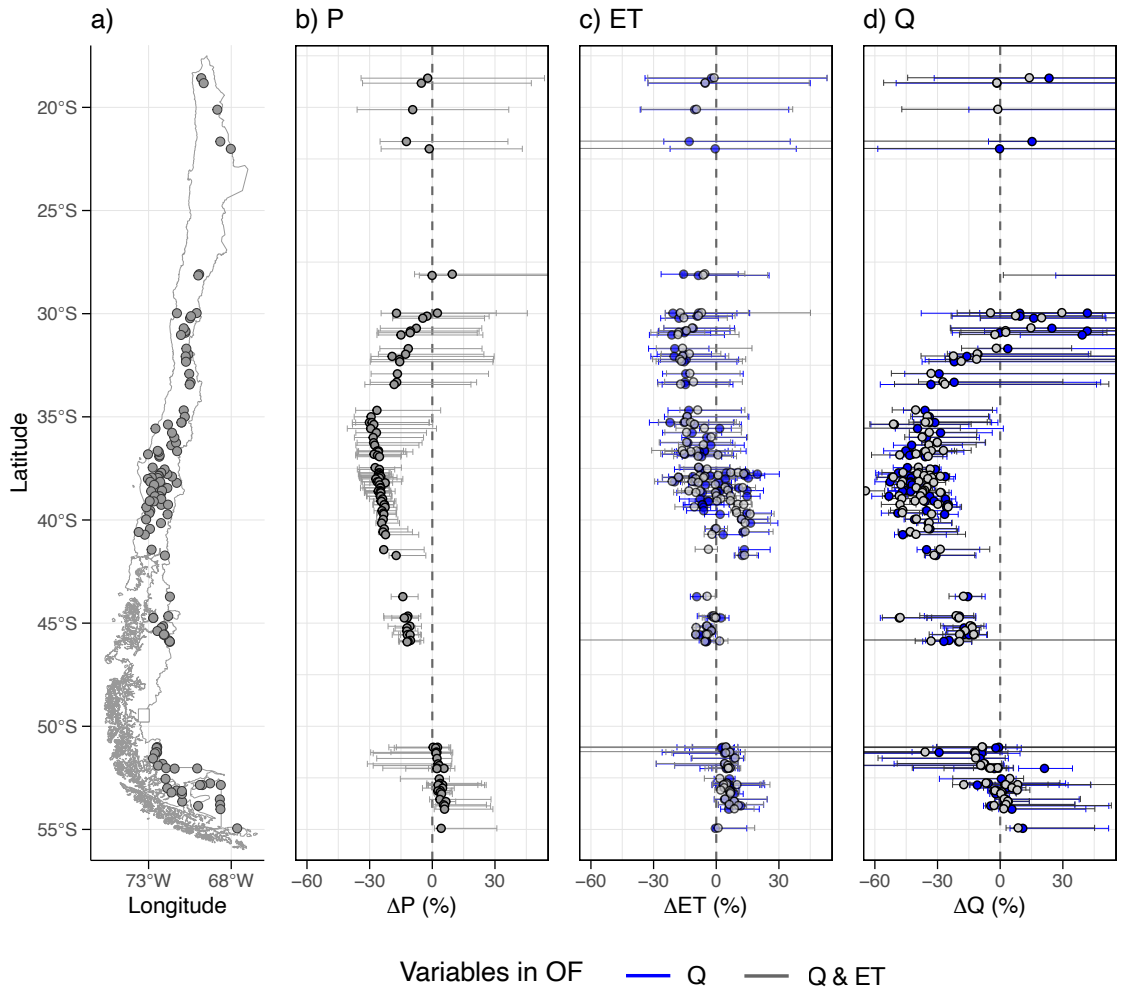

Figure 5.5: (a) Location of case study basins and catchment-scale projected annual changes for (b) P, (c) Q, and (d) ET (2070-2100 vs. 1985-2015). The dots represent the median among GCMs, while whiskers indicate the minimum and maximum among GCMs.

end of the 21th century varies between GCMs.

Figure 5.6 compares catchment-scale projected changes in annual/seasonal Q and ET from OF_Q against the results obtained with $OF_{Q\&ET}$ for the "best" two GCMs. Figure 5.6 shows that projected changes in annual ET (first row) and Q (second row) align, mostly, with the 1:1 line (dashed line), although there are differences between OFs, particularly for ΔQ. Interestingly, selecting different OFs in some basins (dots) may yield different projected change signals (see the dots in the top-left or bottom-right quadrants of each panel). At the seasonal scale, projected changes also align with the 1:1 line, although higher dispersion (lower R^2 values) can be observed, particularly for winter Q and summer ET.

Figure 5.7 shows projected changes in averaged monthly ET and Q for the same basins and GCMs displayed in Figure 5.4, respectively. For simplicity, the basins are referred to by the (i)-(v) identifiers, similar to Figure 5.4. Overall, the basins where differences between observed simulated historical ET seasonalities were obtained (basins ii and iv) yield the largest differences in projected monthly Q and ET changes, whereas the basins where consistent Q and ET observed-historical seasonalities were obtained yield more

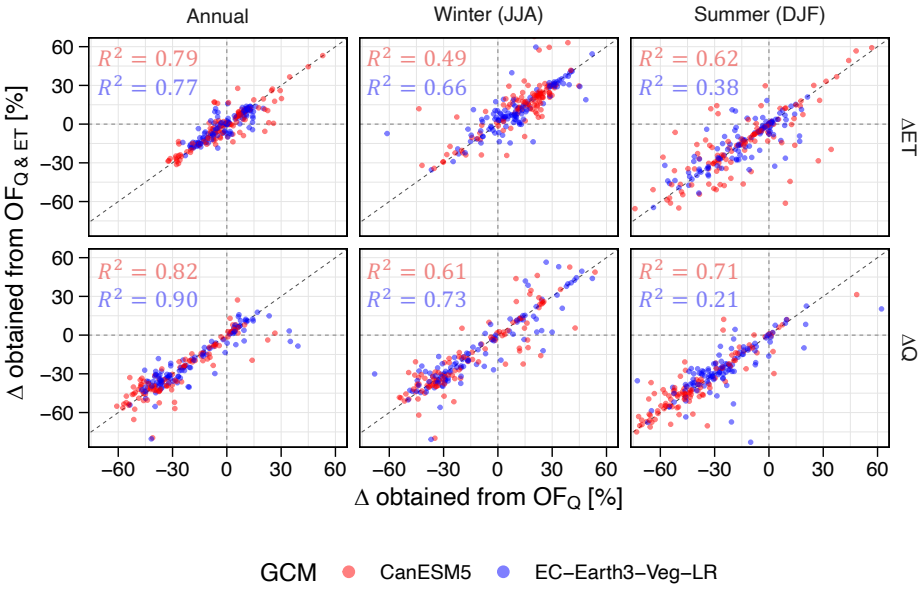

Figure 5.6: Projected percent changes in catchment scale annual and seasonal (Winter and Summer) Q and ET. The results are obtained for the future period 2070-2100 with respect to the reference period 1985-2015. The colors represent the best two GCMs based on the PPI index. (notice that the horizontal and vertical dashed lines represent a 0% change). R^2 represents the coefficient of determination.

robust (i.e., little differences from the choice of OFs) projected changes for particular GCMs. In basin (i), where differences in the observed-historical seasonalities are minimal, no relevant differences in projected changes in Q seasonality are obtained between OF and GCMs, except the dry period (MJJASON) where streamflow values are very low (average monthly values <0.1 m^3/s). In basin (ii), observed-historical differences between OFs can be seen in December for Q, and for the annual cycle of ET (Figure 5.4), producing different projections in Q and ET. Projected changes in Q based on $OF_{Q\&ET}$ yield more drastic reductions for both GCMs in the fall, winter, and summer seasons compared to OF_Q. Such changes may even be different in the projected change signal (see the JJA months for both GCMs). Similarly, projected monthly ET changes are drastic (i.e., larger magnitude) for OF_Q compared to $OF_{Q\&ET}$ (behavior opposed to ΔQ) for both GCMs, with differences in signal in the Fall and Winter months (MJJA months) for the GCM CanESM5. In basin (iii), which has no differences in seasonality in the observed-historical period (Figure 5.4), the choice of OF only introduces slight differences in monthly projected changes, and such differences occur mainly during the Summer-Fall seasons (months JFM). In basin (iv), where the observed-historical Q seasonality is well simulated but with very different annual cycles of ET from the OFs, the projected changes in Q and ET are drastically different in magnitude for both GCMs. For example, when the parameter is selected from OF_Q, more uniform Q changes are obtained along the year for both GCMs, whereas when using the parameter selected from $OF_{Q\&ET}$, less drastic changes are projected for winter compared to the rest of the seasons. In basin (iv), different monthly projected changes in ET are obtained with different OFs. Basin (v) has low-to-none differences between OFs for Q and ET monthly projections.

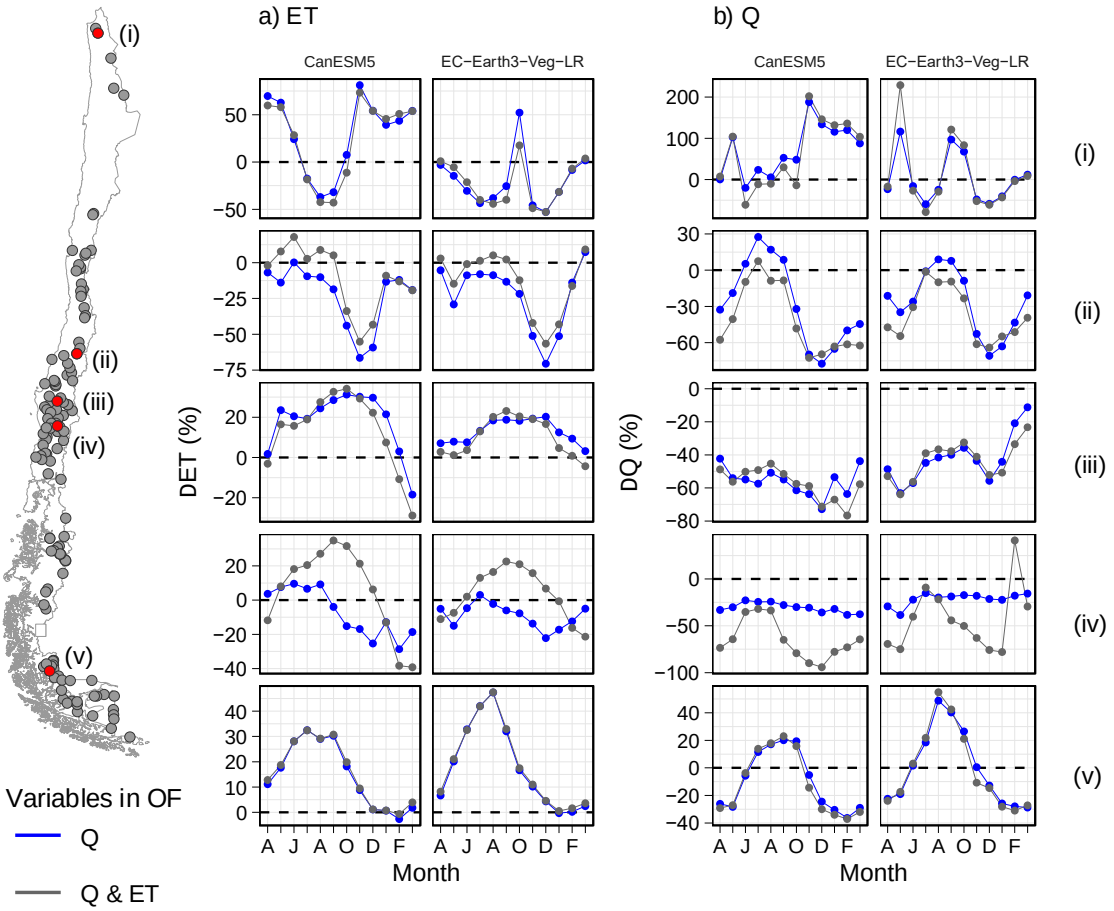

Figure 5.7: Projected monthly changes in Q (panel a) and ET (panel b) for the same selected basins as in Figure 5.4. Hydrologic changes are computed for the period 2070-2100 with respect to 1985-2015. The columns show results for the best two GCMs based on the PPI index. The highlighted basins in red on the map are the same as in Figure 5.4.

An ANOVA analysis is conducted to identify what decision (choice of OF or GCM) is more relevant in each basin to explain the spread in projected changes of Q and ET (Figure 5.8). Figure 5.8b shows that, in most catchments, GCM is the most relevant decision explaining the spread of ΔQ at the annual scale. However, the number of basins where the choice of OF is more relevant increases when moving to the seasonal scale. Similar results are obtained for ΔET (Figure 5.8a), although the number of basins being more sensitive to the choice of OF is smaller compared to the results for ΔQ. Interestingly, most of the basins where the choice of OF is more relevant to explain projected Q changes are located in central-southern Chile (36°S-42°S), a humid area, with low AI values (Figure 5.1) and where the best calibration metrics are obtained (Figure 5.2).

The Cautín at Cajón River basin (ID 9129002 (see location in Figure 5.9) is selected to illustrate how the choice of OF may affect the simulated mean annual runoff ratio (Q/P; Figure 5.9a) and the contribution of baseflow to total runoff (B/Q; Figure 5.9b). In this basin, calibration metrics are KGE(Q)=0.88 and KGE(1/Q)=0.91, with negligible differences between OFs in Q and ET seasonalities (not shown). It can be noted that Q/P values are not affected by the choice of OF in the future and historical periods (Figure 5.9a). However, the contribution of baseflow to total runoff changes between OF_{Q} (B/Q$\approx$0.41) and $OF_{Q\&ET}$ (B/Q$\approx$0.7) for the historical period. Interestingly, for

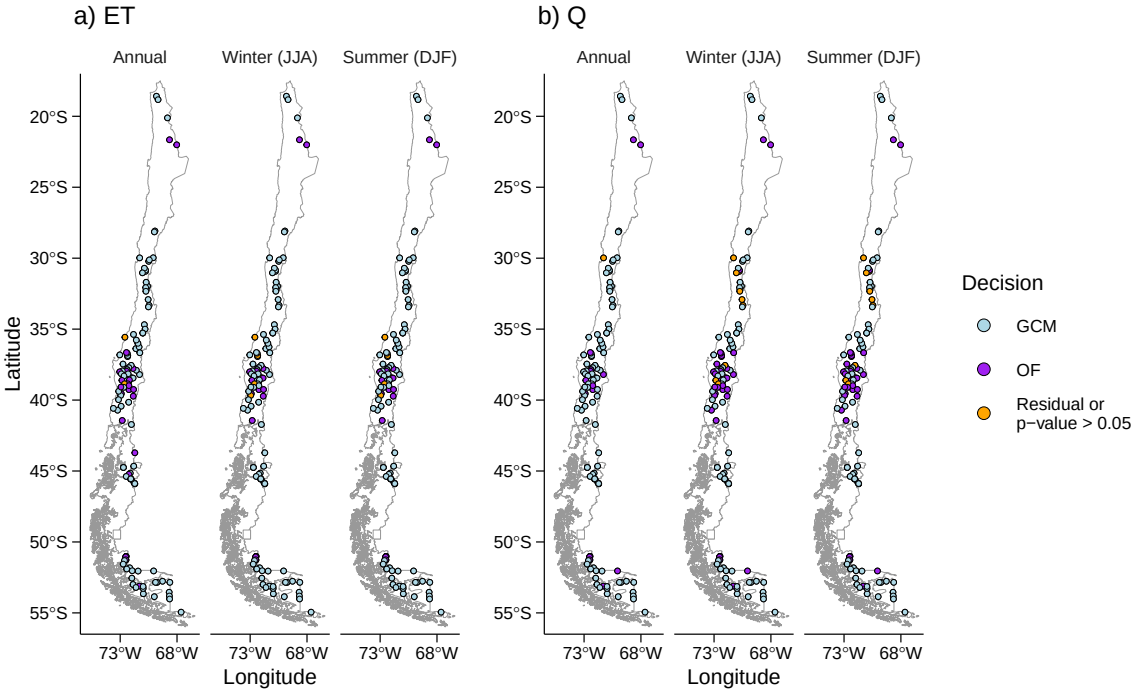

Figure 5.8: Most relevant decision (based on ANOVA analysis) to explain the projected changes in Q (panel a) and ET (panel b) at the annual and seasonal scales. Projected changes were computed comparing the periods 2070-2100 and 1985-2015. The most relevant decision is displayed only for p-values lower than 0.05.

the future period (2070-2100), B/Q increases for OF_Q, but remains nearly the same for $OF_{Q\&ET}$, suggesting that the choice of OF may affect the historical values and the projected changes of B/Q. For the Cautín River basin, results are similar regardless of the GCMs (different lines in Figures 5.9a and 5.9b). Figures 5.9c and 5.9d display the most relevant decision (using ANOVA analysis) explaining the spread of future average B/Q and Δ(B/Q), respectively. The choice of OF arises as the most relevant decision for projected B/Q in most of the basins (n=91), whereas the choice of OF and GCM arises as the most relevant decision For Δ(B/Q) in 55 and 50 basins, respectively.

Figure 5.10 displays the number of basins where the spatial distribution of ΔET (Figure 5.10a) and ΔQ (Figure 5.10b) obtained with OF_Q and $OF_{Q\&ET}$ within each basin is significantly different (p-value$\leq$0.05). Results are displayed for each GCM (y-axis) at the annual and seasonal time scales (x-axis). For ΔQ, the number of basins where the choice of OF yields different spatial patterns of ΔQ is $\sim$70 basins. The number of basins where the spatial distribution of ΔET within each basin is different between OF_Q and $OF_{Q\&ET}$ decreases compared to Figure 5.10b, particularly for the Summer season, although it affects more than half of the number of basins.

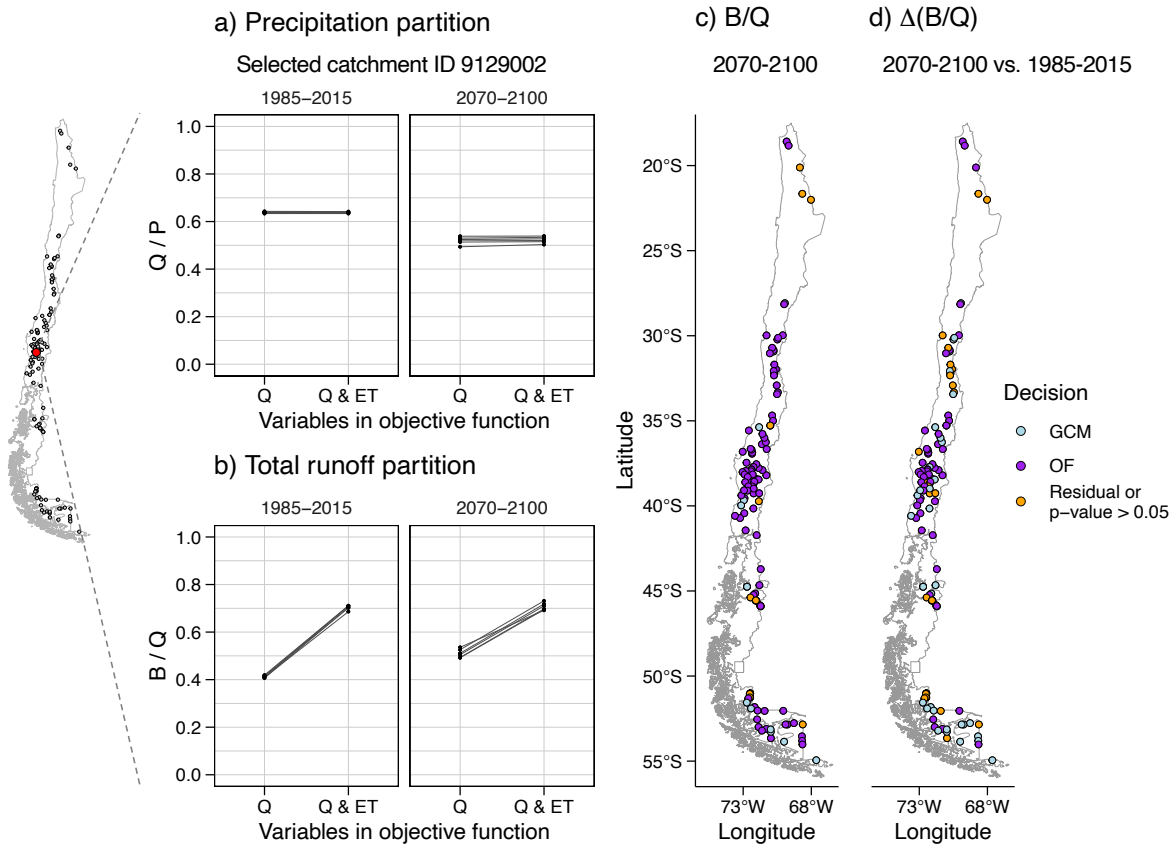

Figure 5.9: (a)Precipitation and (b) runoff partitioning in the historical (1985-2015) and future (2070-2100) periods for the basin Cautín River at Cajón (ID 9129002). Each line represents a GCM. Panels (c) and (d) show the most relevant decisions explaining the spread of the baseflow contribution to total runoff and its projected change, respectively.

5.6 Discussion

5.6.1 Flux equifinality in model calibration

The results presented here show that despite the two objective functions yielding similar streamflow performance metrics in many catchments, the quality of the simulation of other hydrological variables differs considerably (Cinkus et al., 2023). For example, a similar Q performance for the calibration period ($KGE_Q(Q)$=0.7, $KGE_Q(1/Q)$=0.87, $KGE_{Q\&ET}(Q)$=0.72, $KGE_Q(1/Q)$=0.85) is obtained at the Colorado River at Palos —basin (ii) in Figures 5.4 and 5.7 —, but the average seasonality differs (see Figure 5.4), being both incorrect according to the reference dataset. In Allipén in Los Laureles —basin (iv) in Figures 5.4 and 5.7 —, despite the two OFs yield a similar streamflow performance during the calibration period ($KGE_Q(Q)$=0.8, $KGE_Q(1/Q)$=0.88, $KGE_{Q\&ET}(Q)$=0.73, $KGE_Q(1/Q)$=0.9), opposite annual cycles are obtained for ET (maximum ET values in winter and spring vs. maximum ET values in summer for OF_Q and $OF_{Q\&ET}$, respectively). In the rest of the basins shown in Figures 5.4 —basins (i), (iii), and (v) —, similar Q and ET seasonalities are achieved.

It could be hypothesized that this is the result of model internal flux equifinality

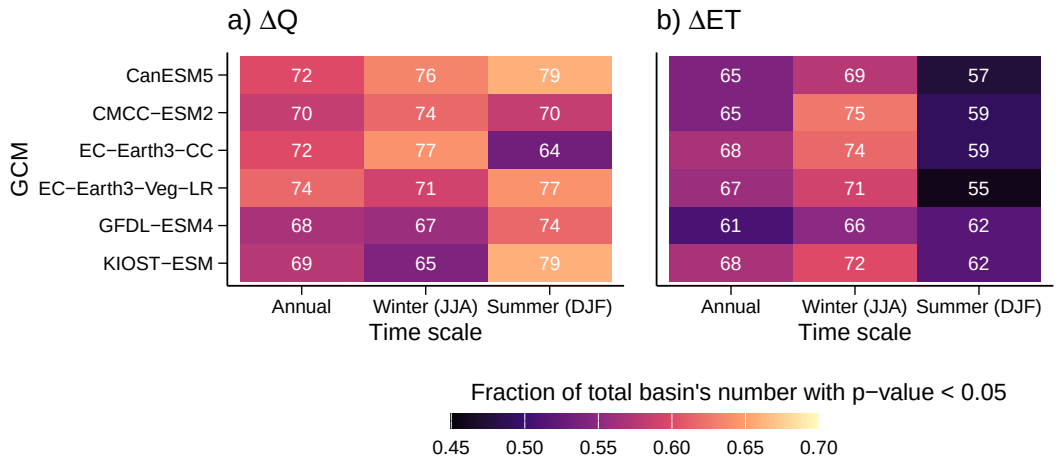

Figure 5.10: Number of basins where the spatial distribution of projected changes in (a) Q and (b) ET are significantly different (based on Kolmogorov-Smirnov tests) between OF_{Q} and $OF_{Q\ \&\ ET}$.

(K. Beven & Binley, 1992; Khatami et al., 2019). In VIC, bare soil evaporation and vegetation transpiration depend on the water stored in the different soil layers, and R and B fluxes contribute to Q. Then, different "R/B" ratios (in the case of VIC) can yield similar Q time series (as illustrated for the Cautín at Cajón river basin in Figure 5.9a and b), due to flux equifinality, as shown by Khatami et al. (2019). More generally, hydrologic models hide an interplay between infiltration, percolation, and water release from different soil layers that will affect soil storage and, therefore water supply for ET. The results of this study illustrate an extreme case of flux equifinality that should be properly diagnosed in order to identify potential errors in forcing data and/or model structure.

5.6.2 On the need to verify annual cycles

Despite the myriad of existing calibration and evaluation approaches that aim to increase model fidelity and/or maintain model performance under changing conditions, no specific requirements (e.g., a minimum performance metric threshold or the evaluation of particular hydrological variables beyond streamflow) have been - to the best of the author's knowledge - proposed to accept a model as "good". However, the extreme cases of flux equifinality that affect ET simulations in some basins show that replicating annual cycles of Q does not warrant realistic catchment-scale water balance portrayals. In this regard, the results of this chapter show that the use of an ET reference dataset for model evaluation is critical to identify model parameter sets that (could) provide inconsistent behaviors. Furthermore, complementing the verification of model simulations using visual inspection or ad-hoc metrics may be necessary since relying on overall performance metric (and their optimization) could be misleading and not necessarily improve model fidelity (Clark et al., 2021; Cinkus et al., 2023; Jiménez-Navarro et al., 2024).

5.6.3 Effects on hydrological projections

In this study, two parameter sets with similar streamflow responses are selected, each obtained from maximizing different objective functions. These parameter sets project different Q and ET annual cycles for the same GCM (Figure 5.7), affecting any subsequent analysis. Despite the use of six GCMs that provide different climate projections, the mean monthly Q and ET change patterns (Figure 5.7) are mostly controlled by the inclusion of ET in the OF for all GCMs. Further, including ET in the OF could also affect the relative importance of different fluxes contributing to streamflow (Figures 5.9a and b), though this could also occur for streamflow-based calibration due to flux equifinality (Khatami et al., 2019; Jiménez-Navarro et al., 2024). Finally, the choice of OF arises as the most relevant decision when analyzing projected changes in B contributions to Q for 50 basins, mostly located in the humid area between 35°S and 43°S. In this regard, most of the basins have maximum Q-values in winter (the rainiest season in this part of the domain). However, in this area, ET maximum values occur in summer (according to the MOD16 ET product; section C). Hence, the partition of precipitation should suffice maximum Q values in winter and maximum ET values in summer. Thus, when only Q is included in the objective function to derive the model parameters, the Q-ET decoupled average seasonalities impose time-delay requirements to the model not included in the calibration. Consequently, verifying the ET average cycle in basins with decoupled Q-ET seasonalities could be relevant for hydrological modeling and climate change assessment (Figure C.2).
It is worth noting that parameter equifinality arises regardless of the variable included in the calibration process. However, this study suggests that verifying the catchment-scale water balance seasonality may be as relevant as the choice of GCM, which is usually one of the main sources of dispersion in hydrological projections (e.g., Hattermann et al., 2018; Chegwidden et al., 2019).

5.6.4 The particular case of continental Chile

Many previous studies have produced hydrological projections for different basins in continental Chile, considering different greenhouse gas emission scenarios, GCMs, hydrological model structures, and parameter estimation methods. Therefore, any comparison with the results presented here should be made cautiously. Vicuña et al. (2011) calibrated the Water Evaluation and Planning System (WEAP; Yates et al., 2005) model to estimate hydrological changes in the Limarí River basin (31°S), obtaining a decrease in mean annual streamflow ranging between -16.6% and -23.5% for scenario B2 and -21% and -41.5% for scenario A2. In this chapter are obtained, for the same latitude, positive and negative projected changes in total runoff, following the mixed projected changes in annual precipitation, although the lowest values of the results presented in this study project a 30% decrease in total annual runoff for the same area. However, for winter and summer streamflow, the results of this work align with Vicuña et al. (2011), projecting a shift in hydrological regime from snowmelt-driven to mixed regime, with an increase in rainfall contributions to runoff. (Demaria et al., 2013) calibrated the VIC model to estimate projected changes in extreme runoff under CMIP3 and CMIP5 models for the Mataquito River basin (35.5°S). They projected precipitation changes of -7% and -20% (depending on the CMIP considered) by the end of the century, whereas the results of this study project (under scenario SSP5-8.5) a decrease in mean annual precipitation between -40% and -20%, although one GCM (KIOST-ESM) projects an increase up to 20%. Bozkurt

et al. (2018) used the VIC model to produce hydrological projections for basins between 34°S and 37°S under scenarios RCP2.6 and RCP8.5. Under the latter scenario —the most similar to SSP5-8.5, used in this study —the results shown in this chapter align well with Bozkurt et al. (2018), projecting a median decrease of -40% in annual total runoff for the area, although the projected changes in monthly Q and (particularly) ET differ. Aguayo et al. (2019) calibrated the WEAP model (using streamflow-only metrics) in basins located between 41°S and 42°S (Puelo River Basin). They projected an increase in winter total runoff (JJA) and a decrease in summer total runoff under the RCP8.5 scenario, whereas the results of this chapter suggest that streamflow will decrease in all seasons and that winter ET is projected to increase in basins around 42°S.

5.6.5 Limitations

In this study, only one hydrological model structure and two parameter sets are used, one parameter for each OF. Further, only the SSP5-8.5 scenario is considered for hydroclimatic projectionsthough more scenarios could be included to complement the analyses presented here. Because previous studies have shown that GCMs are among the main contributors to the spread of hydroclimatic projections (Addor et al., 2014; Hattermann et al., 2018; Chegwidden et al., 2019), the analyses presented in this chapter evaluate the relative importance of the choice of OF compared to the choice of GCM. though more decisions could be included in the ANOVA analysis, such as alternative greenhouse emission scenarios, hydrological models, OFs, downscaling, and bias correction methods.

5.7 Conclusions

This study evaluates the repercussions on simulated monthly Q and ET averages when including ET in the calibration, since both control the overall water balance seasonality. This chapter contributes to the existing literature by demonstrating the added value of incorporating ET in the evaluation of hydrologic models in order to ensure consistent catchment-scale seasonal water balances. To this end, the VIC hydrologic model is calibrated with two objective functions (OF_Q and $OF_{Q\&ET}$) in 120 basins located in continental Chile. The parameter sets obtained were used to evaluate the simulated seasonal water balances and hydrological projections under scenario SSP5-8.5. The main findings are as follows:

1. Parameter sets that produce similar streamflow responses can differ in their simulated annual cycles of ET. Therefore, accurately simulating runoff seasonality does not guarantee a reliable seasonal water balance at the catchment scale. In this context, evaluating ET is crucial for reducing equifinality in the model's internal fluxes and state variables, particularly in basins with decoupled Q and ET average seasonalities. Failing ET simulations (under similar Q model performance) could lead to different projections of both Q and ET under climate change assessment. Thus, assessing ET helps detect potential misrepresentations in the seasonal water balance.
2. Parameter sets with Q simulations but with different ET simulations will lead to different projected changes in annual, seasonal, and monthly Q and ET. This result

is consistent among the GCMs analyzed in this study.

3. Including ET in the OF may become the most relevant decision (compared to the choice of GCM) when analyzing projected changes in Q and ET at the seasonal and monthly time scales. Moreover, nearly half of the basins analyzed show differences in the spatial distribution of projected Q and ET between parameters obtained by OF_{Q} and $OF_{Q\&ET}$.

4. Although the basins that are sensitive to the inclusion of ET in the OF formulation are distributed along the study domain, they are highly concentrated in humid areas, with high annual precipitation amounts and low aridity indices, and decoupled streamflow and evapotranspiration seasonalities.

Chapter 6

Concluding remarks

This thesis estimates how key methodological decisions affect historical simulations and future hydroclimate projections under climate change conditions. First, the role of the temporal stratification used to apply bias correction methods is explored. Secondly, a characterization of climate change impacts on precipitation, temperature, snowfall, contributing areas on rainy days, and the projected spatial distribution of Köppen-Geiger climate classes is provided for continental Chile. Thirdly, this work examined how the choice of streamflow-based calibration metric and parameter regularized affect the annual cycles of simulated variables, including evapotranspiration, soil moisture, land surface temperature, and fractional snow covered area. Finally, it is analyzed to what extent the exclusion of ET from the calibration process may affect hydrological projections under climate change scenarios. .

If bias correction is required for climate change impact assessments, two decisions become relevant: (i) the choice of bias correction method, and (ii) the temporal stratification used to apply the bias correction method. The first decision has been identified as critical because it affects the magnitude and signal of projected changes. Interestingly, the results and conclusions presented in this thesis stress the need for caution when selecting the temporal stratification, not only because it affects the magnitude and signal of projected changes, but also because it is critical to replicate historically observed and projected seasonalities of precipitation and temperature. The assessment of climate change impacts on water resources usually requires forcing hydrological models with downscaled and bias-corrected GCM outputs. Hence, the choice of temporal stratification could not only affect the capability of process-based models to replicate the hydrological regime of observed records, but also how the hydrological regime could change in the future under changing forcing conditions. Moreover, such artificial alterations could produce forcing data that differ from the raw GCM projected changes and seasonalities.

The assessment of climate change for continental Chile considered an ensemble of 30 models, projections under the SSP5-8.5 scenario and a multivariate bias correction method applied with a monthly temporal stratification. Such configuration was preferred to replicate the historical seasonality of precipitation and temperature. Furthermore, the choice of GCM is among the main contributors to the dispersion of hydrological change projections; hence, a particular assessment of an ensemble of models was conducted to identify regions of the study domain where high inter-model agreement arises among model projections. The ensemble of GCMs allowed the examination of the inter-model agreement on the projections of precipitation, temperature, and snowfall (magnitude and

signal of projected changes) along continental Chile. The results indicate increments in projected median changes of 0° C isotherm during rainy days, increasing flood risk. Despite the dispersion in precipitation and temperature changes obtained with different GCMs, there is a high inter-model agreement (> 90%) in the projected spatial distribution of Köppen-Geiger climate classes.

The assessment of hydrological change due to changing climatic conditions is typically conducted through hydrologic modeling, which, in most cases, requires the calibration of model parameters. This thesis has shown that a streamflow-based calibration strategy can miss critical aspects of other hydrological variables than streamflow, affecting model fidelity and, more critically, the catchment scale water balance. This work also illustrates how regularizing different parameters could affect and/or improve the simulated spatial patterns of different hydrological variables, although such improvements could be insufficient to adequately represent seasonal water balances. Hydrological models should be able to adequately simulate fluxes and state variables simultaneously. However, such an aim is challenged by the scarcity of information, forcing uncertainty, model structure, and/or parameter errors. In this regard, it could be argued that replicating the basin scale water balance at the annual and seasonal scales is a minimum requirement for hydrologic models, particularly when good streamflow simulations do not necessarily translate into good evapotranspiration simulations due to flux equifinality.

Hydrological projections are provided for 120 basins in near-natural regimes across continental Chile. To this end, the VIC model was calibrated with two objective functions: (i) using a traditional streamflow-based calibration metric and (ii) using a metric that incorporates streamflow and evapotranspiration. The projected change in annual streamflow follows, in general, the latitudinal pattern obtained for annual precipitation changes. Further, the most drastic reductions (under scenario SSP5-8.5) are expected in central and southern Chile (33°S-45°S), although a higher spread in the magnitude of projected changes is obtained for hydrological projections compared to changes in annual precipitation. Interestingly, the annual evapotranspiration is expected to increase in regions south of 37°S. The choice of the objective function and, more critically, the ability of the model to replicate the seasonal water balance play a key role in hydrological projections, even for equifinal parameter sets (from a streamflow point of view) at the annual, seasonal, and monthly temporal scales. Moreover, it can also affect the spatial distribution of projected runoff and evapotranspiration changes.

This thesis aligns with previous work emphasizing the role of process-based hydrologic model evaluation (e.g., M. C. Demirel et al., 2018; Dembélé, Ceperley, et al., 2020; Clark et al., 2016, 2021; Cinkus et al., 2023; Jiménez-Navarro et al., 2024; Pool et al., 2024; M. C. Demirel et al., 2024), identifying severe pitfalls in hydrological simulations and the application of bias correction methods, affecting hydrological change assessments. Finally, this work expands previous knowledge focused on improving model fidelity and the impacts of methodological decisions on the simulated hydrology of catchments in continental Chile.

In summary, this thesis emphasizes two key aspects: (i) the importance of temporal stratification in applying bias correction methods, which aids in replicating reference precipitation seasonality but may distort projected seasonal patterns, and (ii) the necessity of verifying the seasonality of the simulated water balance. Both factors significantly influence the historical and future projected changes in both signal and magnitude, ultimately impacting the outcomes of climate change assessments.

BIBLIOGRAPHY

Aceituno, P., Boisier, J. P., Garreaud, R., Rondanelli, R., & Rutllant, J. A. (2021). Climate and Weather in Chile. In *Water resources of chile* (pp. 7–29). doi: 10.1007/978-3-030-56901-3-2

Addor, N., Rössler, O., Köplin, N., Huss, M., Weingartner, R., & Seibert, J. (2014). Robust changes and sources of uncertainty in the projected hydrological regimes of Swiss catchments. *Water Resources Research*. doi: 10.1002/2014WR015549

Aguayo, R., León-Muñoz, J., Garreaud, R., & Montecinos, A. (2021, 3). Hydrological droughts in the southern Andes (40–45°S) from an ensemble experiment using CMIP5 and CMIP6 models. *Scientific Reports*, *11*(1), 5530. doi: 10.1038/s41598-021-84807-4

Aguayo, R., León-Muñoz, J., Vargas-Baecheler, J., Montecinos, A., Garreaud, R., Urbina, M., ... Iriarte, J. L. (2019, 8). The glass half-empty: climate change drives lower freshwater input in the coastal system of the Chilean Northern Patagonia. *Climatic Change*, *155*(3), 417–435. doi: 10.1007/s10584-019-02495-6

Akbar, R., Short Gianotti, D. J., Salvucci, G. D., & Entekhabi, D. (2020, 12). Partitioning of Historical Precipitation Into Evaporation and Runoff Based on Hydrologic Dynamics Identified With Recent SMAP Satellite Measurements. *Water Resources Research*, *56*(12). doi: 10.1029/2020WR027307

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*. doi: 10.1029/2018WR023458

Alfieri, L., Avanzi, F., Delogu, F., Gabellani, S., Bruno, G., Campo, L., ... Brocca, L. (2022, 7). High-resolution satellite products improve hydrological modeling in northern Italy. *Hydrology and Earth System Sciences*, *26*(14), 3921–3939. doi: 10.5194/hess-26-3921-2022

Alvarez-Garreton, C., Mendoza, P. A., Pablo Boisier, J., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., ... Ayala, A. (2018). The CAMELS-CL dataset: Catchment attributes and meteorology for large sample studies-Chile dataset. *Hydrology and Earth System Sciences*. doi: 10.5194/hess-22-5817-2018

Andreadis, K. M., & Lettenmaier, D. P. (2006, 6). Assimilating remotely sensed snow observations into a macroscale hydrology model. *Advances in Water Resources*, *29*(6), 872–886. doi: 10.1016/j.advwatres.2005.08.004

Andreadis, K. M., Storck, P., & Lettenmaier, D. P. (2009, 5). Modeling snow accumulation and ablation processes in forested environments. *Water Resources Research*, *45*(5). doi: 10.1029/2008WR007042

Araki, R., Branger, F., Wiekenkamp, I., & McMillan, H. (2022, 4). A signature-based approach to quantify soil moisture dynamics under contrasting land-uses. *Hydrological Processes*, *36*(4). doi: 10.1002/hyp.14553

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*. doi: 10.1007/s00382-020-05231-4

Aryal, Y., & Zhu, J. (2017, 2). On bias correction in drought frequency analysis based on climate models. *Climatic Change*, *140*(3-4), 361–374. doi: 10.1007/s10584-016-1862-3

Avanzi, F., Maurer, T., Glaser, S. D., Bales, R. C., & Conklin, M. H. (2020, 3). Information content of spatially distributed ground-based measurements for hydrologic-

parameter calibration in mixed rain-snow mountain headwaters. *Journal of Hydrology*, *582*, 124478. doi: 10.1016/j.jhydrol.2019.124478

Ayala, A., Farías-Barahona, D., Huss, M., Pellicciotti, F., McPhee, J., & Farinotti, D. (2020, 6). Glacier runoff variations since 1955 in the Maipo River basin, in the semiarid Andes of central Chile. *The Cryosphere*, *14*(6), 2005–2027. doi: 10.5194/tc-14-2005-2020

Bajracharya, A. R., Ahmed, M. I., Stadnyk, T., & Asadzadeh, M. (2023, 6). Process based calibration of a continental-scale hydrological model using soil moisture and streamflow data. *Journal of Hydrology: Regional Studies*, *47*, 101391. doi: 10.1016/j.ejrh.2023.101391

Bambach, N., Meza, F., Gilabert, H., & Miranda, M. (2013, 12). Impacts of climate change on the distribution of species and communities in the Chilean Mediterranean ecosystem. *Regional Environmental Change*, *13*(6), 1245–1257. doi: 10.1007/s10113-013-0425-7

Bambach, N., Rhoades, A., Hatchett, B., Jones, A., Ullrich, P., & Zarzycki, C. (2022, 3). Projecting climate change in South America using variable-resolution Community Earth System Model: An application to Chile. *International Journal of Climatology*, *42*(4), 2514–2542. doi: 10.1002/joc.7379

Barria, I., Carrasco, J., Casassa, G., & Barria, P. (2019, 10). Simulation of Long-Term Changes of the Equilibrium Line Altitude in the Central Chilean Andes Mountains Derived From Atmospheric Variables During the 1958–2018 Period. *Frontiers in Environmental Science*, *7*. doi: 10.3389/fenvs.2019.00161

Barría, P., Sandoval, I. B., Guzman, C., Chadwick, C., Alvarez-Garreton, C., Díaz-Vasconcellos, R., ... Fuster, R. (2021, 5). Water allocation under climate change. *Elementa: Science of the Anthropocene*, *9*(1). doi: 10.1525/elementa.2020.00131

Beck, H. E., McVicar, T. R., Vergopolan, N., Berg, A., Lutsko, N. J., Dufour, A., ... Miralles, D. G. (2023, 10). High-resolution (1 km) Köppen-Geiger maps for 1901–2099 based on constrained CMIP6 projections. *Scientific Data*, *10*(1), 724. doi: 10.1038/s41597-023-02549-6

Beck, H. E., Pan, M., Lin, P., Seibert, J., Dijk, A. I. J. M. v., & Wood, E. F. (2020, 9). Global Fully Distributed Parameter Regionalization Based on Observed Streamflow From 4,229 Headwater Catchments. *Journal of Geophysical Research: Atmospheres*, *125*(17), e2019JD031485. doi: 10.1029/2019JD031485

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018, 10). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*(1), 180214. doi: 10.1038/sdata.2018.214

Bennett, K. E., Cherry, J. E., Balk, B., & Lindsey, S. (2019, 5). Using MODIS estimates of fractional snow cover area to improve streamflow forecasts in interior Alaska. *Hydrology and Earth System Sciences*, *23*(5), 2439–2459. doi: 10.5194/hess-23-2439-2019

Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*. doi: 10.1016/0309-1708(93)90028-E

Beven, K. (2001). How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences*. doi: 10.5194/hess-5-1-2001

Beven, K. (2006). A manifesto for the equifinality thesis. In *Journal of hydrology*. doi: 10.1016/j.jhydrol.2005.07.007

Beven, K. (2012). *Rainfall-Runoff Modelling: The Primer: Second Edition*. doi: 10.1002/9781119951001

Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*. doi: 10.1002/hyp.3360060305

Beven, K. J. (1990). A Discussion of Distributed Hydrological Modelling. , 255–278. Retrieved from https://link.springer.com/chapter/10.1007/978-94-009-0257-2_13 doi: 10.1007/978-94-009-0257-2{_}13

Beven, K. J., & Chappell, N. A. (2021). *Perceptual perplexity and parameter parsimony.* doi: 10.1002/wat2.1530

Blöschl, G., & Sivapalan, M. (1995, 4). Scale issues in hydrological modelling: A review. *Hydrological Processes*, *9*(3-4), 251–290. doi: 10.1002/HYP.3360090305

Boisier, J. P. (2023). CR2MET: A high-resolution precipitation and temperature dataset for the period 1960-2021 in continental Chile. (v2.5) [Data set]. *Zenodo*. doi: 10.5281/zenodo.7529681

Boisier, J. P., Alvarez-Garreton, C., Cordero, R. R., Damiani, A., Gallardo, L., Garreaud, R. D., ... Rondanelli, R. (2018, 1). Anthropogenic drying in central-southern Chile evidenced by long-term observations and climate model simulations. *Elementa: Science of the Anthropocene*, *6*. doi: 10.1525/elementa.328

Bozkurt, D., Rojas, M., Boisier, J. P., Rondanelli, R., Garreaud, R., & Gallardo, L. (2019). Dynamical downscaling over the complex terrain of southwest South America: present climate conditions and added value analysis. *Climate Dynamics*. doi: 10.1007/s00382-019-04959-y

Bozkurt, D., Rojas, M., Boisier, J. P., & Valdivieso, J. (2017). Climate change impacts on hydroclimatic regimes and extremes over Andean basins in central Chile. *Hydrology and Earth System Sciences Discussions*. doi: 10.5194/hess-2016-690

Bozkurt, D., Rojas, M., Boisier, J. P., & Valdivieso, J. (2018, 9). Projected hydroclimate changes over Andean basins in central Chile from downscaled CMIP5 models under the low and high emission scenarios. *Climatic Change 2018 150:3*, *150*(3), 131–147. doi: 10.1007/S10584-018-2246-7

Brooks, R. H., & Corey, A. T. (1964, 3). Hydraulic properties of porous media. *Hydrology Papers, Colorado State University*, *3*.

Cannon, A. J. (2011, 9). Quantile regression neural networks: Implementation in R and application to precipitation downscaling. *Computers & Geosciences*, *37*(9), 1277–1284. doi: 10.1016/j.cageo.2010.07.005

Cannon, A. J. (2016, 10). Multivariate Bias Correction of Climate Model Output: Matching Marginal Distributions and Intervariable Dependence Structure. *Journal of Climate*, *29*(19), 7045–7064. doi: 10.1175/JCLI-D-15-0679.1

Cannon, A. J. (2018). Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dynamics*. doi: 10.1007/s00382-017-3580-6

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*. doi: 10.1175/JCLI-D-14-00754.1

Carrasco, J. F., Casassa, G., & Quintana, J. (2005, 12). Changes of the 0°C isotherm and the equilibrium line altitude in central Chile during the last quarter of the 20th century / Changements de l'isotherme 0°C et de la ligne d'équilibre des neiges dans le Chili central durant le dernier quart du 20ème siècle. *Hydrological Sciences Journal*, *50*(6). doi: 10.1623/hysj.2005.50.6.933

Casper, M. C., Salm, Z., Gronz, O., Hutengs, C., Mohajerani, H., & Vohland, M. (2023, 11). Calibration of Land-Use-Dependent Evaporation Parameters in Distributed Hydrological Models Using MODIS Evaporation Time Series Data. *Hydrology*, *10*(12), 216. doi: 10.3390/hydrology10120216

Catalano, A. J., Loikith, P. C., & Neelin, J. D. (2020, 7). Evaluating CMIP6 model fidelity at simulating non-Gaussian temperature distribution tails. *Environmental*

Research Letters, *15*(7), 074026. doi: 10.1088/1748-9326/ab8cd0

Chaubey, P. K., & Mall, R. K. (2023, 9). Intensification of Extreme Rainfall in Indian River Basin: Using Bias Corrected CMIP6 Climate Data. *Earth's Future*, *11*(9). doi: 10.1029/2023EF003556

Cheeseman, P., John, R., & Nasa, S. (1996). Bayesian Classification (AutoClass): Theory and Results. *Advances in knowledge discovery and data mining*.

Cheeseman, P., Kelly, J., Self, M., Stutz, J., Taylor, W., & Freeman, D. (1988, 1). AutoClass: A Bayesian Classification System. *Machine Learning Proceedings 1988*, 54–64. doi: 10.1016/B978-0-934613-64-4.50011-6

Chegwidden, O. S., Nijssen, B., Rupp, D. E., Arnold, J. R., Clark, M. P., Hamman, J. J., ... Xiao, M. (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*. doi: 10.1029/2018EF001047

Chen, J., Arsenault, R., Brissette, F. P., & Zhang, S. (2021). Climate Change Impact Studies: Should We Bias Correct Climate Model Outputs or Post-Process Impact Model Outputs? *Water Resources Research*. doi: 10.1029/2020WR028638

Chen, J., Brissette, F. P., Chaumont, D., & Braun, M. (2013, 7). Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. *Water Resources Research*, *49*(7), 4187–4205. doi: 10.1002/wrcr.20331

Cheng, Y., Musselman, K. N., Swenson, S., Lawrence, D., Hamman, J., Dagon, K., ... Newman, A. J. (2023, 1). Moving Land Models Toward More Actionable Science: A Novel Application of the Community Terrestrial Systems Model Across Alaska and the Yukon River Basin. *Water Resources Research*, *59*(1). doi: 10.1029/2022WR032204

Cinkus, G., Mazzilli, N., Jourde, H., Wunsch, A., Liesch, T., Ravbar, N., ... Goldscheider, N. (2023, 7). When best is the enemy of good – critical evaluation of performance criteria in hydrological models. *Hydrology and Earth System Sciences*, *27*(13), 2397–2411. doi: 10.5194/hess-27-2397-2023

Clark, M. P., Bierkens, M. F., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., ... Peters-Lidard, C. D. (2017). The evolution of process-based hydrologic models: Historical challenges and the collective quest for physical realism. *Hydrology and Earth System Sciences*. doi: 10.5194/hess-21-3427-2017

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., ... Rasmussen, R. M. (2015). A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resources Research*. doi: 10.1002/2015WR017198

Clark, M. P., Vogel, R. M., Lamontagne, J. R., Mizukami, N., Knoben, W. J., Tang, G., ... Papalexiou, S. (2021, 8). The abuse of popular performance metrics in hydrologic modeling. *Water Resources Research*, e2020WR029001. doi: 10.1029/2020WR029001

Clark, M. P., Wilby, R. L., Gutmann, E. D., Vano, J. A., Gangopadhyay, S., Wood, A. W., ... Brekke, L. D. (2016, 6). Characterizing Uncertainty of the Hydrologic Impacts of Climate Change. *Current Climate Change Reports*, *2*(2), 55–64. doi: 10.1007/s40641-016-0034-x

CONAMA. (2006). *Estudio de la variabilidad climática en Chile para el siglo XXI* (Tech. Rep.). Corporación Nacional del Medio Ambiente. Realizado por Departamento de Geofísica, Universidad de Chile. Retrieved from `http://dgf.uchile.cl/PRECIS/`

Cordero, R. R., Asencio, V., Feron, S., Damiani, A., Llanillo, P. J., Sepulveda, E., ... Casassa, G. (2019, 11). Dry-Season Snow Cover Losses in the Andes (18°–40°S) driven by Changes in Large-Scale Climate Modes. *Scientific Reports*, *9*(1), 16945.

doi: 10.1038/s41598-019-53486-7

Cornwell, E., Molotch, N. P., & McPhee, J. (2016). Spatio-temporal variability of snow water equivalent in the extra-tropical Andes Cordillera from distributed energy balance modeling and remotely sensed snow cover. *Hydrology and Earth System Sciences*. doi: 10.5194/hess-20-411-2016

Coron, L., Thirel, G., Delaigue, O., Perrin, C., & Andréassian, V. (2017, 8). The suite of lumped GR hydrological models in an R package. *Environmental Modelling & Software*, *94*, 166–171. doi: 10.1016/j.envsoft.2017.05.002

Cortés-Salazar, N., Vásquez, N. A., Mizukami, N., Mendoza, P. A., & Vargas, X. (2023, 10). To what extent does river routing matter in hydrological modeling? *Hydrology and Earth System Sciences*, *27*(19), 3505–3524. doi: 10.5194/hess-27-3505-2023

Cortina, M., & Madeira, C. (2023, 6). Exposures to climate change's physical risks in Chile. *Latin American Journal of Central Banking*, *4*(2), 100090. doi: 10.1016/j.latcb.2023.100090

Covián, F., & Stowhas, L. (2015). Modelación probabilística de la línea de nieves durante eventos de tormenta entre los 28.5°S and 46.5°S. *Sociedad Chilena de Ingeniería Hidráulica - XXII Congreso Chileno de Ingeniería Hidráulica*.

Coxon, G., Freer, J., Lane, R., Dunne, T., Knoben, W. J. M., Howden, N. J. K., ... Woods, R. (2019, 6). DECIPHeR v1: Dynamic fluxEs and ConnectIvity for Predictions of HydRology. *Geoscientific Model Development*, *12*(6), 2285–2306. doi: 10.5194/gmd-12-2285-2019

CR2. (2018). *"Simulaciones climáticas regionales. Proyecto "Simulaciones climáticas regionales y marco de evaluación de la vulnerabilidad" mandatado por el Ministerio del Medio Ambiente (FONDAP 15110009)* (Tech. Rep.). Centro de Ciencia del Clima y la Resiliencia (CR)2. Retrieved from `https://www.cr2.cl/wp-content/uploads/2019/06/Simulaciones-climticas-regionales-2018.pdf`

Craig, J. R., Brown, G., Chlumsky, R., Jenkinson, R. W., Jost, G., Lee, K., ... Tolson, B. A. (2020, 7). Flexible watershed simulation with the Raven hydrological modelling framework. *Environmental Modelling & Software*, *129*, 104728. doi: 10.1016/j.envsoft.2020.104728

Crow, B. R., Tarasov, L., Schulz, M., & Prange, M. (2024, 2). Uncertainties originating from GCM downscaling and bias correction with application to the MIS-11c Greenland Ice Sheet. *Climate of the Past*, *20*(2), 281–296. doi: 10.5194/cp-20-281-2024

Cuartas, L. A., Tomasella, J., Nobre, A. D., Nobre, C. A., Hodnett, M. G., Waterloo, M. J., ... Ferreira, M. (2012, 9). Distributed hydrological modeling of a micro-scale rainforest watershed in Amazonia: Model evaluation and advances in calibration using the new HAND terrain model. *Journal of Hydrology*, *462-463*, 15–27. doi: 10.1016/j.jhydrol.2011.12.047

Cuntz, M., Mai, J., Samaniego, L., Clark, M., Wulfmeyer, V., Branch, O., ... Thober, S. (2016). The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model. *Journal of Geophysical Research*. doi: 10.1002/2016JD025097

de Lavenne, A., Andréassian, V., Thirel, G., Ramos, M. H., & Perrin, C. (2019). A Regularization Approach to Improve the Sequential Calibration of a Semidistributed Hydrological Model. *Water Resources Research*. doi: 10.1029/2018WR024266

Demaria, E. M., 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*. doi: 10.1016/j.jhydrol.2013.08.027

Dembélé, M., Ceperley, N., Zwart, S. J., Salvadore, E., Mariethoz, G., & Schaefli, B. (2020). Potential of satellite and reanalysis evaporation datasets for hydrological

modelling under various model calibration strategies. *Advances in Water Resources*. doi: 10.1016/j.advwatres.2020.103667

Dembélé, M., Hrachowitz, M., Savenije, H. H., 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*. doi: 10.1029/2019WR026085

Demirel, M., Koch, J., Mendiguren, G., & Stisen, S. (2018, 9). Spatial Pattern Oriented Multicriteria Sensitivity Analysis of a Distributed Hydrologic Model. *Water*, *10*(9), 1188. doi: 10.3390/w10091188

Demirel, M. C., Koch, J., Rakovec, O., Kumar, R., Mai, J., Müller, S., ... Stisen, S. (2024, 1). Tradeoffs Between Temporal and Spatial Pattern Calibration and Their Impacts on Robustness and Transferability of Hydrologic Model Parameters to Ungauged Basins. *Water Resources Research*, *60*(1). doi: 10.1029/2022WR034193

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*. doi: 10.5194/hess-22-1299-2018

Denager, T., Sonnenborg, T. O., Looms, M. C., Bogena, H., & Jensen, K. H. (2023, 7). Point-scale multi-objective calibration of the Community Land Model (version 5.0) using in situ observations of water and energy fluxes and variables. *Hydrology and Earth System Sciences*, *27*(14), 2827–2845. doi: 10.5194/hess-27-2827-2023

Dettinger, Cayan, D., Meyer, M., & Jeton, A. (2004). Simulated Hydrologic Responses To Climate Variations. *Climatic Change*.

DGA. (2017). Actualización del Balance Hídrico Nacional. Informe Final. SIT N° 417, Ministerio de Obras Públicas, Dirección General de Aguas, División de Estudios y Planificación, Santiago, Chile, Realizado por: Universidad de Chile & Pontificia Universidad Católica. , 378. Retrieved from `https://snia.mop.gob.cl/sad/REH5796v1.pdf%0Ahttps://snia.mop.gob.cl/sad/USO5795v2.pdf%0Ahttps://snia.mop.gob.cl/sad/USO5795v3.pdf`

DGA. (2018). *Aplicación de la Metodología de Actualización del Balance Hídrico Nacional en las cuencas de las Macrozonas Norte y Centro, SIT N° 435* (Tech. Rep.). Ministerio de Obras Públicas, Dirección General de Aguas, División de Estudios y Planificación, Santiago, Chile. Elaborado por Fundación para la Transferencia Tecnológica y Pontificia Universidad Católica de Chile.

DGA. (2019a). *Aplicación de la Metodología de Actualización del Balance Hídrico Nacional en la Macrozonas Sur y parte Norte de la Macrozona Austral, SIT N° 441* (Tech. Rep.). Ministerio de Obras Públicas, Dirección General de Aguas, DIvisión de Estudios y Planificación, Santiago, Chile. Elaborado por Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.

DGA. (2019b). *Aplicación de la metodología de actualización del balance hídrico nacional en las cuencas de la parte sur de la Macrozona Autral e Isla de Pascua, SIT N° 444.* (Tech. Rep.). Santiago, Chile: Dirección General de Aguas.

DGA. (2022). *Homologación del cálculo hidrológico para la estimación de la oferta natural del agua histórica y futura en Chile.* (Tech. Rep.). SIT N° 524. Ministerio de Obras Públicas, Dirección General de Aguas, División de Estudios y Planificación, Chile. Elaborado por Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas. Retrieved from `https://snia.mop.gob.cl/repositoriodga/handle/20.500.13000/126394`

Dickinson, R. E. (1984). Modeling evapotranspiration for three-dimensional global climate models. In J. E. Hansen & T. Takahashi (Eds.), *Climate processes and climate*

sensitivities (Vol. 29, pp. 58–72). doi: 10.1029/GM029p0058

Di Virgilio, G., Ji, F., Tam, E., Nishant, N., Evans, J. P., Thomas, C., ... Delage, F. (2022, 4). Selecting CMIP6 GCMs for CORDEX Dynamical Downscaling: Model Performance, Independence, and Climate Change Signals. *Earth's Future*, *10*(4). doi: 10.1029/2021EF002625

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., ... Lecomte, P. (2017, 12). ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sensing of Environment*, *203*, 185–215. doi: 10.1016/J.RSE.2017.07.001

dos Santos Araujo, D. C., Gico Lima Montenegro, S. M., Ribeiro Neto, A., & da Silva, S. F. (2024, 1). Evaluation of satellite-based soil moisture for agricultural drought monitoring in the Brazilian semiarid region. *Remote Sensing Applications: Society and Environment*, *33*, 101111. doi: 10.1016/j.rsase.2023.101111

Draper, C. S., Walker, J. P., Steinle, P. J., de Jeu, R. A., & Holmes, T. R. (2009, 4). An evaluation of AMSR–E derived soil moisture over Australia. *Remote Sensing of Environment*, *113*(4), 703–710. doi: 10.1016/j.rse.2008.11.011

Duethmann, D., Bloschl, G., & Parajka, J. (2020). Why does a conceptual hydrological model fail to correctly predict discharge changes in response to climate change? *Hydrology and Earth System Sciences*. doi: 10.5194/hess-24-3493-2020

Duethmann, D., Peters, J., Blume, T., Vorogushyn, S., & Güntner, A. (2014, 3). The value of satellite-derived snow cover images for calibrating a hydrological model in snow-dominated catchments in Central Asia. *Water Resources Research*, *50*(3), 2002–2021. doi: 10.1002/2013WR014382

Dussaillant, I., Berthier, E., Brun, F., Masiokas, M., Hugonnet, R., Favier, V., ... Ruiz, L. (2019). Two decades of glacier mass loss along the Andes. *Nature Geoscience*. doi: 10.1038/s41561-019-0432-5

Eshel, G., Araus, V., Undurraga, S., Soto, D. C., Moraga, C., Montecinos, A., ... Gutiérrez, R. A. (2021, 11). Plant ecological genomics at the limits of life in the Atacama Desert. *Proceedings of the National Academy of Sciences*, *118*(46). doi: 10.1073/pnas.2101177118

Fan, Y., Clark, M., Lawrence, D. M., Swenson, S., Band, L. E., Brantley, S. L., ... Yamazaki, D. (2019). Hillslope Hydrology in Global Change Research and Earth System Modeling. *Water Resources Research*. doi: 10.1029/2018WR023903

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., ... Alsdorf, D. (2007, 5). The Shuttle Radar Topography Mission. *Reviews of Geophysics*, *45*(2), RG2004. doi: 10.1029/2005RG000183

Fatolazadeh, F., Eshagh, M., & Goïta, K. (2022, 12). New spectro-spatial downscaling approach for terrestrial and groundwater storage variations estimated by GRACE models. *Journal of Hydrology*, *615*, 128635. doi: 10.1016/j.jhydrol.2022.128635

Fisher, R. A., & Koven, C. D. (2020, 4). Perspectives on the Future of Land Surface Models and the Challenges of Representing Complex Terrestrial Systems. *Journal of Advances in Modeling Earth Systems*, *12*(4). doi: 10.1029/2018MS001453

Fowler, H. J., Blenkinsop, S., & Tebaldi, C. (2007, 10). Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology*, *27*(12), 1547–1578. doi: 10.1002/JOC.1556

Fowler, K., Coxon, G., Freer, J., Peel, M., Wagener, T., Western, A., ... Zhang, L. (2018). Simulating Runoff Under Changing Climatic Conditions: A Framework for Model Improvement. *Water Resources Research*. doi: 10.1029/2018WR023989

Fowler, K., Peel, M., Western, A., & Zhang, L. (2018). Improved Rainfall-Runoff Calibra-

tion for Drying Climate: Choice of Objective Function. *Water Resources Research*. doi: 10.1029/2017WR022466

Franchini, M., & Pacciani, M. (1991, 1). Comparative analysis of several conceptual rainfall-runoff models. *Journal of Hydrology*, *122*(1-4), 161–219. doi: 10.1016/0022-1694(91)90178-K

François, B., Vrac, M., Cannon, A. J., Robin, Y., & Allard, D. (2020, 6). Multivariate bias corrections of climate simulations: which benefits for which losses? *Earth System Dynamics*, *11*(2), 537–562. doi: 10.5194/esd-11-537-2020

Frêne, C., Armesto, J. J., Nespolo, R. F., Gaxiola, A., Navarrete, S. A., Troncoso, A., ... Corcuera, L. J. (2023, 3). Chilean long-term Socio-Ecological Research Network: progresses and challenges towards improving stewardship of unique ecosystems. *Revista Chilena de Historia Natural*, *96*(1), 1. doi: 10.1186/s40693-023-00114-4

Gao, M., Kim, S.-J., Yang, J., Liu, J., Jiang, T., Su, B., ... Huang, J. (2021, 6). Historical fidelity and future change of Amundsen Sea Low under 1.5 °C–4 °C global warming in CMIP6. *Atmospheric Research*, *255*, 105533. doi: 10.1016/j.atmosres.2021.105533

Garcia, F., Folton, N., & Oudin, L. (2017, 5). Which objective function to calibrate rainfall–runoff models for low-flow index simulations? *Hydrological Sciences Journal*, *62*(7), 1149–1166. doi: 10.1080/02626667.2017.1308511

Garreaud, R., Vuille, M., & Clement, A. C. (2003, 5). The climate of the Altiplano: observed current conditions and mechanisms of past changes. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *194*(1-3), 5–22. doi: 10.1016/S0031-0182(03)00269-4

Garreaud, R., Vuille, M., Compagnucci, R., & Marengo, J. (2009, 10). Present-day South American climate. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *281*(3-4), 180–195. doi: 10.1016/j.palaeo.2007.10.032

Garreaud, R. D., Alvarez-Garreton, C., Barichivich, J., Boisier, J. P., Christie, D., Galleguillos, M., ... Zambrano-Bigiarini, M. (2017, 12). The 2010–2015 megadrought in central Chile: impacts on regional hydroclimate and vegetation. *Hydrology and Earth System Sciences*, *21*(12), 6307–6327. doi: 10.5194/hess-21-6307-2017

Garreaud, R. D., Boisier, J. P., Rondanelli, R., Montecinos, A., Sepúlveda, H. H., & Veloso-Aguila, D. (2020, 1). The Central Chile Mega Drought (2010–2018): A climate dynamics perspective. *International Journal of Climatology*, *40*(1), 421–439. doi: 10.1002/joc.6219

Gateño, F., Mendoza, P. A., Vásquez, N., Lagos-Zúñiga, M., Jiménez, H., Jerez, C., ... Montserrat, S. (2024, 6). Screening CMIP6 models for Chile based on past performance and code genealogy. *Climatic Change*, *177*(6), 87. doi: 10.1007/s10584-024-03742-1

Geophysics Department, U. d. C., & Ministerio de Energía, G. d. C. (2018). *Explorador Eólico* (Tech. Rep.). Geophysic Deparment, Universidad de Chile. Retrieved from `https://eolico.minenergia.cl/inicio`

Ghimire, U., Srinivasan, G., & Agarwal, A. (2019, 3). Assessment of rainfall bias correction techniques for improved hydrological simulation. *International Journal of Climatology*, *39*(4), 2386–2399. doi: 10.1002/joc.5959

Güntner, A. (2008, 10). Improvement of Global Hydrological Models Using GRACE Data. *Surveys in Geophysics*, *29*(4-5), 375–397. doi: 10.1007/s10712-008-9038-y

Guo, Q., Chen, J., Zhang, X. J., Xu, C., & Chen, H. (2020, 5). Impacts of Using State-of-the-Art Multivariate Bias Correction Methods on Hydrological Modeling Over North America. *Water Resources Research*, *56*(5). doi: 10.1029/2019WR026659

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009, 10). Decomposition of the mean squared error and NSE performance criteria: Implications for improving

hydrological modelling. *Journal of Hydrology*, *377*(1-2), 80–91. doi: 10.1016/j.jhydrol.2009.08.003

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*. doi: 10.1029/97WR03495

Gutiérrez, J. M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., ... Pagé, C. (2019, 7). An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment. *International Journal of Climatology*, *39*(9), 3750–3785. doi: 10.1002/joc.5462

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*. doi: 10.1002/2014WR015559

Haerter, J. O., Hagemann, S., Moseley, C., & Piani, C. (2011, 3). Climate model bias correction and the role of timescales. *Hydrology and Earth System Sciences*, *15*(3), 1065–1079. doi: 10.5194/hess-15-1065-2011

Hagemann, S., Chen, C., Haerter, J. O., Heinke, J., Gerten, D., & Piani, C. (2011, 8). Impact of a Statistical Bias Correction on the Projected Hydrological Changes Obtained from Three GCMs and Two Hydrology Models. *Journal of Hydrometeorology*, *12*(4), 556–578. doi: 10.1175/2011JHM1336.1

Hakala, K., Addor, N., & Seibert, J. (2018, 8). Hydrological Modeling to Evaluate Climate Model Simulations and Their Bias Correction. *Journal of Hydrometeorology*, *19*(8), 1321–1337. doi: 10.1175/JHM-D-17-0189.1

Hall, D. K., & Riggs, G. A. (2016). MODIS/Terra snow cover Daily L3 Global 500m SIN GRID, version 6. *Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MOD10A1.006*.

Han, P., Long, D., Han, Z., Du, M., Dai, L., & Hao, X. (2019, 4). Improved understanding of snowmelt runoff from the headwaters of China's Yangtze River using remotely sensed snow products and hydrological modeling. *Remote Sensing of Environment*, *224*, 44–59. doi: 10.1016/j.rse.2019.01.041

Hanus, S., Hrachowitz, M., Zekollari, H., Schoups, G., Vizcaino, M., & Kaitna, R. (2021). Future changes in annual, seasonal and monthly runoff signatures in contrasting Alpine catchments in Austria. *Hydrology and Earth System Sciences*. doi: 10.5194/hess-25-3429-2021

Hao, D., Bisht, G., Huang, M., Ma, P., Tesfa, T., Lee, W., ... Leung, L. R. (2022, 4). Impacts of Sub-Grid Topographic Representations on Surface Energy Balance and Boundary Conditions in the E3SM Land Model: A Case Study in Sierra Nevada. *Journal of Advances in Modeling Earth Systems*, *14*(4). doi: 10.1029/2021MS002862

Hattermann, F. F., Vetter, T., Breuer, L., Su, B., Daggupati, P., Donnelly, C., ... Krysnaova, V. (2018). Sources of uncertainty in hydrological climate impact assessment: A cross-scale study. *Environmental Research Letters*. doi: 10.1088/1748-9326/aa9938

Hawkins, E., Frame, D., Harrington, L., Joshi, M., King, A., Rojas, M., & Sutton, R. (2020, 3). Observed Emergence of the Climate Change Signal: From the Familiar to the Unknown. *Geophysical Research Letters*, *47*(6). doi: 10.1029/2019GL086259

Hawkins, E., & Sutton, R. (2012, 1). Time of emergence of climate signals. *Geophysical Research Letters*, *39*(1). doi: 10.1029/2011GL050087

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., & Piontek, F. (2013, 7). A trend-preserving bias correction – the ISI-MIP approach. *Earth System Dynamics*, *4*(2), 219–236. doi: 10.5194/esd-4-219-2013

Her, Y., & Chaubey, I. (2015, 9). Impact of the numbers of observations and calibration parameters on equifinality, model performance, and output and parameter uncertainty. *Hydrological Processes*, *29*(19), 4220–4237. doi: 10.1002/hyp.10487

Her, Y., Yoo, S. H., Cho, J., Hwang, S., Jeong, J., & Seong, C. (2019). Uncertainty in hydrological analysis of climate change: multi-parameter vs. multi-GCM ensemble predictions. *Scientific Reports*. doi: 10.1038/s41598-019-41334-7

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... Thépaut, J. (2020, 7). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. doi: 10.1002/qj.3803

Hess, P., Lange, S., Schötz, C., & Boers, N. (2023, 10). Deep Learning for Bias-Correcting CMIP6-Class Earth System Models. *Earth's Future*, *11*(10). doi: 10.1029/2023EF004002

Heusser, C. J. (1974, 9). Vegetation and Climate of the Southern Chilean lake District During and Since the last Interglaciation. *Quaternary Research*, *4*(3), 290–315. doi: 10.1016/0033-5894(74)90018-0

IPCC. (2023, 6). Technical Summary. In *Climate change 2022 – impacts, adaptation and vulnerability* (pp. 37–118). Cambridge University Press. doi: 10.1017/9781009325844.002

Iziomon, M. G., Mayer, H., & Matzarakis, A. (2003, 7). Downward atmospheric longwave irradiance under clear and cloudy skies: Measurement and parameterization. *Journal of Atmospheric and Solar-Terrestrial Physics*, *65*(10), 1107–1116. doi: 10.1016/j.jastp.2003.07.007

Jain, S., Scaife, A. A., Shepherd, T. G., Deser, C., Dunstone, N., Schmidt, G. A., ... Turkington, T. (2023, 6). Importance of internal variability for climate model assessment. *npj Climate and Atmospheric Science*, *6*(1), 68. doi: 10.1038/s41612-023-00389-0

Jennings, K. S., Winchell, T. S., Livneh, B., & Molotch, N. P. (2018, 3). Spatial variation of the rain–snow temperature threshold across the Northern Hemisphere. *Nature Communications*, *9*(1), 1148. doi: 10.1038/s41467-018-03629-7

Jiménez-Navarro, I. C., Pierson, D., & Senent-Aparicio, J. (2024, 8). The Implications of the Use of ET Remote Sensing Data for Calibrating Hydrological Models: A Comparison of Single and Multi-criteria Calibration in SWAT+. *Earth Systems and Environment*. doi: 10.1007/s41748-024-00438-5

Johnson, F., & Sharma, A. (2011, 4). Accounting for interannual variability: A comparison of options for water resources climate change impact assessments. *Water Resources Research*, *47*(4). doi: 10.1029/2010WR009272

Keller, A. A., Garner, K., Rao, N., Knipping, E., & Thomas, J. (2023, 4). Hydrological models for climate-based assessments at the watershed scale: A critical review of existing hydrologic and water quality models. *Science of The Total Environment*, *867*, 161209. doi: 10.1016/j.scitotenv.2022.161209

Khatami, S., Peel, M. C., Peterson, T. J., & Western, A. W. (2019). Equifinality and Flux Mapping: A New Approach to Model Evaluation and Process Representation Under Uncertainty. *Water Resources Research*. doi: 10.1029/2018WR023750

Khorrami, B., Pirasteh, S., Ali, S., Sahin, O. G., & Vaheddoost, B. (2023, 9). Statistical downscaling of GRACE TWSA estimates to a 1-km spatial resolution for a local-scale surveillance of flooding potential. *Journal of Hydrology*, *624*, 129929. doi: 10.1016/j.jhydrol.2023.129929

Kim, Y., Evans, J. P., Sharma, A., & Rocheta, E. (2021, 6). Spatial, Temporal, and Multivariate Bias in Regional Climate Model Simulations. *Geophysical Research Letters*, *48*(11). doi: 10.1029/2020GL092058

Kinar, N. J., & Pomeroy, J. W. (2015, 6). Measurement of the physical properties of the snowpack. *Reviews of Geophysics*, *53*(2), 481–544. doi: 10.1002/2015RG000481

Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*. doi: 10.1029/2005WR004362

Kirchner, J. W., & Allen, S. T. (2020, 1). Seasonal partitioning of precipitation between streamflow and evapotranspiration, inferred from end-member splitting analysis. *Hydrology and Earth System Sciences*, *24*(1), 17–39. doi: 10.5194/hess-24-17-2020

KlemeŠ, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal*. doi: 10.1080/02626668609491024

Kling, H., Fuchs, M., & Paulin, M. (2012, 3). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, *424-425*, 264–277. doi: 10.1016/J.JHYDROL.2012.01.011

Knoben, W. J., Woods, R. A., & Freer, J. E. (2018). A Quantitative Hydrological Climate Classification Evaluated With Independent Streamflow Data. *Water Resources Research*. doi: 10.1029/2018WR022913

Knoben, W. J. M., Freer, J. E., Fowler, K. J. A., Peel, M. C., & Woods, R. A. (2019, 6). Modular Assessment of Rainfall–Runoff Models Toolbox (MARRMoT) v1.2: an open-source, extendable framework providing implementations of 46 conceptual hydrologic models as continuous state-space formulations. *Geoscientific Model Development*, *12*(6), 2463–2480. doi: 10.5194/gmd-12-2463-2019

Koch, J., Demirel, M. C., & Stisen, S. (2018). The SPAtial EFficiency metric (SPAEF): Multiple-component evaluation of spatial patterns for optimization of hydrological models. *Geoscientific Model Development*. doi: 10.5194/gmd-11-1873-2018

Koch, J., Mendiguren, G., Mariethoz, G., & Stisen, S. (2017). Spatial sensitivity analysis of simulated land surface patterns in a catchment model using a set of innovative spatial performance metrics. *Journal of Hydrometeorology*. doi: 10.1175/JHM-D-16-0148.1

Kollat, J. B., Reed, P. M., & Wagener, T. (2012, 3). When are multiobjective calibration trade-offs in hydrologic models meaningful? *Water Resources Research*, *48*(3). doi: 10.1029/2011WR011534

Koppa, A., Gebremichael, M., & Yeh, W. W. (2019). Multivariate calibration of large scale hydrologic models: The necessity and value of a Pareto optimal approach. *Advances in Water Resources*. doi: 10.1016/j.advwatres.2019.06.005

Kusumastuti, C., Jiang, Z., Mehrotra, R., & Sharma, A. (2022, 10). Correcting Systematic Bias in Climate Model Simulations in the Time-Frequency Domain. *Geophysical Research Letters*, *49*(19). doi: 10.1029/2022GL100550

Kwon, S., Kim, J., Boo, K., Shim, S., Kim, Y., & Byun, Y. (2019, 3). Performance-based projection of the climate-change effects on precipitation extremes in East Asia using two metrics. *International Journal of Climatology*, *39*(4), 2324–2335. doi: 10.1002/joc.5954

Lafon, T., Dadson, S., Buys, G., & Prudhomme, C. (2013, 5). Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods. *International Journal of Climatology*, *33*(6), 1367–1381. doi: 10.1002/joc.3518

Lehner, F., Wood, A. W., Vano, J. A., Lawrence, D. M., Clark, M. P., & Mankin, J. S. (2019). *The potential to reduce uncertainty in regional runoff projections from climate models.* doi: 10.1038/s41558-019-0639-x

Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., & Wood, E. F. (2015, 9). Inroads of remote sensing into hydrologic science during the WRR era. *Water Resources Research*, *51*(9), 7309–7342. doi: 10.1002/2015WR017616

Li, H., Sheffield, J., & Wood, E. F. (2010, 5). Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *Journal of Geophysical Research: Atmospheres*, *115*(D10). doi: 10.1029/2009JD012882

Li, X., Cheng, G., Ge, Y., Li, H., Han, F., Hu, X., ... Cai, X. (2018, 1). Hydrological Cycle in the Heihe River Basin and Its Implication for Water Resource Management in Endorheic Basins. *Journal of Geophysical Research: Atmospheres*, *123*(2), 890–914. doi: 10.1002/2017JD027889

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*. doi: 10.1029/94jd00483

López López, P., Sutanudjaja, E. H., Schellekens, J., Sterk, G., & Bierkens, M. F. P. (2017, 6). Calibration of a large-scale hydrological model using satellite-based soil moisture and evapotranspiration products. *Hydrology and Earth System Sciences*, *21*(6), 3125–3144. doi: 10.5194/hess-21-3125-2017

Madeira, C. (2022, 12). A review of the future impact of climate change in Chile: economic output and other outcomes. *Mitigation and Adaptation Strategies for Global Change*, *27*(8), 56. doi: 10.1007/s11027-022-10034-5

Maity, R., Suman, M., Laux, P., & Kunstmann, H. (2019, 4). Bias Correction of Zero-Inflated RCM Precipitation Fields: A Copula-Based Scheme for Both Mean and Extreme Conditions. *Journal of Hydrometeorology*, *20*(4), 595–611. doi: 10.1175/JHM-D-18-0126.1

Maraun, D. (2016, 12). Bias Correcting Climate Change Simulations - a Critical Review. *Current Climate Change Reports*, *2*(4), 211–220. doi: 10.1007/s40641-016-0050-x

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., ... Thiele-Eich, I. (2010, 9). Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, *48*(3), RG3003. doi: 10.1029/2009RG000314

Mardones, P., & Garreaud, R. D. (2020, 11). Future Changes in the Free Tropospheric Freezing Level and Rain–Snow Limit: The Case of Central Chile. *Atmosphere*, *11*(11), 1259. doi: 10.3390/atmos11111259

Marquet, P. A., Buschmann, A. H., Corcoran, D., Díaz, P. A., Fuentes-Castillo, T., Garreaud, R., ... Salazar, A. (2023). Global Change and Acceleration of Anthropic Pressures on Patagonian Ecosystems. In (pp. 33–65). doi: 10.1007/978-3-031-39408-9{_}2

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., ... Verhoest, N. E. C. (2017, 5). GLEAM v3: satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, *10*(5), 1903–1925. doi: 10.5194/gmd-10-1903-2017

Martínez-Retureta, R., Aguayo, M., Abreu, N. J., Stehr, A., Duran-Llacer, I., Rodríguez-López, L., ... Sánchez-Pérez, J.-M. (2021, 3). Estimation of the Climate Change Impact on the Hydrological Balance in Basins of South-Central Chile. *Water*, *13*(6), 794. doi: 10.3390/w13060794

Matiu, M., & Hanzer, F. (2022, 6). Bias adjustment and downscaling of snow cover fraction projections from regional climate models using remote sensing for the European Alps. *Hydrology and Earth System Sciences*, *26*(12), 3037–3054. doi: 10.5194/hess-26-3037-2022

Matott, L. (2017). *OSTRICH: an Optimization Software Tool, Documentation and User's Guide, Version 17.12.19.* Retrieved from http://www.civil.uwaterloo.ca/envmodelling/Ostrich.html

Maurer, E. P., & Pierce, D. W. (2014, 3). Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean. *Hydrology and Earth System Sciences*, *18*(3), 915–925. doi: 10.5194/hess-18-915-2014

McCabe, M. F., Rodell, M., Alsdorf, D. E., Miralles, D. G., Uijlenhoet, R., Wagner, W., ... Wood, E. F. (2017). The future of Earth observation in hydrology. *Hydrology and Earth System Sciences*. doi: 10.5194/hess-21-3879-2017

McInerney, D., Thyer, M., Kavetski, D., Lerat, J., & Kuczera, G. (2017). Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. *Water Resources Research*. doi: 10.1002/2016WR019168

McInerney, D., Thyer, M., Kavetski, D., Westra, S., Maier, H. R., Shanafield, M., ... Leonard, M. (2024, 5). Neglecting hydrological errors can severely impact predictions of water resource system performance. *Journal of Hydrology*, *634*, 130853. doi: 10.1016/j.jhydrol.2024.130853

McMillan, H. K., Gnann, S. J., & Araki, R. (2022, 6). Large Scale Evaluation of Relationships Between Hydrologic Signatures and Processes. *Water Resources Research*, *58*(6). doi: 10.1029/2021WR031751

Mehboob, M. S., & Kim, Y. (2024, 1). Impact of climate change on the hydrological projections over a western Himalayan river basin and the associated uncertainties. *Journal of Hydrology*, *628*, 130460. doi: 10.1016/j.jhydrol.2023.130460

Mei, Y., Mai, J., Do, H. X., Gronewold, A., Reeves, H., Eberts, S., ... Hunt, R. J. (2023, 2). Can Hydrological Models Benefit From Using Global Soil Moisture, Evapotranspiration, and Runoff Products as Calibration Targets? *Water Resources Research*, *59*(2). doi: 10.1029/2022WR032064

Melsen, L., Teuling, A., Torfs, P., Zappa, M., Mizukami, N., Clark, M., & Uijlenhoet, R. (2016, 6). Representation of spatial and temporal variability in large-domain hydrological models: Case study for a mesoscale pre-Alpine basin. *Hydrology and Earth System Sciences*, *20*(6), 2207–2226. doi: 10.5194/hess-20-2207-2016

Melsen, L. A., Addor, N., Mizukami, N., Newman, A. J., Torfs, P. J., Clark, M. P., ... Teuling, A. J. (2018). Mapping (dis)agreement in hydrologic projections. *Hydrology and Earth System Sciences*. doi: 10.5194/hess-22-1775-2018

Melsen, L. A., Teuling, A. J., Torfs, P. J., Zappa, M., Mizukami, N., Mendoza, P. A., ... Uijlenhoet, R. (2019, 1). Subjective modeling decisions can significantly impact the simulation of flood and drought events. *Journal of Hydrology*, *568*, 1093–1104. doi: 10.1016/J.JHYDROL.2018.11.046

Mendoza, P. A., Clark, M. P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G., & Gupta, H. (2015). *Are we unnecessarily constraining the agility of complex process-based models?* doi: 10.1002/2014WR015820

Mendoza, P. A., Clark, M. P., Mizukami, N., Gutmann, E. D., Arnold, J. R., Brekke, L. D., & Rajagopalan, B. (2016). How do hydrologic modeling decisions affect the portrayal of climate change impacts? *Hydrological Processes*. doi: 10.1002/hyp.10684

Mendoza, P. A., Clark, M. P., Mizukami, N., Newman, A. J., Barlage, M., Gutmann, E. D., ... Arnold, J. R. (2015, 4). Effects of hydrologic model choice and calibration on the portrayal of climate change impacts. *Journal of Hydrometeorology*, *16*(2), 762–780. Retrieved from https://journals.ametsoc.org/view/journals/hydr/16/2/jhm-d-14-0104_1.xml doi: 10.1175/JHM-D-14-0104.1

Mendoza, P. A., McPhee, J., & Vargas, X. (2012, 9). Uncertainty in flood forecasting: A distributed modeling approach in a sparse data catchment. *Water Resources Research*, *48*(9). doi: 10.1029/2011WR011089

Meresa, H., Murphy, C., Fealy, R., & Golian, S. (2021, 9). Uncertainties and their interaction in flood hazard assessment with climate change. *Hydrology and Earth System Sciences*, *25*(9), 5237–5257. doi: 10.5194/hess-25-5237-2021

Merz, R., Parajka, J., & Blöschl, G. (2011). Time stability of catchment model parameters: Implications for climate impact analyses. *Water Resources Research*. doi: 10.1029/2010WR009505

Meyer, J., Kohn, I., Stahl, K., Hakala, K., Seibert, J., & Cannon, A. J. (2019, 3). Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments. *Hydrology and Earth System Sciences*, *23*(3), 1339–1354. doi: 10.5194/hess-23-1339-2019

Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B., ... Samaniego, L. (2017, 9). Towards seamless large-domain parameter estimation for hydrologic models. *Water Resources Research*, *53*(9), 8020–8040. doi: 10.1002/2017WR020401

Mizukami, N., Clark, M. P., Sampson, K., Nijssen, B., Mao, Y., McMillan, H., ... Brekke, L. D. (2016, 6). mizuRoute version 1: a river network routing tool for a continental domain water resources applications. *Geoscientific Model Development*, *9*(6), 2223–2238. doi: 10.5194/gmd-9-2223-2016

Mizukami, N., Rakovec, O., Newman, A. J., Clark, M. P., Wood, A. W., Gupta, H. V., & Kumar, R. (2019, 6). On the choice of calibration metrics for "high-flow" estimation using hydrologic models. *Hydrology and Earth System Sciences*, *23*(6), 2601–2614. doi: 10.5194/hess-23-2601-2019

MMA. (2016). *Chile's Third National Communication on Climate Change to the United Nations Framework Convention on Climate Change* (Tech. Rep.). Ministry of Environment, Government of Chile. Retrieved from `https://snichile.mma.gob.cl/wp-content/uploads/2022/06/2016_es3nc_chile.pdf`

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*.

Mu, Q., Zhao, M., & Running, S. W. (2011, 8). Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment*, *115*(8), 1781–1800. doi: 10.1016/j.rse.2011.02.019

Muñoz-Castro, E., Mendoza, P. A., Vásquez, N. A., & Vargas, X. (2023, 6). Exploring parameter (dis)agreement due to calibration metric selection in conceptual rainfall-runoff models. *Hydrological Sciences Journal*. doi: 10.1080/02626667.2023.2231434

Muñoz Sabater, J. (2019). *ERA5-Land hourly data from 1981 to present.*

Muñoz-Sáez, A., Choe, H., Boynton, R. M., Elsen, P. R., & Thorne, J. H. (2021, 9). Climate exposure shows high risk and few climate refugia for Chilean native vegetation. *Science of The Total Environment*, *785*, 147399. doi: 10.1016/j.scitotenv.2021.147399

Murillo, O., Mendoza, P. A., Vásquez, N. A., Mizukami, N., & Ayala, A. (2022). Impacts of subgrid elevation bands on hydrological portrayals: insights from a suite of hydroclimatically diverse mountainous catchments. *Earth and Space Science Open Archive, submitted to Water Resources Research*, 31. Retrieved from `https://doi.org/10.1002/essoar.10510847.1` doi: 10.1002/essoar.10510847.1

Musselman, K. N., Clark, M. P., Liu, C., Ikeda, K., & Rasmussen, R. (2017, 3). Slower snowmelt in a warmer world. *Nature Climate Change*, *7*(3), 214–219. doi: 10.1038/

nclimate3225

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology*. doi: 10.1016/0022-1694(70)90255-6

Nijzink, R. C., Almeida, S., Pechlivanidis, I. G., Capell, R., Gustafssons, D., Arheimer, B., ... Hrachowitz, M. (2018). Constraining Conceptual Hydrological Models With Multiple Information Sources. *Water Resources Research*. doi: 10.1029/2017WR021895

Nijzink, R. C., Samaniego, L., Mai, J., Kumar, R., Thober, S., Zink, M., ... Hrachowitz, M. (2016, 3). The importance of topography-controlled sub-grid process heterogeneity and semi-quantitative prior constraints in distributed hydrological models. *Hydrology and Earth System Sciences*, *20*(3), 1151–1176. doi: 10.5194/hess-20-1151-2016

Niño, Y. (2002, 10). Simple Model for Downstream Variation of Median Sediment Size in Chilean Rivers. *Journal of Hydraulic Engineering*, *128*(10), 934–941. doi: 10.1061/(ASCE)0733-9429(2002)128:10(934)

Odusanya, A. E., Schulz, K., Biao, E. I., Degan, B. A., & Mehdi-Schulz, B. (2021, 10). Evaluating the performance of streamflow simulated by an eco-hydrological model calibrated and validated with global land surface actual evapotranspiration from remote sensing at a catchment scale in West Africa. *Journal of Hydrology: Regional Studies*, *37*, 100893. doi: 10.1016/j.ejrh.2021.100893

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., ... Sanderson, B. M. (2016, 9). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482. doi: 10.5194/gmd-9-3461-2016

Ortega, G., Arias, P. A., Villegas, J. C., Marquet, P. A., & Nobre, P. (2021, 12). Present-day and future climate over central and South America according to ¡scp¿CMIP5¡/scp¿ / ¡scp¿CMIP6¡/scp¿ models. *International Journal of Climatology*, *41*(15), 6713–6735. doi: 10.1002/joc.7221

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005, 3). Which potential evapotranspiration input for a lumped rainfall–runoff model? *Journal of Hydrology*, *303*(1-4), 290–306. doi: 10.1016/j.jhydrol.2004.08.026

Parajka, J., & Blöschl, G. (2008, 9). The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models. *Journal of Hydrology*, *358*(3-4), 240–258. doi: 10.1016/j.jhydrol.2008.06.006

Peel, M. C., & Blöschl, G. (2011, 3). Hydrological modelling in a changing world:. *http://dx.doi.org/10.1177/0309133311402550*, *35*(2), 249–261. doi: 10.1177/0309133311402550

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007, 10). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, *11*(5), 1633–1644. doi: 10.5194/hess-11-1633-2007

Pelletier, A., & Andréassian, V. (2022, 5). On constraining a lumped hydrological model with both piezometry and streamflow: results of a large sample evaluation. *Hydrology and Earth System Sciences*, *26*(10), 2733–2758. doi: 10.5194/hess-26-2733-2022

Pierce, D. W., Cayan, D. R., Maurer, E. P., Abatzoglou, J. T., & Hegewisch, K. C. (2015, 12). Improved Bias Correction Techniques for Hydrological Simulations of Climate Change*. *Journal of Hydrometeorology*, *16*(6), 2421–2442. doi: 10.1175/JHM-D-14-0236.1

Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G. B., Kempen, B., Ribeiro, E., & Rossiter, D. (2021, 6). SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *SOIL*, *7*(1), 217–240. doi: 10.5194/SOIL-7-217-2021

Pokhrel, P., & Gupta, H. V. (2010). On the use of spatial regularization strategies to improve calibration of distributed watershed models. *Water Resources Research*. doi: 10.1029/2009WR008066

Pokhrel, P., Gupta, H. V., & Wagener, T. (2008). A spatial regularization approach to parameter estimation for a distributed watershed model. *Water Resources Research*. doi: 10.1029/2007WR006615

Pool, S., Fowler, K., & Peel, M. (2024, 4). Benefit of Multivariate Model Calibration for Different Climatic Regions. *Water Resources Research*, *60*(4). doi: 10.1029/2023WR036364

Pool, S., Vis, M., & Seibert, J. (2018). Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency. *Hydrological Sciences Journal*. doi: 10.1080/02626667.2018.1552002

R Core Team. (2023). *R: A Language and Environment for Statistical Computing.* Vienna: https://www.R-project.org/.

Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., ... Samaniego, L. (2016). Multiscale and multivariate evaluation of water fluxes and states over european river Basins. *Journal of Hydrometeorology*. doi: 10.1175/JHM-D-15-0054.1

Rakovec, O., Mizukami, N., Kumar, R., Newman, A. J., Thober, S., Wood, A. W., ... Samaniego, L. (2019). Diagnostic Evaluation of Large-Domain Hydrologic Models Calibrated Across the Contiguous United States. *Journal of Geophysical Research: Atmospheres*. doi: 10.1029/2019JD030767

Rastogi, D., Kao, S., & Ashfaq, M. (2022, 8). How May the Choice of Downscaling Techniques and Meteorological Reference Observations Affect Future Hydroclimate Projections? *Earth's Future*, *10*(8). doi: 10.1029/2022EF002734

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., & Seo, D. J. (2004, 10). Overall distributed model intercomparison project results. *Journal of Hydrology*, *298*(1-4), 27–60. doi: 10.1016/J.JHYDROL.2004.03.031

Refsgaard, J. C. (1997, 11). Parameterisation, calibration and validation of distributed hydrological models. *Journal of Hydrology*, *198*(1-4), 69–97. doi: 10.1016/S0022-1694(96)03329-X

Refsgaard, J. C., & Knudsen, J. (1996, 7). Operational validation and intercomparison of different types of hydrological models. *Water Resources Research*, *32*(7), 2189–2202. doi: 10.1029/96WR00896

Reiter, P., Gutjahr, O., Schefczyk, L., Heinemann, G., & Casper, M. (2018, 3). Does applying quantile mapping to subsamples improve the bias correction of daily precipitation? *International Journal of Climatology*, *38*(4), 1623–1633. doi: 10.1002/joc.5283

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., ... Tavoni, M. (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*. doi: 10.1016/j.gloenvcha.2016.05.009

Rijsberman, F. R. (2006, 2). Water scarcity: Fact or fiction? *Agricultural Water Management*, *80*(1-3), 5–22. doi: 10.1016/J.AGWAT.2005.07.001

Ruffault, J., Martin-StPaul, N. K., Duffet, C., Goge, F., & Mouillot, F. (2014, 7). Projecting future drought in Mediterranean forests: bias correction of climate models matters! *Theoretical and Applied Climatology*, *117*(1-2), 113–122. doi:

10.1007/s00704-013-0992-z

Saavedra, D., Mendoza, P. A., Addor, N., Llauca, H., & Vargas, X. (2022). A multi-objective approach to select hydrological models and constrain structural uncertainties for climate impact assessments. *Hydrological Processes*. doi: 10.1002/hyp.14446

Salazar, A., Thatcher, M., Goubanova, K., Bernal, P., Gutiérrez, J., & Squeo, F. (2023, 12). CMIP6 precipitation and temperature projections for Chile. *Climate Dynamics*. doi: 10.1007/s00382-023-07034-9

Samaniego, L., Kumar, R., & Attinger, S. (2010, 5). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*(5), 5523. doi: 10.1029/2008WR007327

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., & Carrillo, G. (2011). Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA. *Hydrology and Earth System Sciences Discussions*. doi: 10.5194/hessd-8-4495-2011

Schoups, G., Lee Addams, C., & Gorelick, S. M. (2005, 12). Multi-objective calibration of a surface water-groundwater flow model in an irrigated agricultural region: Yaqui Valley, Sonora, Mexico. *Hydrology and Earth System Sciences*, *9*(5), 549–568. doi: 10.5194/hess-9-549-2005

Schwemmle, R., Demand, D., & Weiler, M. (2020). Technical note: Diagnostic efficiency – specific evaluation of model performance. *Hydrology and Earth System Sciences Discussions*. doi: 10.5194/hess-2020-237

SEIA. (2023a, 11). *Criterio de evaluación en el SEIA: Cambio climático en la evaluación ambiental del recurso hídrico* (Tech. Rep.). Santiago de Chile: Servicio de Evaluación de Impacto Ambiental. Retrieved from `https://www.sea.gob.cl/documentacion/guias-y-criterios/criterio-de-evaluacion-en-el-seia-cambio-climatico-en-la-evaluacion`

SEIA. (2023b, 11). *Guía Metodológica para la Consideración del Cambio Climático en el SEIA* (Tech. Rep.). Santiago de Chile: Servicio de Evaluación de Impacto Ambiental. Retrieved from `https://www.sea.gob.cl/documentacion/guias-y-criterios/guia-metodologica-para-la-consideracion-del-cambio-climatico-en-0`

Senatore, A., Fuoco, D., Maiolo, M., Mendicino, G., Smiatek, G., & Kunstmann, H. (2022, 8). Evaluating the uncertainty of climate model structure and bias correction on the hydrological impact of projected climate change in a Mediterranean catchment. *Journal of Hydrology: Regional Studies*, *42*, 101120. doi: 10.1016/j.ejrh.2022.101120

Sepúlveda, U. M., Mendoza, P. A., Mizukami, N., & Newman, A. J. (2022, 7). Revisiting parameter sensitivities in the variable infiltration capacity model across a hydroclimatic gradient. *Hydrology and Earth System Sciences*, *26*(13), 3419–3445. doi: 10.5194/hess-26-3419-2022

Shafii, M., & Tolson, B. A. (2015). Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. *Water Resources Research*. doi: 10.1002/2014WR016520

Shah, S., Duan, Z., Song, X., Li, R., Mao, H., Liu, J., ... Wang, M. (2021, 12). Evaluating the added value of multi-variable calibration of SWAT with remotely sensed evapotranspiration data for improving hydrological modeling. *Journal of Hydrology*, *603*, 127046. doi: 10.1016/j.jhydrol.2021.127046

Sheffield, J., Wood, E. F., Pan, M., Beck, H., Coccia, G., Serrat-Capdevila, A., & Verbist, K. (2018, 12). Satellite Remote Sensing for Water Resources Management: Potential for Supporting Sustainable Development in Data-Poor Regions. *Water Resources*

Research, *54*(12), 9724–9758. doi: 10.1029/2017WR022437

Soltani, S. S., Ataie-Ashtiani, B., & Simmons, C. T. (2021, 2). Review of assimilating GRACE terrestrial water storage data into hydrological models: Advances, challenges and opportunities. *Earth-Science Reviews*, *213*, 103487. doi: 10.1016/j.earscirev.2020.103487

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

S.U., S. L., Singh, D., & Shojaei Baghini, M. (2014, 8). A critical review of soil moisture measurement. *Measurement*, *54*, 92–105. doi: 10.1016/j.measurement.2014.04.007

Sulla-Menashe, D., & Friedl, M. A. (2018). *User Guide to Collection 6 MODIS Land Cover (MCD12Q1 and MCD12C1) Product* (Tech. Rep.). Retrieved from `https://doi.org/10.5067/MODIS/MCD12Q1.006` doi: 10.5067/MODIS/MCD12Q1

Switanek, M. B., Troch, P. A., Castro, C. L., Leuprecht, A., Chang, H.-I., Mukherjee, R., & Demaria, E. M. C. (2017, 6). Scaled distribution mapping: a bias correction method that preserves raw climate model projected changes. *Hydrology and Earth System Sciences*, *21*(6), 2649–2666. doi: 10.5194/hess-21-2649-2017

Széles, B., Parajka, J., Hogan, P., Silasari, R., Pavlin, L., Strauss, P., & Blöschl, G. (2020). The Added Value of Different Data Types for Calibrating and Testing a Hydrologic Model in a Small Catchment. *Water Resources Research*. doi: 10.1029/2019WR026153

Tang, G., Clark, M. P., Knoben, W. J. M., Liu, H., Gharari, S., Arnal, L., ... Papalexiou, S. M. (2023, 6). The Impact of Meteorological Forcing Uncertainty on Hydrological Modeling: A Global Analysis of Cryosphere Basins. *Water Resources Research*, *59*(6). doi: 10.1029/2022WR033767

Taylor, G. P., Loikith, P. C., Aragon, C. M., Lee, H., & Waliser, D. E. (2023, 4). CMIP6 model fidelity at simulating large-scale atmospheric circulation patterns and associated temperature and precipitation over the Pacific Northwest. *Climate Dynamics*, *60*(7-8), 2199–2218. doi: 10.1007/s00382-022-06410-1

Taylor, K. (2001, 4). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, *106*(D7), 7183–7192. doi: 10.1029/2000JD900719

Taylor, K., Stouffer, R., & Meehl, G. (2012, 4). An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. doi: 10.1175/BAMS-D-11-00094.1

Teng, J., Potter, N. J., Chiew, F. H. S., Zhang, L., Wang, B., Vaze, J., & Evans, J. P. (2015, 2). How does bias correction of regional climate model precipitation affect modelled runoff? *Hydrology and Earth System Sciences*, *19*(2), 711–728. doi: 10.5194/hess-19-711-2015

Teutschbein, C., & Seibert, J. (2010, 7). Regional Climate Models for Hydrological Impact Studies at the Catchment Scale: A Review of Recent Modeling Strategies. *Geography Compass*, *4*(7), 834–860. doi: 10.1111/j.1749-8198.2010.00357.x

Todorović, A., Grabs, T., & Teutschbein, C. (2022, 9). Advancing traditional strategies for testing hydrological model fitness in a changing climate. *Hydrological Sciences Journal*, *67*(12), 1790–1811. doi: 10.1080/02626667.2022.2104646

Tolson, B. A., & Shoemaker, C. A. (2007, 1). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, *43*(1), 1413. doi: 10.1029/2005WR004723

Tong, R., Parajka, J., Salentinig, A., Pfeil, I., Komma, J., Szeles, B., ... Bloschl, G. (2021, 3). The value of ASCAT soil moisture and MODIS snow cover data for

calibrating a conceptual hydrologic model. *Hydrology and Earth System Sciences*, *25*(3), 1389–1410. doi: 10.5194/hess-25-1389-2021

Torres-Rojas, L., Vergopolan, N., Herman, J. D., & Chaney, N. W. (2022, 12). Towards an Optimal Representation of Sub-Grid Heterogeneity in Land Surface Models. *Water Resources Research*, *58*(12). doi: 10.1029/2022WR032233

Troch, P. A., Carrillo, G., Sivapalan, M., Wagener, T., & Sawicz, K. (2013, 6). Climate-vegetation-soil interactions and long-term hydrologic partitioning: signatures of catchment co-evolution. *Hydrology and Earth System Sciences*, *17*(6), 2209–2217. doi: 10.5194/hess-17-2209-2013

Ukkola, A. M., De Kauwe, M. G., Roderick, M. L., Abramowitz, G., & Pitman, A. J. (2020, 6). Robust Future Changes in Meteorological Drought in ¡scp¿CMIP6¡/scp¿ Projections Despite Uncertainty in Precipitation. *Geophysical Research Letters*, *47*(11). doi: 10.1029/2020GL087820

Urrutia-Jalabert, R., González, M. E., Gonzalez-Reyes, A., Lara, A., & Garreaud, R. (2018, 4). Climate variability and forest fires in central and south-central Chile. *Ecosphere*, *9*(4). doi: 10.1002/ecs2.2171

Vano, J. A., Kim, J. B., Rupp, D. E., & Mote, P. W. (2015). Selecting climate change scenarios using impact-relevant sensitivities. *Geophysical Research Letters*. doi: 10.1002/2015GL063208

Vásquez, N., Cepeda, J., Gómez, T., Mendoza, P. A., Lagos, M., Boisier, J. P., ... Vargas, X. (2021). Catchment-Scale Natural Water Balance in Chile. In (pp. 189–208). Retrieved from `http://link.springer.com/10.1007/978-3-030-56901-3_9` doi: 10.1007/978-3-030-56901-3{_}9

Vásquez, N., & Mendoza, P. A. (2024). *Statistically downscaled and bias-corrected CMIP6 models for Continental Chile under scenario SSP5-8.5 [Dataset].* doi: https://doi.org/10.7910/DVN/O3YBOT

Vásquez, N., Mendoza, P. A., Knoben, W. J. M., Arnal, L., Lagos-Zúñiga, M., Clark, M., & Vargas, X. (2024, 8). The Key Role of Temporal Stratification for GCM Bias Correction in Climate Impact Assessments. *Earth's Future*, *12*(8). doi: 10.1029/2023EF004242

Viale, M., Bianchi, E., Cara, L., Ruiz, L. E., Villalba, R., Pitte, P., ... Zalazar, L. (2019, 5). Contrasting Climates at Both Sides of the Andes in Argentina and Chile. *Frontiers in Environmental Science*, *7*. doi: 10.3389/fenvs.2019.00069

Vicuña, S., Garreaud, R. D., Mcphee, J., Vicuña, S., Garreaud, R. D., & Mcphee, J. (2011). Climate change impacts on the hydrology of a snowmelt driven basin in semiarid Chile. *Climatic Change*, *105*, 469–488. doi: 10.1007/s10584-010-9888-4

Vicuña, S., Gironás, J., Meza, F. J., Cruzat, M. L., Jelinek, M., Bustos, E., ... Bambach, N. (2013, 11). Exploring possible connections between hydrological extreme events and climate change in central south Chile. *Hydrological Sciences Journal*, *58*(8), 1598–1619. doi: 10.1080/02626667.2013.840380

Vicuña, S., McPhee, J., & Garreaud, R. D. (2012). Agriculture Vulnerability to Climate Change in a Snowmelt-Driven Basin in Semiarid Chile. *Journal of Water Resources Planning and Management*. doi: 10.1061/(asce)wr.1943-5452.0000202

Vicuña, S., Vargas, X., Boisier, J. P., Mendoza, P. A., Gómez, T., Vásquez, N., & Cepeda, J. (2021). Impacts of Climate Change on Water Resources in Chile. In B. Fernández & J. Gironás (Eds.), *Water resources of chile* (pp. 347–363). Cham: Springer International Publishing. Retrieved from `https://doi.org/10.1007/978-3-030-56901-3_19` doi: 10.1007/978-3-030-56901-3{_}19

Vishwakarma, B. D., Zhang, J., & Sneeuw, N. (2021, 3). Downscaling GRACE total water storage change using partial least squares regression. *Scientific Data*, *8*(1),

95. doi: 10.1038/s41597-021-00862-6

Vogel, E., Johnson, F., Marshall, L., Bende-Michl, U., Wilson, L., Peter, J. R., ... Duong, V. C. (2023, 7). An evaluation framework for downscaling and bias correction in climate change impact studies. *Journal of Hydrology*, *622*, 129693. doi: 10.1016/j.jhydrol.2023.129693

Vrac, M., & Thao, S. (2020, 11). R2D2; v2.0: accounting for temporal dependences in multivariate bias correction via analogue rank resampling. *Geoscientific Model Development*, *13*(11), 5367–5387. doi: 10.5194/gmd-13-5367-2020

Vrac, M., Thao, S., & Yiou, P. (2022, 7). Should Multivariate Bias Corrections of Climate Simulations Account for Changes of Rank Correlation Over Time? *Journal of Geophysical Research: Atmospheres*, *127*(14). doi: 10.1029/2022JD036562

Vuille, M., Franquist, E., Garreaud, R., Lavado Casimiro, W. S., & Cáceres, B. (2015, 5). Impact of the global warming hiatus on Andean temperature. *Journal of Geophysical Research: Atmospheres*, *120*(9), 3745–3757. doi: 10.1002/2015JD023126

Wan, Z. (2014, 1). New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. *Remote Sensing of Environment*, *140*, 36–45. doi: 10.1016/j.rse.2013.08.027

Wan, Z., Hook, S., & Hulley, G. (1999). *MOD11A1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC.* Retrieved from `https://lpdaac.usgs.gov/products/mod11a1v006/`

Wang, K., & Dickinson, R. E. (2012, 6). A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Reviews of Geophysics*, *50*(2). doi: 10.1029/2011RG000373

Wang, S., Peng, H., Hu, Q., & Jiang, M. (2022, 8). Analysis of runoff generation driving factors based on hydrological model and interpretable machine learning method. *Journal of Hydrology: Regional Studies*, *42*, 101139. doi: 10.1016/j.ejrh.2022.101139

Wen, S., Su, B., Wang, Y., Zhai, J., Sun, H., Chen, Z., ... Jiang, T. (2020, 12). Comprehensive evaluation of hydrological models for climate change impact assessment in the Upper Yangtze River Basin, China. *Climatic Change*, *163*(3), 1207–1226. doi: 10.1007/s10584-020-02929-6

Werner, A. T., & Cannon, A. J. (2016, 4). Hydrologic extremes – an intercomparison of multiple gridded statistical downscaling methods. *Hydrology and Earth System Sciences*, *20*(4), 1483–1508. doi: 10.5194/hess-20-1483-2016

Werth, S., & Güntner, A. (2010, 1). Calibration analysis for water storage variability of the global hydrological model WGHM. *Hydrology and Earth System Sciences*, *14*(1), 59–78. doi: 10.5194/hess-14-59-2010

Werth, S., Güntner, A., Petrovic, S., & Schmidt, R. (2009, 1). Integration of GRACE mass variations into a global hydrological model. *Earth and Planetary Science Letters*, *277*(1-2), 166–173. doi: 10.1016/j.epsl.2008.10.021

Westerberg, I. K., Guerrero, J. L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., ... Xu, C. Y. (2011). Calibration of hydrological models using flow-duration curves. *Hydrology and Earth System Sciences*. doi: 10.5194/hess-15-2205-2011

Widmann, M., Bretherton, C. S., & Salathé, E. P. (2003, 3). Statistical Precipitation Downscaling over the Northwestern United States Using Numerically Simulated Precipitation as a Predictor*. *Journal of Climate*, *16*(5), 799–816. doi: 10.1175/1520-0442(2003)016⟨0799:SPDOTN⟩2.0.CO;2

Wilby, R. L., & Dessai, S. (2010). Robust adaptation to climate change. *Weather*. doi: 10.1002/wea.543

Wood, E. F., Lettenmaier, D. P., & Zartarian, V. G. (1992). A land-surface hydrology parameterization with subgrid variability for general circulation models. *Journal of Geophysical Research*, *97*(D3), 2717. doi: 10.1029/91JD01786

Woods, R. A. (2009, 10). Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks. *Advances in Water Resources*, *32*(10), 1465–1481. doi: 10.1016/j.advwatres.2009.06.011

Wootten, A. M., Dixon, K. W., Adams-Smith, D. J., & McPherson, R. A. (2021, 2). Statistically downscaled precipitation sensitivity to gridded observation data and downscaling technique. *International Journal of Climatology*, *41*(2), 980–1001. doi: 10.1002/joc.6716

Wu, Y., Miao, C., Fan, X., Gou, J., Zhang, Q., & Zheng, H. (2022, 11). Quantifying the Uncertainty Sources of Future Climate Projections and Narrowing Uncertainties With Bias Correction Techniques. *Earth's Future*, *10*(11). doi: 10.1029/2022EF002963

Xavier, A. C. F., Martins, L. L., Rudke, A. P., de Morais, M. V. B., Martins, J. A., & Blain, G. C. (2022, 1). Evaluation of Quantile Delta Mapping as a bias-correction method in maximum rainfall dataset from downscaled models in São Paulo state (Brazil). *International Journal of Climatology*, *42*(1), 175–190. doi: 10.1002/joc.7238

Yan, H., Sun, N., Eldardiry, H., Thurber, T. B., Reed, P. M., Malek, K., ... Rice, J. S. (2023, 5). Large Ensemble Diagnostic Evaluation of Hydrologic Parameter Uncertainty in the Community Land Model Version 5 (CLM5). *Journal of Advances in Modeling Earth Systems*, *15*(5). doi: 10.1029/2022MS003312

Yáñez-Morroni, G., Suárez, F., Muñoz, J. F., & Lagos, M. S. (2023, 10). Hydrological modeling of the Silala River basin. 2. Validation of hydrological fluxes with contemporary data. *WIREs Water*. doi: 10.1002/wat2.1696

Yates, S., Siebert, J., Turkey, D., & Huber-Lee, A. (2005, 7). WEAP21—A demand-, priority-, and preference-driven water planning model: part 1: model characteristics. *Water international*, *30*(4).

Yilmaz, K. K., Gupta, H. V., & Wagener, T. (2008). A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resources Research*. doi: 10.1029/2007WR006716

Yin, W., Hu, L., Zhang, M., Wang, J., & Han, S. (2018, 6). Statistical Downscaling of GRACE-Derived Groundwater Storage Using ET Data in the North China Plain. *Journal of Geophysical Research: Atmospheres*, *123*(11), 5973–5987. doi: 10.1029/2017JD027468

Zaitchik, B. F., Rodell, M., Biasutti, M., & Seneviratne, S. I. (2023, 5). Wetting and drying trends under climate change. *Nature Water*, *1*(6), 502–513. doi: 10.1038/s44221-023-00073-w

Zhang, H., Wang, B., Liu, D. L., Zhang, M., Leslie, L. M., & Yu, Q. (2020, 6). Using an improved SWAT model to simulate hydrological responses to land use change: A case study of a catchment in tropical Australia. *Journal of Hydrology*, *585*, 124822. doi: 10.1016/j.jhydrol.2020.124822

Zhao, R., Zhang, Y. L., Fang, X., Liu, R., & Zhang, Q. S. (1980). The Xinanjiang Model. In *Hydrological forecasting proceedings oxford symposium, iahs 129* (pp. 351–356).

Zink, M., Mai, J., Cuntz, M., & Samaniego, L. (2018, 4). Conditioning a Hydrologic Model Using Patterns of Remotely Sensed Land Surface Temperature. *Water Resources Research*, *54*(4), 2976–2998. doi: 10.1002/2017WR021346

ANNEXES

Annex A

The key role of temporal stratification for GCM bias correction in climate impact assessments

A.1 Selected GCMs

Table A.1 shows the GCMs included in this study.

A.2 Remaining biases after bias correction

Figures A.1 to A.8 display the remaining bias of each climate index disaggregated by BCM. When not shown, the unit of the bias corresponds to the difference between the bias-corrected GCM and the reference ($X_{GCM} - X_{ref}$).

Table A.1: GCMs considered in this study

GCM	Δlat	Δlon	Institution
ACCESS-CM2	1.25	1.88	Australian Research Council Centre of Excellence for Climate Science, Australia.
ACCESS-ESM1-5	1.25	1.88	
BCC-CSM2-MR	1.11	1.13	Beijing Climate Center, China.
CanESM5	2.77	2.81	Canadian Centre for Climate Modelling and Analysis, Canada.
CMCC-ESM2	0.94	1.25	Euro-Mediterranean Centre on Climate Change Coupled Climate Model, Italy.
CNRM-CM6-1-HR	0.50	0.50	Centre National de Recherches Météorologiques (CNRM), France.
CNRM-CM6-1	1.40	1.40	
CNRM-ESM2-1	1.40	1.41	
E3SM-1-0	1.00	1.00	Lawrence Livermore National Laboratory, USA.
EC-Earth3-CC	0.70	0.70	EC-Earth Consortium, Europe.
EC-Earth3-Veg-LR	1.12	1.13	
EC-Earth3-Veg	0.70	0.70	
EC-Earth3	0.70	0.70	
FGOALS-g3	2.18	2.00	Chinese Academy of Sciences Flexible Global Ocean-Atmosphere-Land System Model, China.
GFDL-CM4	1.00	1.25	Geophysical Fluid Dynamics Laboratory, USA.
GFDL-ESM4	1.00	1.25	
INM-CM4-8	1.50	2.00	Institute for Numerical Mathematics, Russia.
INM-CM5-0	1.50	2.00	
IPSL-CM6A-LR	1.27	2.50	Institute Pierre Simon Laplace (IPSL), France.
KACE-1-0-G	1.25	1.88	National Institute of Meteorological Sciences (NIMS) and Korea Meteorological Administration (KMA), South Korea.
KIOST-ESM	1.88	1.88	Korea Institute of Ocean Science and Technology Earth System Model and Its Simulation Characteristics, South Korea.
MIROC-ES2L	2.79	2.81	Japan Agency for Marina-Earth Science and Technology (JAMSTEC), Japan.
MIROC6	1.39	1.41	
MPI-ESM1-2-HR	0.93	0.94	Max Planck Institute for Meteorology (MPI-M), Germany.
MPI-ESM1-2-LR	1.87	1.88	
MRI-ESM2-0	1.11	1.13	Meteorological Research Institute, Japan.
NESM3	1.85	1.88	Nanjing University of Information Science and Technology Earth System Model, China.
NorESM2-MM	0.94	1.25	NorESM Climate modeling Consortium, Oslo, Norway.
TaiESM1	0.94	1.25	Research Center for Environmental Changes, Academia Sinica, Nankang, Taipei, Taiwan.

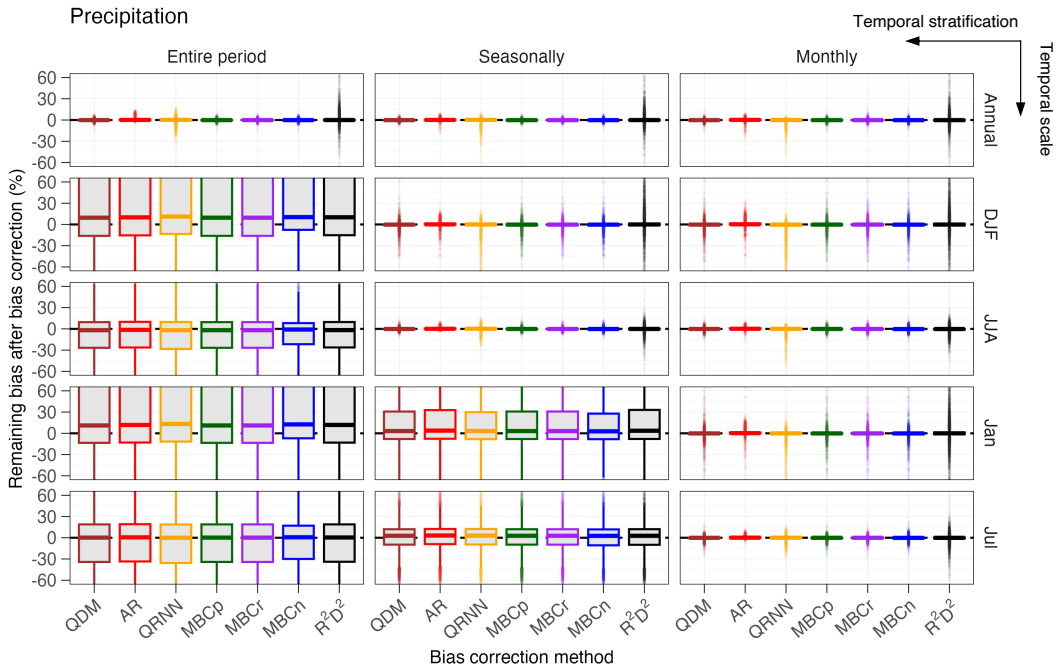

Figure A.1: Remaining biases after bias correction for precipitation at different time scales. Different columns represent different temporal stratifications. Values are disaggregated by bias correction method. The dispersion within each boxplot derives from 29 GCMs and 1000 grid cells.

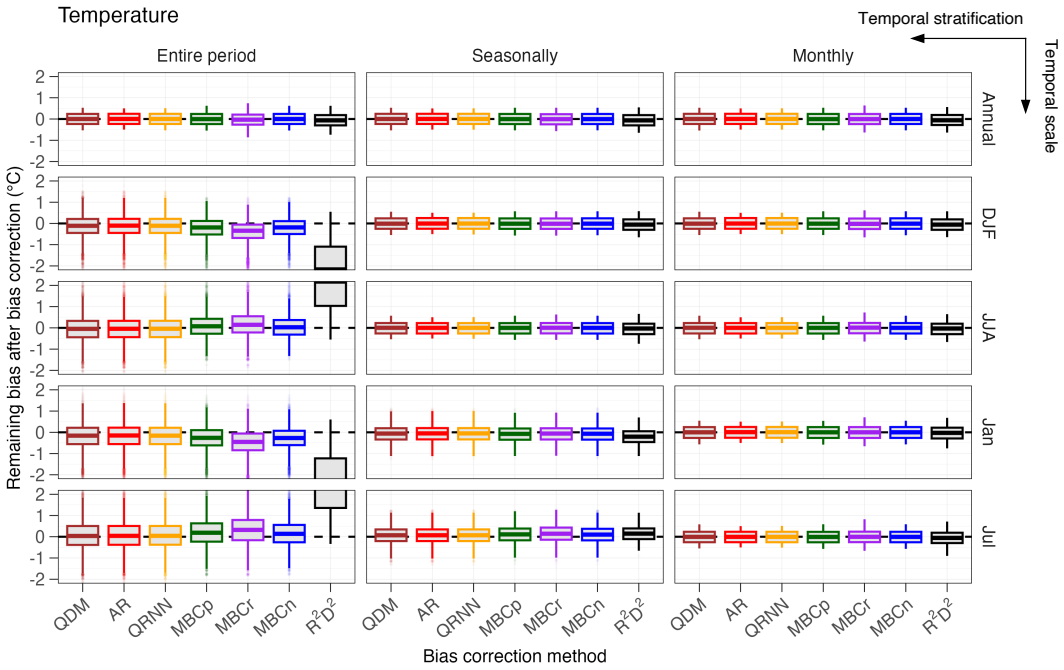

Figure A.2: Same as in Fig. A.1, but for temperature.

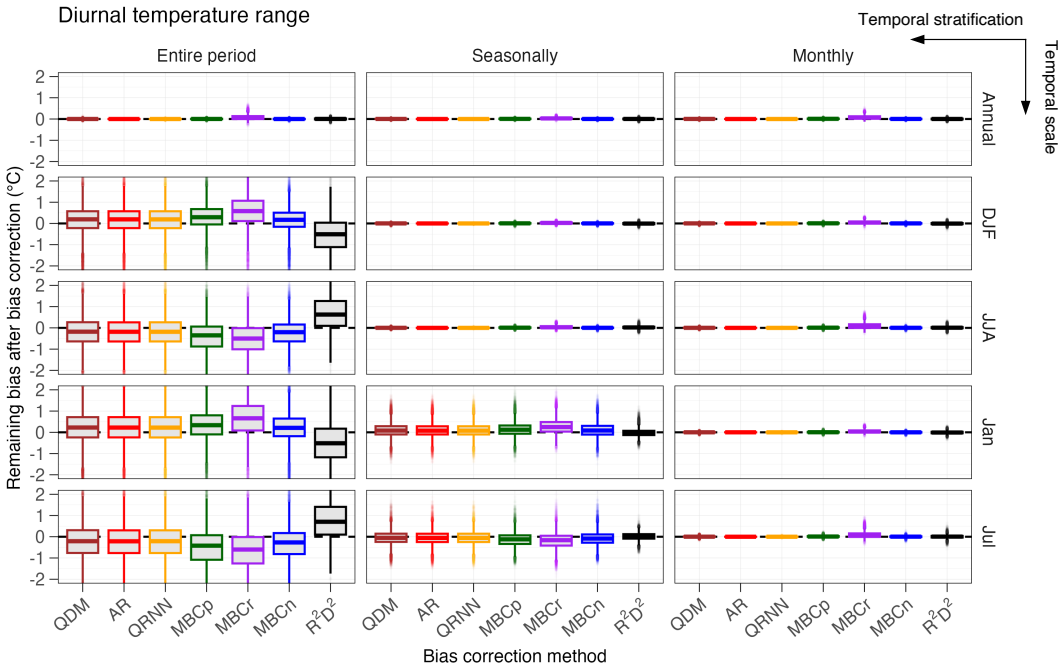

Figure A.3: Same as in Fig. A.1, but for diurnal temperature range.

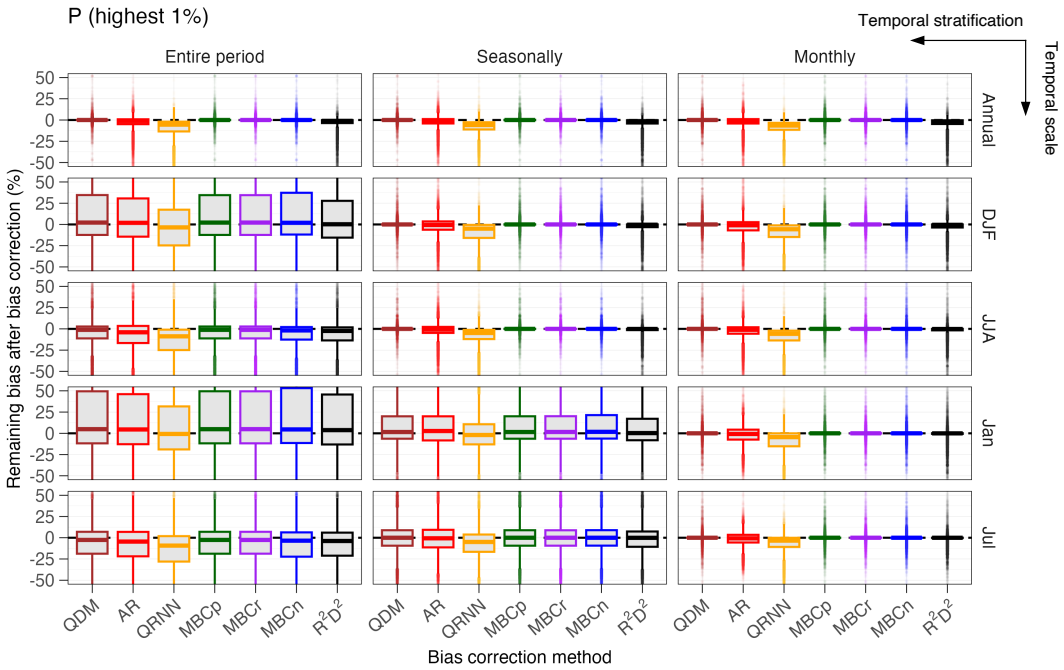

Figure A.4: Same as in Fig. A.1, but for highest 1% daily precipitation.

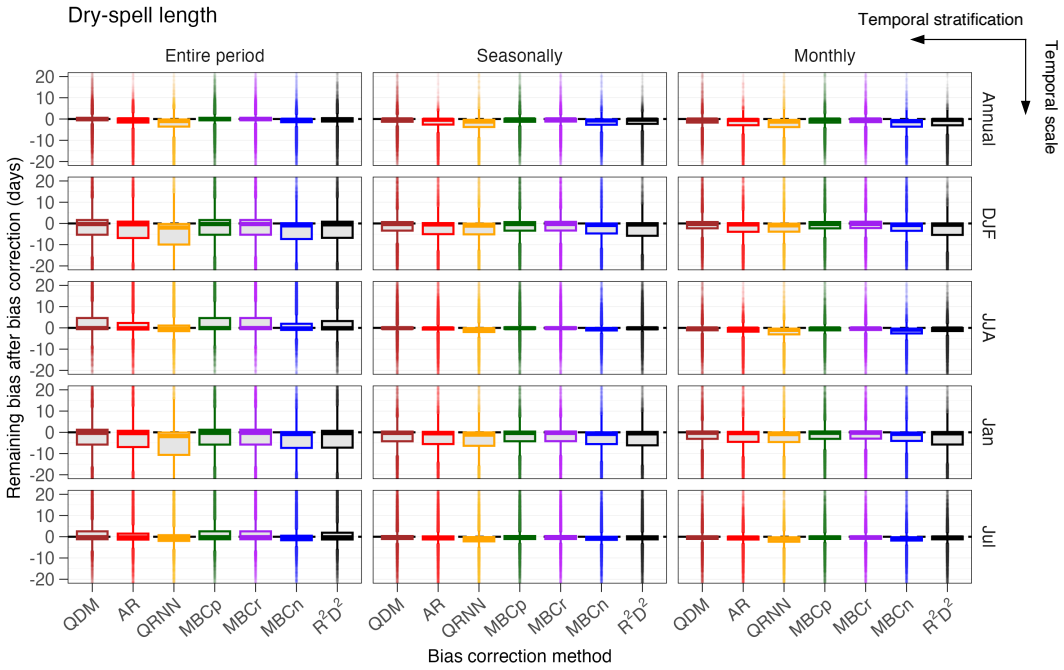

Figure A.5: Same as in Fig. A.1, but for dry-spell length.

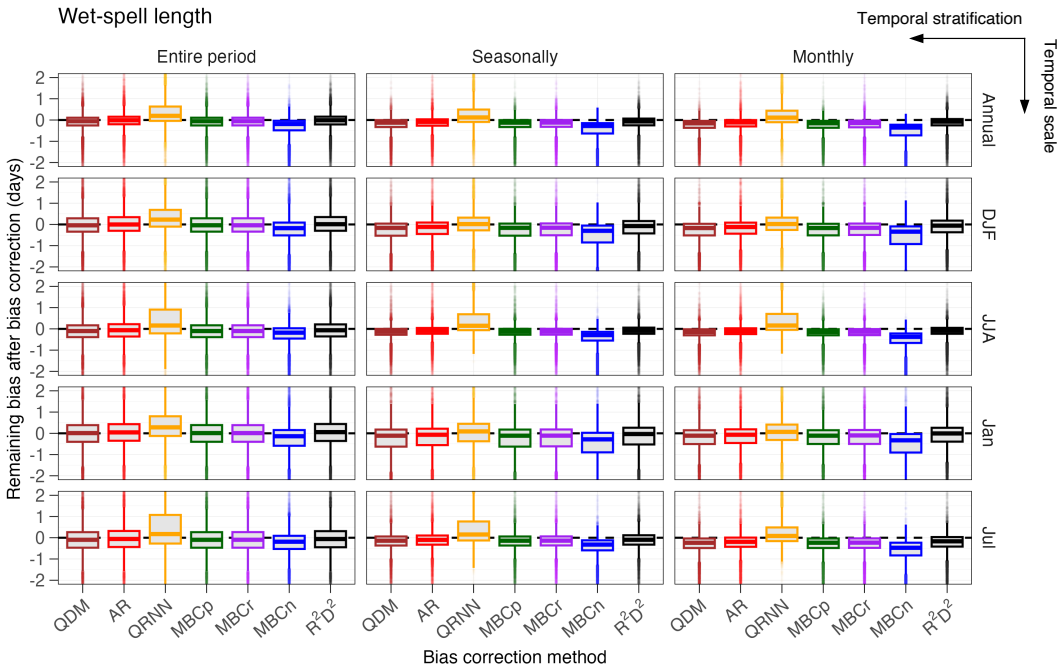

Figure A.6: Same as in Fig. A.1, but for wet-spell length.

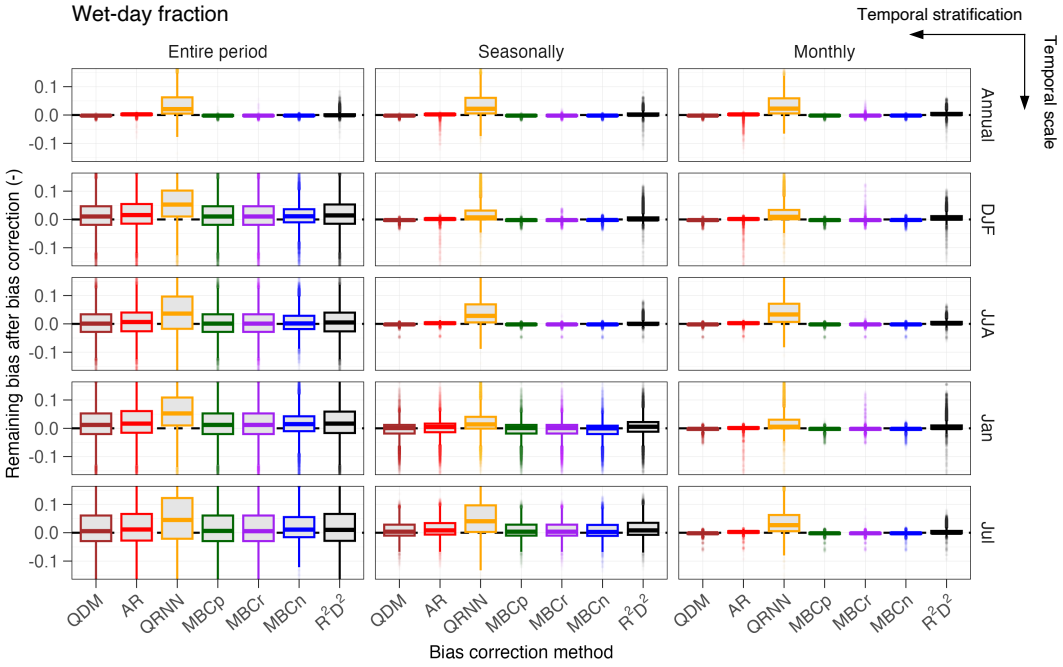

Figure A.7: Same as in Fig. A.1, but for wet-day fraction.

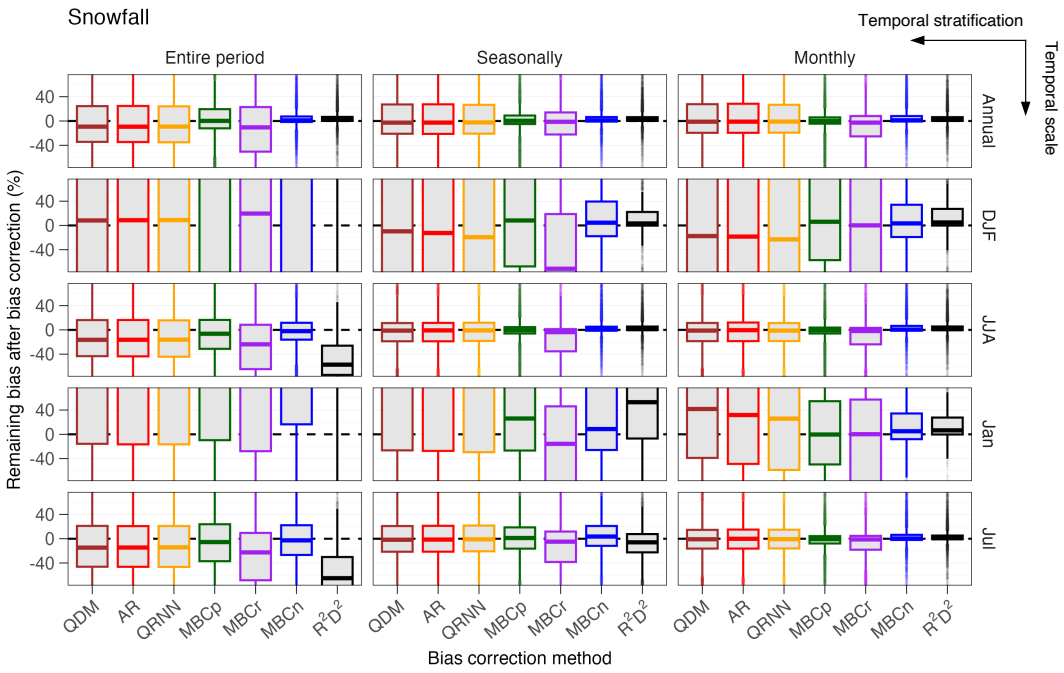

Figure A.8: Same as in Fig. A.1, but for snowfall.

A.3 Precipitation dismissed by bias-correcting the wet-day fraction

Figure A.9 shows the fraction of raw GCM precipitation dismissed (pr$< \tau$; median among models) when the wet-day fraction is bias-correct firstly.

A.4 Relative importance of the choice of Bias Correction Method and Temporal Stratification

Figure A.10 shows the relative importance of the bias correction method (BCM) and the temporal stratification to explain the variance of errors in bias-corrected climate indices during the historical period for Continental Chile, based on ANOVA analysis. The Total Variance (TV) is formulated as $TV = BCM + TS + Residuals$. Results from the ANOVA analysis (BCM/TV, TS/TV, and $Residuals/TV$) are computed for each grid cell and GCM and subsequently averaged for continental Chile.

A.5 Scaling factor example

We illustrate the effects of the temporal stratification by applying the linear scaling method (LSM) (Maraun et al., 2010) for one grid cell-GCM combination. Figure A.11a shows monthly precipitation averages from raw GCM outputs, whereas Figure A.11b-d shows the bias-corrected GCM values for three different temporal stratifications. Monthly values were obtained from the daily corrected time series.
Note that when the entire period is used to bias-correct the GCM, only one factor is ap-

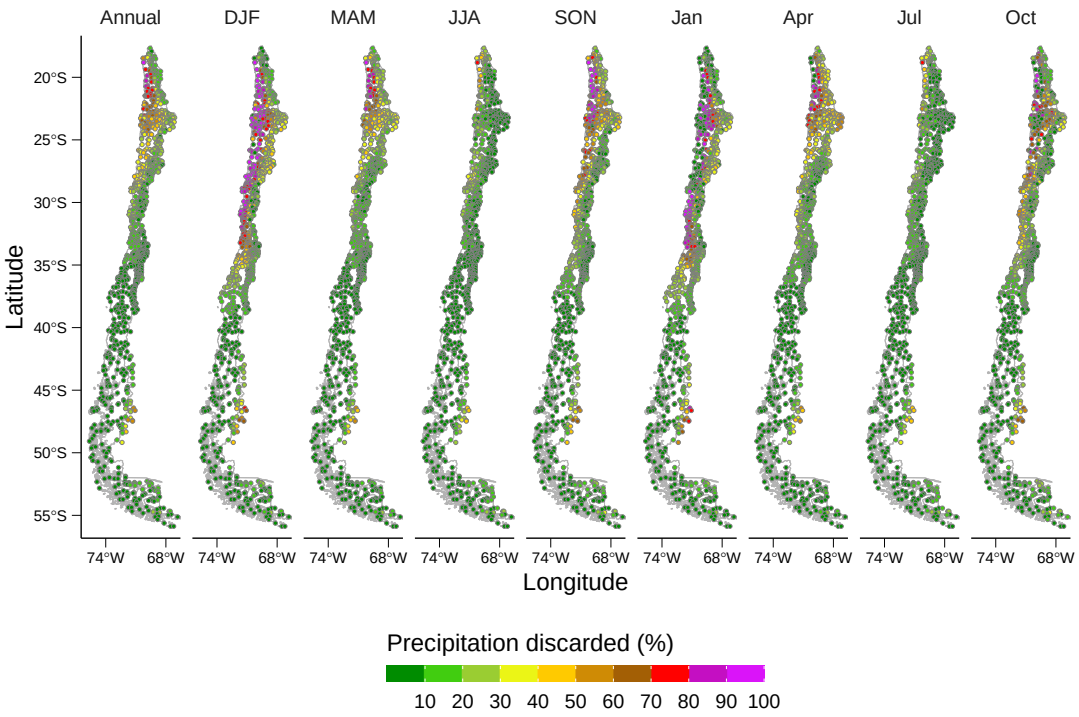

Figure A.9: Amount of precipitation dismissed to adjust the number of rainy days. Annual corresponds to the entire period temporal stratification, DJF, MAM, JJA and SON to a seasonal temporal stratification, while each month corresponds to a monthly temporal stratification.

plied. In the grid cell analyzed, the reference annual precipitation is 4,371 mm/yr, which is below the historical raw GCM amount for the same period (5,020 mm/yr). Hence, the raw GCM precipitation time series is multiplied by the factor $f = 4731/5020 = 0.87$, which removes the annual SDBC bias; nevertheless, monthly SDBC-biases persist (see differences between black and blue lines in Figure A.11b). When the LSM is applied seasonally, four factors are used to multiply the raw GCM time series. For example, daily values from March, April, and May are bias-corrected by the seasonal factor obtained from the reference (1134 mm/season) and the raw GCM (1498 mm/season) precipitation amounts. In this case, the factor used to bias-correct daily precipitation from March, April, and May is $f_{MAM} = 1134/1498 = 0.76$. Similarly, if the LSM is applied monthly, daily precipitation amounts from March are bias-corrected using the reference (374 mm/-month) and raw GCM (498 mm/month), which yields a factor $f = 374/498 = 0.75$. For the monthly TS, the black and blue lines are the same. Note that the projected maximum monthly precipitation is October for the three TS, which is the same as the raw GCM projection. However, the projected minimum monthly precipitation is September, March, and March for the entire period, season, and monthly application of the LSM, respectively.

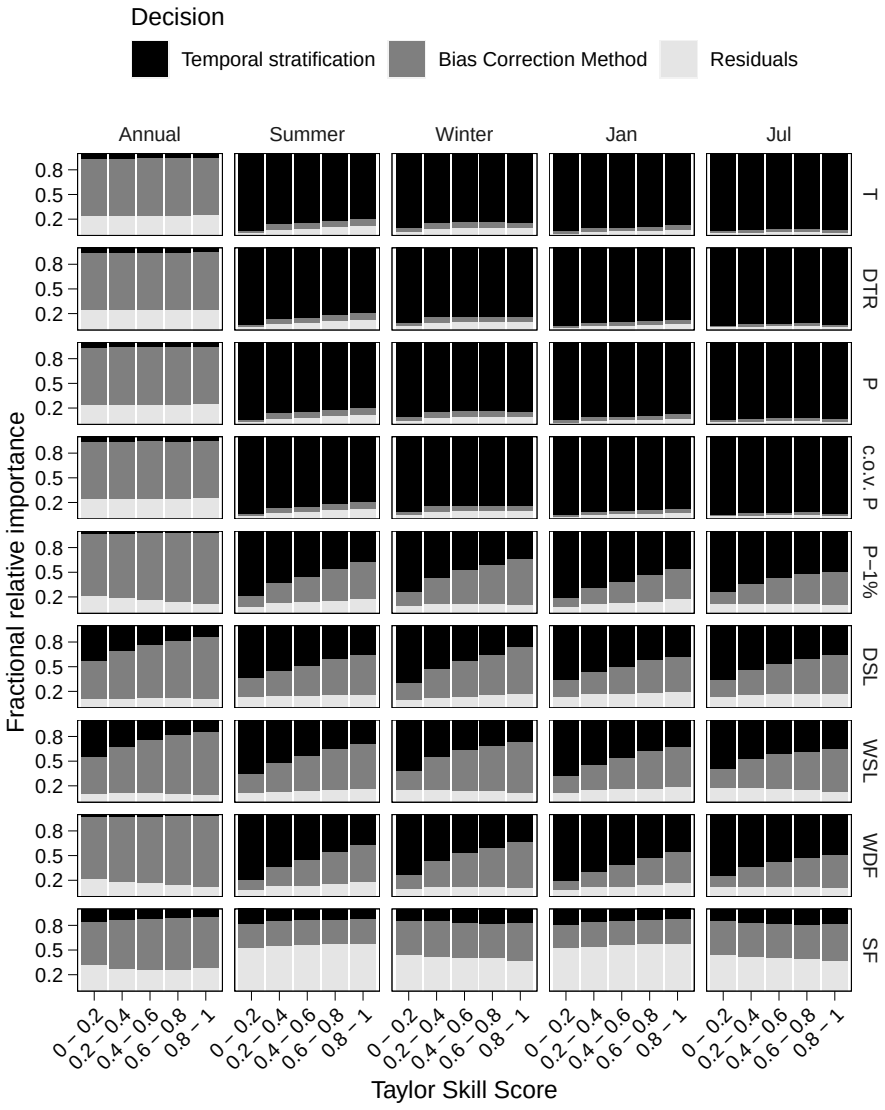

Figure A.10: Relative importance (averaged across all grid cells and GCMs) of the bias correction method and the temporal stratification to explain the dispersion of biases with respect to the reference dataset at the annual, seasonal (DJF and JJA), and monthly (January and July) time scales during the historical period (1980-2014). Results are stratified according to the historical raw GCM performance (measured by the TSS; x-axis). Biases are computed after applying the BCMs, and results are displayed for temperature (T), diurnal temperature range (DTR), precipitation (P), coefficient of variation of inter-annual precipitation (c.o.v. P), highest 1% daily precipitation amount (P-1%), dry spell length (DSL), wet spell length (WSL) and snowfall fraction (SF).

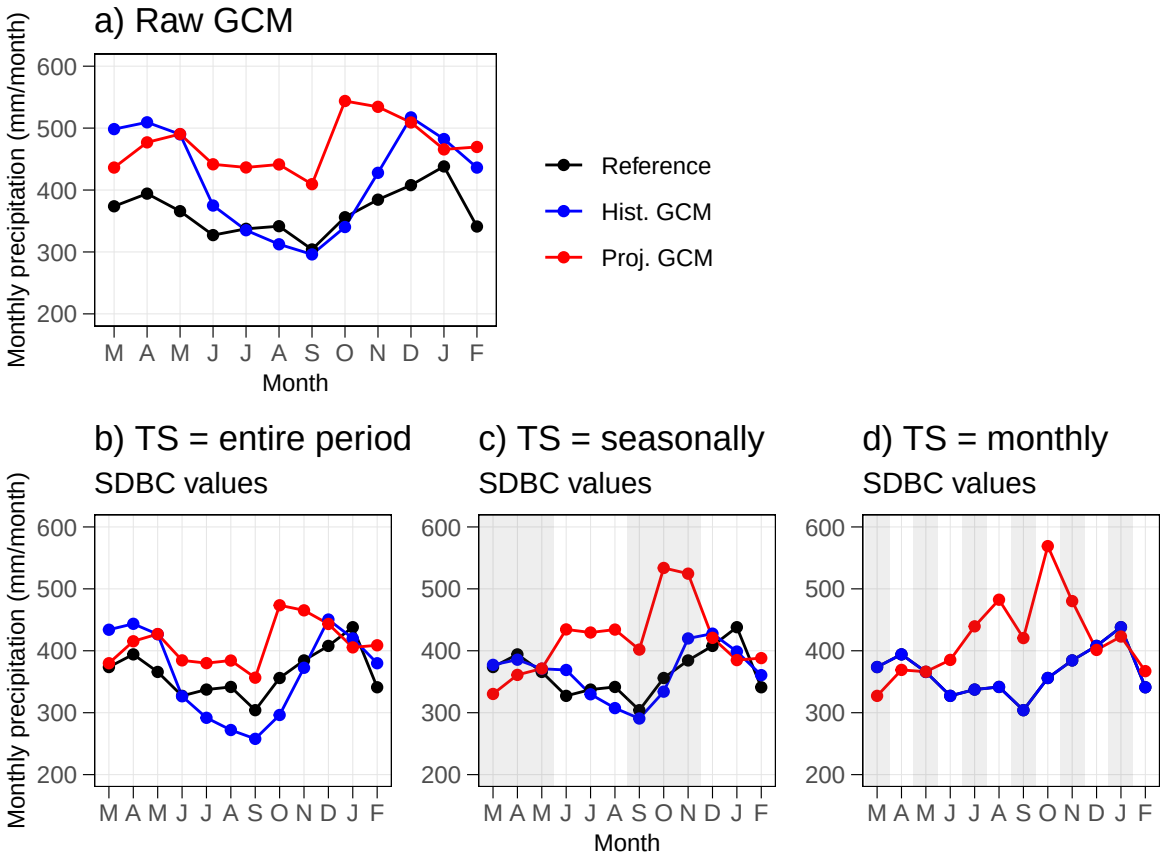

Figure A.11: Illustration of the linear scaling method, applied to one grid cell-GCM combination, and its effects on the SDBC-biases and projections. (a) Reference (observational) and raw GCM seasonality during the period 1980-2014 (black and blue lines). The projected raw seasonality is also shown in red (2065-2099). (b), (c) and (d) show the bias-corrected precipitation amounts using the entire period, seasons, and months, respectively, for temporal stratification. The reference value is shown in all panels for completeness, and the shaded areas represent the temporal stratification.

Annex B

Principal component analysis to spatially distribute model parameters

To derive *a priori* parameters, we conduct a Principal Component Analysis between (i) elevation, (ii) slope, (iii) clay and sand contents, and (iv) bulk density. Figure B.1a illustrates the fraction of the total variance explained by each component, while Figure B.1b contrasts the PC_1 against the attributes.

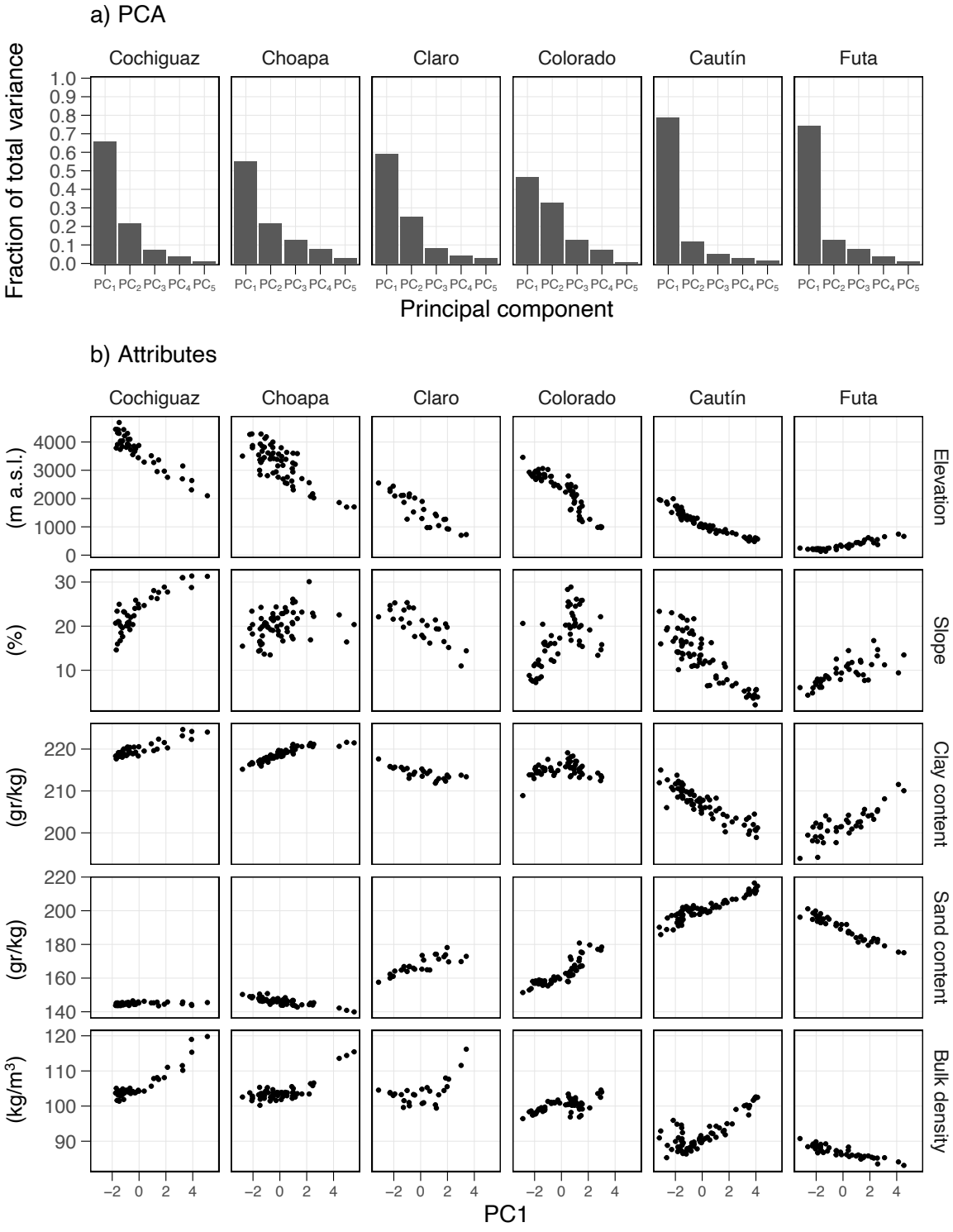

Figure B.1: Principal component analysis to derive *a priori* parameters. a) Fraction of total variance explained by each component. b) Relationship between PC_1 and the attributes. Each dot represents a grid cell (0.05° horizontal resolution).

Annex C

Q-ET seasonality index

This section explores why the choice of the objective function — whether to include Q or Q+ET – is, in general, the most critical decision in the wettest part of the domain explaining the dispersion of simulated hydrological changes. Since "good" Q simulation does not translate into "good" ET simulations, the seasonality of the water balance at the catchment scale could be poorly simulated, impacting hydrological projections under different conditions. In this regard, reasonably simulating Q and ET average seasonalities is, arguably, a minimum requirement for hydrological simulations.
In the wettest part of the domain, most of the basins considered in this study are rainfall-dominated, i.e., most of the runoff occurs during winter, while the maximum ET values are in summer. This requires the model to partition the precipitation into P and ET for two different periods. To test if this is the case for the basins of this work (i.e., Q and ET maximum values in different seasons or months), the precipitation seasonality index (δ_P^* Woods, 2009) is applied to Q and ET seasonality (δ_{Q-ET}^*). Figure C.1 illustrates the Q-ET seasonality index for two cases. Similar to δ_P^*, positive values indicate Q and ET seasonalities coupled (Figure C.1a), while negative values indicate decoupled and ET seasonalities (Figure C.1b).

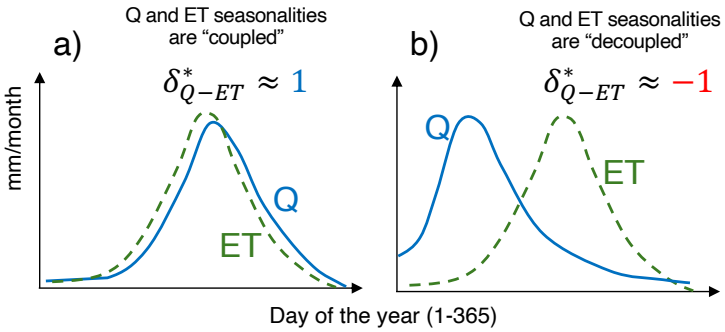

Figure C.1: Scheme with the Q-ET seasonality index for two basins with different streamflow average seasonalities.

To verify if the inclusion of Q or Q+ET in the objective function is more relevant in basins with negative δ_{Q-ET}^* values, Figure C.2 displays the most relevant decision (choice of GCM or inclusion of Q or Q+ET in the OF) explaining the dispersion of projected hydrological changes. Overall, most of the basins where including Q or Q+ET in the OF is the most relevant decision (compared to the choice of GCM) have negative Q-ET seasonality index values (i.e., decoupled Q and ET seasonalities).

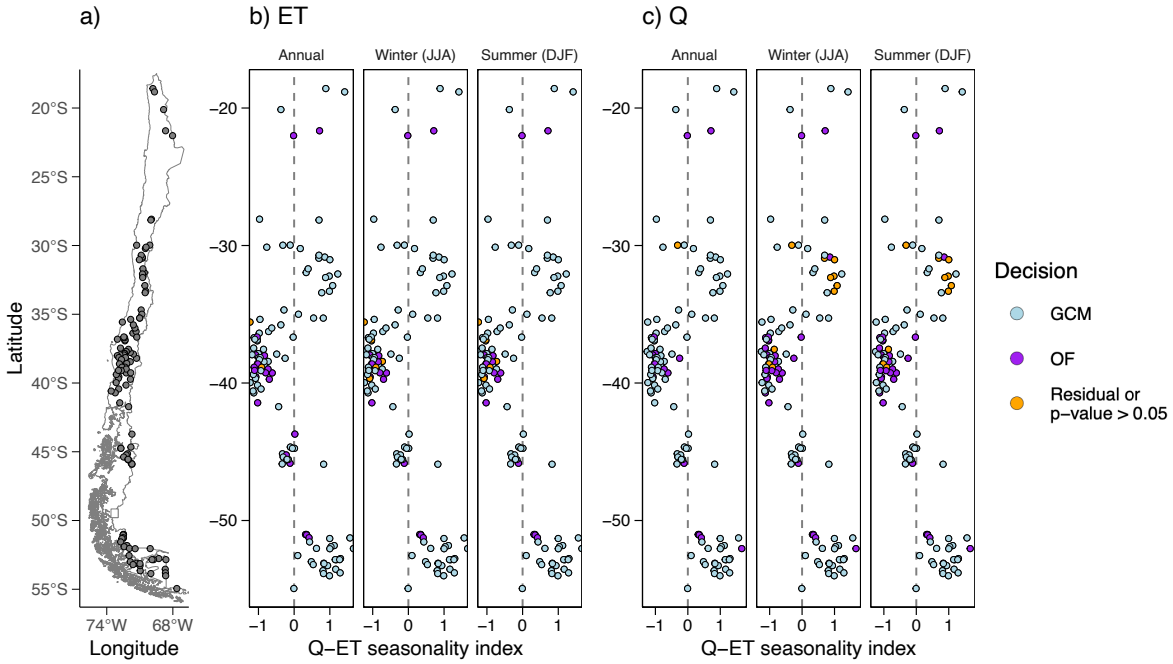

Figure C.2: Most relevant decision (color) explaining the dispersion of projected changes in ET (panel b) and ET (panel c) at the annual and seasonal time scales. The Q-ET seasonality index (x-axis) is computed as in Woods (2009) for observed runoff and MOD16 ET annual cycles. To reduce the effect of biases in MOD16 ET values, both Q and ET are normalized as $X'=(X-X_{min})/(X_{max} - X_{min})$ before computing the Q-ET seasonality index.