



**STREAMFLOW SIMULATION IN DATA-SCARCE REGIONS USING REMOTE
SENSING DATA IN COMBINATION WITH GROUND-BASED MEASUREMENTS**

Oscar M. Baez-Villanueva

Page left intentionally blank.

Streamflow simulation in data-scarce regions using remote sensing data in combination with ground-based measurements

by

Oscar Manuel Baez Villanueva

A doctorate dissertation submitted to the Department of Spatial Planning at the TU Dortmund University in partial fulfilment of the requirements for the Degree of Doctor of Engineering (Dr.-Ing.)

Examination Board:

Chairman:	Univ.-Prof. Dr.-Ing. Stefan Siedentop
Supervisor and Reviewer:	Univ.-Prof. Dr. habil. Nguyen Xuan Thinh
Supervisor and Reviewer:	Prof. Dr. Lars Ribbe

Technische Universität Dortmund

Dortmund, Germany

May 2022

© 2022 — Oscar Manuel Baez Villanueva
All rights reserved.

Declaration:

I declare that this cumulative dissertation has been composed by myself and that the work has not been submitted for any other degree or professional qualification. My contribution and those of the other authors to this work have been explicitly indicated below. I confirm that appropriate credit has been given within this dissertation where reference has been made to the work of others. The work presented in Chapter 3 is composed by a three-step procedure: *i*) Evaluation; *ii*) Merging; and *iii*) Regionalisation. Each step corresponds to a scientific article previously published in peer-reviewed scientific journals.

1. The Evaluation step was published in *Atmospheric Research* as "*Temporal and spatial evaluation of satellite rainfall estimates over different regions in Latin-America*" by Oscar M. Baez-Villanueva (author of the declaration), Mauricio Zambrano-Bigiarini, Lars Ribbe (supervisor), Alexandra Nauditt, Juan Diego Giraldo-Osorio, Nguyen Xuan Thinh (supervisor). OMBV led the investigation, conducted the analysis, and wrote the original draft; OMBV and MZB conceived and developed the methodology, and supervised the project; JDGO provided methodological feedback; and all authors reviewed the manuscript (DOI: 10.1016/j.atmosres.2018.05.011).

2. The Merging step was published in *Remote Sensing of Environment* as "*RF-MEP: A novel Random Forest method for merging gridded precipitation products and ground-based measurements*" by Oscar M. Baez-Villanueva (author of the declaration), Mauricio Zambrano-Bigiarini, Hylke E. Beck, Ian McNamara, Lars Ribbe (supervisor), Alexandra Nauditt, Christian Birkel, Koen Verbist, Juan Diego Giraldo-Osorio, and Nguyen Xuan Thinh (supervisor). OMBV led the investigation, conducted the analysis, and wrote the original draft; OMBV and MZB conceived and developed the methodology, and supervised the project; HEB, IM, AN, CB, and KV provided methodological feedback; and all authors reviewed the manuscript (DOI: 10.1016/j.rse.2019.111606).

3. *The Regionalisation step was published in Hydrology and Earth System Sciences as "On the selection of precipitation products for the regionalisation of hydrological model parameters" by Oscar M. Baez-Villanueva (author of the declaration), Mauricio Zambrano-Bigiarini, Pablo A. Mendoza, Ian McNamara, Hylke E. Beck, Joschka Thurner, Alexandra Nauditt, Lars Ribbe (supervisor), and Nguyen Xuan Thinh (supervisor). OMBV led the investigation, conducted the analysis, and wrote the original draft; OMBV, MZB, and PAM conceived and developed the methodology, and supervised the project; HEB and IM provided methodological feedback; JT supported in the development of the algorithms; all authors reviewed the manuscript (DOI: 10.5194/hess-25-5805-2021).*

Streamflow simulation in data-scarce regions using remote sensing data in combination with ground-based measurements

Abstract

Global water resources are currently under unprecedented stress, which is projected to increase due to the influence of multiple factors. Therefore, changes in governance are urgently required to improve water management and water use efficiency while maintaining the health of river systems and their water quantity and quality. Data is crucial in this process; however, most rivers in the world remain ungauged, and in data-scarce regions, the hydrometric and hydrometeorological networks of stations have been decreasing during the last decades. This hinders the implementation of proactive water management approaches that strive towards informed-based decision-making.

This cumulative thesis shows how open access global precipitation products can be evaluated, corrected, and used to predict streamflow at the daily temporal scale in data-scarce regions in combination with ground-based measurements by following a three-step approach: *i*) performance evaluation of different precipitation products over regions with different climates and at multiple temporal scales; *ii*) development of a novel merging method to improve the representation of precipitation at the daily scale; and *iii*) assessment of the ability of the novel merged product altogether with state-of-the-art precipitation products to predict daily streamflow over ungauged catchments through the implementation of regionalisation approaches.

This thesis showed that the precipitation products perform differently depending on the temporal scale, elevation, and climate; and that these products still have errors in detecting particular precipitation events. These insights served as a basis to develop a novel merging procedure named RF-MEP, which combines data from precipitation products, ground-based measurements, and topographical features to improve the characterisation of precipitation. RF-MEP proved to be a powerful method as the precipitation errors at different temporal scales were substantially reduced, outperforming state-of-the-art precipitation products and merging procedures. The

precipitation product derived with RF-MEP has been included in a Chilean precipitation monitor platform from the Center for Climate and Resilience Research ([Mawün](#)) and users can apply this method in a friendly manner using the R package [RFmerge](#). This merged product altogether with three state-of-the-art precipitation products was used to implement three regionalisation approaches by calibrating an HBV-like hydrological model over 100 near-natural catchments in Chile. The results showed that although these methods yielded relatively good performances, the precipitation products corrected with daily gauge observations did not necessarily yield the best hydrological and regionalisation performance.

Additionally, the hydrological regime of the catchments influenced the performance of the evaluated regionalisation techniques, with the pluvio-nival and rain-dominated catchments yielding the best and worst performance, respectively. This cumulative dissertation shows that precipitation datasets can help to strive towards informed-based decision-making in data-scarce regions. However, these regions often lack the infrastructure and human capacity to use this type of information efficiently. Therefore, an informed-based decision-making process requires institutional transitions and changes that help address water resources management's present and future challenges. In this sense, there is a need to move towards data-driven water resources management by implementing strategic approaches that systematically build the capacities and infrastructure of such regions.

Keywords: data scarcity, hydrological modelling, machine learning, merging procedures, PUB, precipitation, precipitation products, Random Forest, regionalisation, remote sensing, RF-MEP

Contents

Abstract	iv
List of Figures	x
List of Tables	xii
Acknowledgments	xv
Dedication	xvi
1 Introduction	1
1.1 Background	1
1.2 Information to improve water resources management in data-scarce environments	3
1.3 The role of streamflow and precipitation	6
1.3.1 Precipitation products and their uncertainties	7
1.3.2 Streamflow prediction in ungauged catchments	8
1.4 Thesis objectives and outline	10
2 Study Areas	13
2.1 Selected catchments for the evaluation of precipitation products . . .	13
2.1.1 The Magdalena	14
2.1.2 The Paraiba do Sul	15

CONTENTS

2.1.3	The Imperial	16
2.2	Continental Chile	18
3	Data and Methods	21
3.1	Evaluation	21
3.2	Merging	27
3.2.1	RF-MEP	27
3.2.2	Application of RF-MEP to continental Chile	29
3.3	Regionalisation	33
3.3.1	Meteorological forcings	33
3.3.2	Hydrological model	34
3.3.3	Independent catchment calibration and verification	35
3.3.4	Regionalisation techniques	36
3.3.5	Influence of nested catchments	39
3.3.6	Influence of donor catchments for feature similarity	40
4	Results	41
4.1	Evaluation	41
4.1.1	Performance of the evaluated precipitation products at different temporal scales	41
4.1.2	Does the upscaling process affect the performance of the products?	44
4.1.3	Lessons learned from the evaluation of precipitation products	46
4.2	Merging	48
4.2.1	Spatio-temporal performance of the merged products	48
4.2.2	Correction of mismatches from the original precipitation products	51
4.2.3	Impact of density, spatial resolution, and limitations of RF-MEP	53
4.2.4	Comparison with other merging methods	54
4.3	Regionalisation	55

CONTENTS

4.3.1	Performance of precipitation products	55
4.3.2	Evaluation of regionalisation techniques	59
4.3.3	Impact of nested catchments	62
4.3.4	Impact of donors in feature similarity	63
5	Discussion	65
5.1	Evaluation of precipitation products and its influence on water re- sources management	65
5.2	Implications of merging precipitation datasets	69
5.3	On the selection of precipitation products to predict streamflow in ungauged catchments	71
5.4	Future research and recommendations	73
6	Conclusion	77
	Appendices	81
A	Evaluation step	83
B	Merging step	101
B.1	Supplement material	121
C	Regionalisation step	125
C.1	Supplement material	159
	References	175

CONTENTS

List of Figures

1.1	Conceptualisation of the three different components of the methodology followed in this cumulative dissertation: <i>i</i>) the evaluation step (Baez-Villanueva et al. 2018); <i>ii</i>) the merging step (Baez-Villanueva et al. 2020); and <i>iii</i>) the regionalisation step (Baez-Villanueva et al. 2021).	12
2.1	The three selected study areas; <i>a</i>) the Magdalena, located in Colombia; <i>b</i>) the Paraiba do Sul, located in Brazil; and <i>c</i>) the Imperial, located in Chile.	14
2.2	Köppen-Geiger climate classification over the three study areas according to Beck et al. (2018); the Imperial (top-left); the Paraiba do Sul (top-right); and the Magdalena (bottom-left).	15
2.3	Mean annual precipitation of the three study regions: <i>a</i>) the Imperial (during 2001–2015); <i>b</i>) the Paraiba do Sul (during 2001–2015); and <i>c</i>) the Magdalena (during 2001–2014). The blue line represents the mean values for each catchment and the red contour represents one standard deviation above and below the mean value.	17
2.4	Study area: <i>a</i>) elevation (SRTMv4.1; Jarvis et al. 2008); <i>b</i>) land cover classification (Zhao et al. 2016); and <i>c</i>) Köppen-Geiger climate classification (Beck et al. 2018) over the five major macroclimatic zones described in Zambrano-Bigiarini et al. (2017).	19
3.1	Flowchart summarising the three-step procedure followed in the methodology presented in this cumulative dissertation: <i>i</i>) evaluation; <i>ii</i>) merging; and <i>iii</i>) regionalisation. For a more detailed explanation please see Baez-Villanueva et al. (2018, 2020, 2021).	22

LIST OF FIGURES

List of Tables

1.1	Summary of available open-access gridded products.	4
1.2	Summary of water resources management decisions that can be made with hydrometeorological data. This table has been adapted from Sheffield et al. (2018).	5
3.1	Evaluated precipitation products.	23
3.2	Classification of precipitation events based on daily intensity (i) according to Zambrano-Bigiarini et al. (2017).	26
3.3	Precipitation products used in the case study, where S stands for satellite, R for reanalysis, and G for gauge data.	29
3.4	Gridded precipitation products used in this study.	33
3.5	Summary of the TUWmodel parameters considered for calibration, following the conceptualisation presented in Széles et al. (2020).	35
3.6	Selected climatic and physiographic characteristics to quantify feature similarity between catchments. All variables related to precipitation were computed using the corresponding precipitation product used as an input to the TUWmodel for 1990–2018.	38
4.1	Results of the evaluation of performance using the KGE' and the RMSE over the study areas.	42
4.2	Results of the evaluation of performance over the study areas using the categorical indices.	45
4.3	Results of the assessment of the temporal, spatial, and categorical performance of the merged products.	50

LIST OF TABLES

4.4	Calibration, verification, and regionalisation results over 100 near-natural catchments.	57
4.5	Summary of the results of the overall performance of the regionalisation techniques and the impact of hydrological regimes.	61

Acknowledgments

Firstly, I would like to express my sincere gratitude and appreciation to Prof. Dr Lars Ribbe and Prof. Dr Nguyen Xuan Think. Thank you for your guidance and support. I want to express special thanks to my unofficial supervisor Mauricio Zambrano-Bigiarini, who taught me a lot about hydrology, remote sensing, and life. Thank you for your unconditional help, mentorship, and friendship. Thanks to Juan Diego (el parce) Giraldo-Osorio, who helped me from the beginning of my PhD scientifically and emotionally. Thanks to Alex Nauditt for all the fruitful discussions and help. Also, I want to thank Hylke E. Beck for his openness, feedback, support, and humility.

I would also like to thank my friends for their support and encouragement during this journey. Thank you, Sergio, Edder, and Panguito, for the unforgettable moments, incredible food, and support in my difficult moments. To Ian and Lisa for the lovely memories, the city safaris, and the real ones. To Tere, Zryab, Santiago, Alicia, Alex K., Nora, Pedro, and Justyna for the incredible moments at the office. To Joschka for his unconditional support, beers, and Mexican dinners. To Aric for the nice moments and the misunderstood jokes. To Nathaly, my sister from Colombia. To Pia for her outright support along the way. To Alex and Lars for the fantastic parties on their terrace. Finally, to the "pachangos" (Ale, Andy, Josue, Marce, Maye, Nacho, Rafa, Rob, Rodrigo y Suko) for always being there.

Last but not least, I want to express my eternal gratitude to my family (Yolanda, Oscar, Diego, Abue, and Chayo). This is our achievement because I couldn't have made it without your love, guidance, and support.

"Gratitude is the memory of the heart."

Jean Baptiste Massieu

To my family and friends.

Chapter 1

Introduction

1.1 Background

Global water resources are expected to come under increased stress due to the influence of multiple factors, such as *i*) overexploitation of surface and groundwater (Rockström et al. 2012; Richey et al. 2015; Van Ginkel et al. 2018); *ii*) rising demands of agricultural production and energy (Vörösmarty et al. 2000; Brauman et al. 2016; Qin et al. 2019); *iii*) climate change impacts (Konapala et al. 2020; Woolway et al. 2020), the projected increase in the frequency and severity of droughts, heatwaves, and floods in some regions (Vörösmarty et al. 2000; Whitfield 2012; Madsen et al. 2014; Prudhomme et al. 2014; Blöschl et al. 2015; Mazdiyasn and AghaKouchak 2015; Roudier et al. 2016; Van Vliet et al. 2016); and *iv*) population growth (Schlosser et al. 2014; Wada et al. 2014; Kiguchi et al. 2015), among others. Irrigated agriculture is the largest water consumer at the global scale, accounting for around 85–90% of water consumption (Vörösmarty et al. 2000; Oki and Kanae 2006; Zhang et al. 2016; Brauman et al. 2016), followed by industrial and domestic water use, which has quadrupled in the last 60 years (Flörke et al. 2018), and evaporative losses from lakes and reservoirs (Brauman et al. 2016; Qin et al. 2019).

Recent studies indicate that the levels of water extraction will continue to degrade riverine ecosystems, reduce river’s streamflow, hindering the possibility to strive towards sustainable water management (Wan et al. 2017; Sabater et al. 2018; Bond et al. 2019). Therefore, changes in governance are urgently required to maintain the health of river systems, streamflow levels, and water quality (Grafton et al. 2013), which highlights the increased importance of water resources management and water use efficiency.

CHAPTER 1. INTRODUCTION

There is a need to move from sectorial-based management of water resources and strive for an integrated approach that involves all sectors in an informed-based decision-making process. Different efforts have been started in this direction, such as the 2030 Agenda for Sustainable development (UN 2015), which aims to ensure access to water and sanitation for the world’s population (Goal 6) and take urgent action to combat climate change and its impacts (Goal 13). Similarly, there are a plethora of national and multinational water management plans that strive to increase access to drinking water and sanitation, and protect freshwater resources. Some examples of these national and international plans are the EU Framework Directive in Europe (Kallis and Butler 2001); the Clean Water Act in USA (Houck 2002); the Report of Three Decades of Water Policy in Australia (Doolan et al. 2016); the Urban Sanitation Policy in India (MoUD 2008); and the Report on the National Water Reserves Program in Mexico (Barrios-Ordóñez et al. 2015). The implementation of these plans and the involvement of key stakeholders is crucial as it is projected that the world’s water demand will increase by 40% in 2030 (Connor et al. 2017). Therefore, the execution of robust water resources management solutions that tackle these complex water challenges requires the incorporation of cutting-edge information and knowledge.

However, there are several challenges related to the availability of accurate sources of data and information that can be used to strive towards an informed-based decision making, such as *i*) the data reliability, which requires quality control, a uncertainty evaluation, and an assessment of their sources (Zambrano-Bigiarini et al. 2017; Baez-Villanueva et al. 2018; Kundzewicz et al. 2018); *ii*) the conversion of data into information in a timely manner (Liu et al. 2015; Senay et al. 2015; Atmoko et al. 2017); *iii*) the generation of useful information for users and stakeholders (Henriksen et al. 2018; Ahmad and Hossain 2019); *iv*) the reduction of operational costs (Calera et al. 2017; Andres et al. 2018); and *v*) the lack of general standards and protocols for data management (Beal and Flynn 2015; Cominola et al. 2015; Cheong et al. 2016). Additionally, in many regions (especially in developing countries), relevant data that have the potential to be used for water resources management (e.g., hydrometeorological and agricultural monitoring networks) are often sparse or non-existent and have large latency. Therefore, their use is impractical for real-time decision-making (Sheffield et al. 2018). Several open-access global gridded products derived through remote sensing data and reanalysis models have become operational in the past decades to overcome this challenge. These products are often used as complementary information for water resources management over diverse regions, although they are often the only source of information over data-scarce settings (Sheffield et al. 2018).

1.2 Information to improve water resources management in data-scarce environments

In many regions, the hydrometeorological monitoring networks have large latency and are sparse, which hinders their use for real-time decision-making (Sheffield et al. 2018) and have been declining over the past 30 years due to an ongoing lack of investment and infrastructure maintenance (Lorenz and Kunstmann 2012; Fay et al. 2017; Sheffield et al. 2018). The inclusion of additional information and knowledge such as satellite data, complementary airborne monitoring systems, and the application of tools that can generate runoff predictions over a range of spatio-temporal scales is pivotal to improving water resources management (Blöschl et al. 2013; Cosgrove and Loucks 2015).

Satellite-based and reanalysis products can be used to account for all components of the hydrological cycle over data-scarce settings (McCabe et al. 2017), such as precipitation, streamflow, actual and potential evaporation, soil moisture, snow cover, and surface and subsurface water levels. Therefore, these products can be used to provide valuable information for supporting water management, developing early warning systems, and planning and designing associated infrastructure (Sheffield et al. 2018). Table 1.1 presents a summary of state-of-the-art hydrometeorological datasets that can be used for water resources management.

Table 1.2 is adapted from Sheffield et al. (2018) and includes more examples for water management (Hong et al. 2007a; Hellegers et al. 2010; Rossa et al. 2010; Hirpa et al. 2013; Bello and Aina 2014; Brown et al. 2014; Senay et al. 2015; Eggimann et al. 2017; Swain and Patra 2017; Herman et al. 2018; Zambrano-Bigiarini and Baez-Villaneuva 2019; Beck et al. 2020a; Gerlitz et al. 2020; Baez-Villanueva et al. 2021; Jurečka et al. 2021; Salvia et al. 2021) focusing on the water resources management decisions that can be taken with the use of these hydrometeorological datasets. However, despite the potential that these products have for operational applications, they are still subject to errors and uncertainties that hinder their use (Scanlon et al. 2016; Baez-Villanueva et al. 2018; McNamara et al. 2021; Beck et al. 2021b); and therefore, they must be evaluated before any hydrological, ecological, or agricultural application (Zambrano-Bigiarini et al. 2017).

CHAPTER 1. INTRODUCTION

Table 1.1: Summary of available open-access gridded products.

Product	Spatio-temporal res.	Spatial coverage	Availability	References
Precipitation				
CHIRPSv2	0.05°; daily	50°N–50°S	1981–present	Funk et al. (2015)
CMORPHv1	0.072°; 30-minutes	60°N–60°S	2002–present	Xie et al. (2017)
ERA5	0.25°; hourly	Global	1950–present	Hersbach et al. (2020)
ERA5 Land	0.10°; hourly	Global	1950–present	Muñoz Sabater et al. (2021)
IMERG-FRv06	0.10°; 30-minutes	Global	2000–present	Huffman et al. (2015)
MSWEPv2.8	0.10°; 3-hourly	Global	1979–present	Beck et al. (2017a)
MSWXv1	0.10°; 3-hourly	Global	1979–7 months in future	Beck et al. (2021a)
PERSIANN	0.25°; hourly	60°N–60°S	2000–present	Sorooshian et al. (2000)
PERSIANN-CCS	0.04°; hourly	60°N–60°S	2003–present	Hong et al. (2007b)
PERSIANN-CDR	0.25°; hourly	60°N–60°S	1983–present	Ashouri et al. (2015)
Actual evaporation				
ERA5	0.25°; hourly	Global	1950–present	Hersbach et al. (2020)
ERA5 Land	0.10°; hourly	Global	1950–present	Muñoz Sabater et al. (2021)
GLEAMv3.5a	0.25°; daily	Global	1980–2020	Martens et al. (2017)
GLEAMv3.5b	0.25°; daily	Global	2003–2020	Martens et al. (2017)
MOD16A2 aqua	0.0043°; 10-day	Global	2008–present	Mu et al. (2011)
MOD16A2 terra	0.0043°; 10-day	Global	2001–present	Mu et al. (2011)
PMLv2	500 m; 8-day	Global	2002–present	Zhang et al. (2019)
SSEBop	0.0092°; 10-day	Global	2003–present	Senay et al. (2013)
Potential evaporation				
ERA5	0.25°; hourly	Global	1950–present	Hersbach et al. (2020)
ERA5 Land	0.10°; hourly	Global	1950–present	Muñoz Sabater et al. (2021)
GLEAMv3.5a	0.25°; daily	Global	1980–2020	Martens et al. (2017)
GLEAMv3.5b	0.25°; daily	Global	2003–2020	Martens et al. (2017)
hPET	0.10°; hourly	Global	1981–2020	Singer et al. (2021)
Soil moisture				
ERA5	0.25°; hourly	Global	1950–present	Hersbach et al. (2020)
ERA5 Land	0.10°; hourly	Global	1950–present	Muñoz Sabater et al. (2021)
ESA-CCI	0.25°; daily	Global	1978–2020	Dorigo et al. (2017)
SMAP-L3SMPE	9 km; daily	Global	2015–present	Entekhabi et al. (2010)
SMAP-L4SM	9 km; 3-hourly	Global	2015–present	Entekhabi et al. (2010)
SMOS	0.25°; 3-hourly	Global	2010–present	Kerr et al. (2010)
Snow cover				
ERA5	0.25°; hourly	Global	1950–present	Hersbach et al. (2020)
ERA5 Land	0.10°; hourly	Global	1950–present	Muñoz Sabater et al. (2021)
ESA-SCFV	5 km; daily	Global	1982–2019	Chen et al. (2021b)
MOD10A1	500 m; daily	Global	2000–present	Hall and Riggs (2007)
Storage				
GRACE	~300 km; monthly	Global	2002–2017	Tapley et al. (2004)
GRACE-FO	~300 km; monthly	Global	2018–present	Kornfeld et al. (2019)

CHAPTER 1. INTRODUCTION

Table 1.2: Summary of water resources management decisions that can be made with hydrometeorological data. This table has been adapted from Sheffield et al. (2018).

Water resources management practice	Hydrometeorological data and references
Planning and design	
Design of flood controls	Long-term precipitation, streamflow, and snow water equivalent in cold areas (Rossa et al. 2010; Brown et al. 2014; Sheffield et al. 2018).
Design of hydropower systems	Long-term precipitation, streamflow, and snow water equivalent in cold areas (Sheffield et al. 2018).
Design of irrigation systems	Long-term precipitation, streamflow, groundwater, and crop water use (Sheffield et al. 2018).
Design of wastewater treatment systems	Streamflow and groundwater (Eggimann et al. 2017; Sheffield et al. 2018).
Design of water supply systems	Long-term precipitation, streamflow, and groundwater (Sheffield et al. 2018).
Transboundary water agreements (to analyse water uses and needs, and design management strategies)	Long-term precipitation, streamflow, groundwater, and water use records (Sheffield et al. 2018).
Prediction of streamflow in ungauged catchments	Long-term precipitation, temperature, potential evaporation, and streamflow (Beck et al. 2020a; Baez-Villanueva et al. 2021).
Improvement of hydrological modelling	Long-term precipitation, temperature, potential evaporation, and streamflow (Herman et al. 2018).
Prediction of crop yield to support decision-making	Precipitation and actual evaporation (Jurečka et al. 2021).
Management and operations	
Water resources management (to satisfy water demand given supply), water availability and consumption	Real-time precipitation, evaporation, streamflow, and groundwater (Hellegers et al. 2010; Sheffield et al. 2018).
Water supply operations	Real-time precipitation, evaporation, streamflow, groundwater, and precipitation forecasts (Sheffield et al. 2018).
Maximisation of hydropower production	Real-time precipitation, evaporation, streamflow, groundwater, and precipitation forecasts (Sheffield et al. 2018).
Reservoir operations	Real-time precipitation, evaporation, streamflow, snow water equivalent, groundwater, and precipitation forecasts (Sheffield et al. 2018).
Wastewater management	Real-time streamflow and groundwater data (Sheffield et al. 2018).
Irrigation system operations	Real-time precipitation, evaporation, streamflow, groundwater, and soil moisture (Sheffield et al. 2018).
Ecosystem management	Near real-time precipitation, streamflow, water levels, soil moisture, and water quality (Sheffield et al. 2018).
Water bodies management	Precipitation and streamflow Eggimann et al. (2017).
Propagation of river flow waves	Real-time and forecasts of precipitation (Hirpa et al. 2013).
Production of river flow nowcasts and forecasts	Real-time and forecasts of precipitation, temperature, potential evaporation, and snow water equivalent (Bello and Aina 2014; Gerlitz et al. 2020).
Disaster management	
Crop monitoring and early warning systems	Real-time precipitation, soil moisture, crop water needs, vegetation characteristics, and precipitation forecasts (Sheffield et al. 2018).
Drought monitoring and early warning systems	Real-time precipitation, temperature, evaporation, streamflow, groundwater, soil moisture, precipitation forecasts (Senay et al. 2015; Swain and Patra 2017; Sheffield et al. 2018; Zambrano-Bigiarini and Baez-Villanueva 2019; Salvia et al. 2021).
Detection of precipitation-triggered landslides	Real-time precipitation (Hong et al. 2007a).

1.3 The role of streamflow and precipitation

Daily streamflow data are crucial for a diverse range of operational and scientific applications such as water resources management (Mohammadi et al. 2020; Niu and Feng 2021); climate change assessment (e.g., Kling et al. 2012; Rojas et al. 2013; Mendoza et al. 2016); and flood forecasting (e.g., Coughlan et al. 2016; Sharma et al. 2018), among others. An accurate representation of daily streamflow can help to shift towards proactive approaches of water management by relying on an information-based decision-making process (see Table 1.2). This is of particular importance over arid and semi-arid regions as water resources are more limited in such regions, which may exacerbate the competition for water among different sectors. As streamflow is the variable that draws the most attention from a water management perspective (McCabe et al. 2017), a deep understanding of the water resources dynamics at the catchment scale is crucial to achieve water security and sustainable water management (Sheffield et al. 2018). However, the vast majority of rivers worldwide remain ungauged (Young 2006; Beck et al. 2016); and therefore, there is a need to maintain and improve the global and regional monitoring systems.

Additionally, to streamflow data, there is a need for information related to the main variables that constitute the hydrological cycle and their inter-annual and intra-annual variability (Walker 2000; García et al. 2016; Sheffield et al. 2018). In this sense, precipitation is the main component of the water cycle; and therefore, its accurate representation is crucial for water resources management (Eggimann et al. 2017). Precipitation can be used for a wide range of applications related to water resources management, such as *i*) flood monitoring and early warning systems through implementing hydrological models, threshold runoff values, or numerical models (Liu et al. 2018; Corral et al. 2019; Hofmann and Schüttrumpf 2019); *ii*) drought monitoring and early warning systems, by automating the computation of meteorological, soil moisture, and hydrological indices (Zambrano-Bigiarini and Baez-Villaneuva 2019; Salvia et al. 2021); *iii*) water demand and availability analysis, through the estimation of water demand and supply (Hellegers et al. 2010; Sheffield et al. 2018); *iv*) dams and hydropower plants operations, through the application of hydrological modelling, forecasts, and climate projections (Wei and Xun 2019; Koppa et al. 2021; Zhao et al. 2021); and *v*) ecosystems and disaster management, to derive the hazard component or river corridors (Wright et al. 2017; Tomsett and Leyland 2019).

1.3.1 Precipitation products and their uncertainties

Rain gauges provide direct measurements of precipitation on a given spot and are typically considered to be the source of reference data for precipitation observations (Tapiador et al. 2012). However, in data-scarce settings, an accurate representation of the spatio-temporal patterns of precipitation when only ground-based measurements are used is subject to large uncertainties (Woldemeskel et al. 2013; Rana et al. 2015; Adhikary et al. 2015; Xavier et al. 2016; Kidd et al. 2017). These uncertainties are related to *i*) the lack of precipitation information over large areas, which is crucial to evaluate the extent of a particular precipitation event, and *ii*) the lack of precipitation stations at high altitudes, which may result in an underestimation of precipitation. Therefore, complementary information can improve the characterisation of precipitation at different temporal scales for further hydrological, ecological, and agricultural applications.

In the last decades, several gridded precipitation products have been developed using different data sources such as ground-based data, satellite-based information, and reanalysis models. The products developed using ground-based data rely on point-based measurements (Chen et al. 2008), which are interpolated through various techniques and methods to produce continuous precipitation fields (e.g., CPC Unified; Chen et al. 2008). These products could present larger errors in *i*) areas where ground-based measurements were not used in their development (Wei et al. 2021) and *ii*) areas far from the UTC ± 00 due to the difference in reporting times between gauge stations and gridded products (Beck et al. 2019a).

Satellite estimates are mainly based on infrared (IR), passive microwave (PMW) data, or a combination of both. The methods that use IR data are based on the assumption that cold cloud-top temperature is related to rain processes. Despite their high temporal resolution, the estimates derived solely from IR data may face some problems such as *i*) the interference of multi-layer cloud systems in the detection of the cloud layer that is precipitating (Tapiador et al. 2012) and *ii*) the dependence of the statistical relationship between cloud-top temperature and rainfall, seasonality, and storm type (Vicente et al. 1998). On the other hand, PMW is a more direct approach to measure precipitation from satellites because PMW sensors use frequencies at a range in which precipitation-sized particles absorb, emit, and scatter radiation (Tapiador et al. 2012). Some challenges related to this approach are *i*) the revisiting time of low-earth-orbit satellites that reduce the temporal resolution of the measurements and *ii*) the beam-filling effect related to convective precipitation (Kummerow 1998).

The reanalysis products provide a large number of atmospheric, land, and oceanic climate variables and are based on forecast models and data assimilation systems to generate a numerical description of the climate based on the combination of model outputs with observations (ECMWF 2021). Some examples of reanalysis-based precipitation products are the reanalysis JRA-55 (Kobayashi et al. 2015), ERA-Interim (Dee et al. 2011), and ERA5 (Hersbach et al. 2020). However, reanalysis products present better performances over high latitudes as they present persistent, large-scale stratiform precipitation systems (Beck et al. 2017b). Additionally, some products use more than one of the aforementioned techniques to produce their estimates. For example, the Multi-Source Weighted-Ensemble Precipitation (MSWEPv2.8; Beck et al. 2017a) and the Climate Hazards InfraRed Precipitation with Stations data version 2.0 (CHIRPSv2.0 Funk et al. 2015) incorporate satellite imagery, reanalysis data, and ground-based measurements at the daily and monthly temporal scales, respectively.

Despite that these products provide an unprecedented opportunity to account for the spatio-temporal distribution of precipitation, many studies that have evaluated their performance have found that they are still subject to uncertainties and errors (Espinoza Villar et al. 2009; Dinku et al. 2010; Melo et al. 2015; Beck et al. 2017b; Zambrano-Bigiarini et al. 2017), which are related to the detection of single precipitation events, precipitation totals, and the distribution of precipitation intensity (Beck et al. 2019a). In this sense, there is still not a global best-performing product; therefore, these products must be evaluated before any operational application (Zambrano-Bigiarini et al. 2017).

1.3.2 Streamflow prediction in ungauged catchments

As the vast majority of rivers worldwide remain ungauged (Young 2006; Beck et al. 2016), streamflow is often estimated through the implementation of hydrological models. These models are calibrated using observed streamflow and require precipitation, potential evaporation, and temperature data as inputs (among other variables related to the geomorphological and physical characteristics of the catchment). The scientific initiative Prediction in Ungauged Basins (PUB; see review by Hrachowitz et al. 2013) has fostered the development and application of diverse methods to predict streamflow in ungauged catchments through the transfer of model parameters from controlled to ungauged catchments in a process known as regionalisation (Oudin et al. 2008). According to Blöschl et al. (2013) there is a need for streamflow predictions in ungauged catchments for diverse water management applications, such as:

CHAPTER 1. INTRODUCTION

1. water allocation, long term planning, and groundwater recharge (related to long-term streamflow volume);
2. water supply, hydropower production and potential, planning, restoration measures, ecological purposes, and irrigation (related to the inter-annual and intra-annual variability of streamflow);
3. environmental flows for ecological health, drought management, river restoration, water quality, design of spillways, culverts and dams, reservoir management, and risk management (related to the occurrence of dry and wet periods and future projections of extremes); and
4. runoff dynamics, sediments and nutrients prediction (related to all of the above points, as well as sediments and nutrients transportation).

There are three main regionalisation approaches to predict daily streamflow in ungauged catchments by transferring the calibrated parameters of a hydrological model: *i*) spatial proximity, which assumes that neighbouring catchments exhibit similar hydrological behaviour because physiographic and climatic characteristics are relatively homogeneous in a region (Vandewiele and Elias 1995; Oudin et al. 2008); *ii*) feature similarity, which transfers the model parameters of gauged to ungauged catchments based on their degree of climatic and geomorphological similarity (McIntyre et al. 2005; Beck et al. 2016; Carrillo et al. 2011); and *iii*) parameter regression, in which statistical relationships between model parameters and catchment characteristics are developed and used to estimate the model parameters of ungauged catchments (Fernandez et al. 2000; Carrillo et al. 2011).

The application of these methods require reliable meteorological information and catchment-specific data; therefore, most studies have been conducted over regions with dense networks of meteorological stations (e.g., McIntyre et al. 2005; Parajka et al. 2005; Oudin et al. 2008; Bao et al. 2012; Singh et al. 2012; Zelelew and Alfredsen 2014; Garambois et al. 2015; Athira et al. 2016; Rakovec et al. 2016; Swain and Patra 2017; Saadi et al. 2019; Neri et al. 2020).

1.4 Thesis objectives and outline

Precipitation is the major component of the water cycle and its accurate representation is crucial for hydrological, ecological, and agricultural applications. As observed, there is not a best performing precipitation product on a general basis, and their errors and uncertainties hinder their use. Therefore, they must be evaluated before any operational application to assess whether they are suitable to be used in a specific region (Zambrano-Bigiarini et al. 2017).

In this sense, there is a need for generating better precipitation estimates. However, the big challenge lies in integrating different sources of data into coherent and usable information (Eggimann et al. 2017). Interpolation and measurement uncertainties must be adequately considered (Villarini and Krajewski 2008; Rios Gaona et al. 2015) and large amounts of data must be processed. Several approaches have been implemented to derive gridded precipitation and other climatological variables using point-based information and gridded products (Li and Shao 2010; Xie and Xiong 2011; Jewell and Gaussiat 2015; Fu et al. 2016; Manz et al. 2016; Yang et al. 2017; Beck et al. 2019b).

Despite the improvements in the spatio-temporal representation of precipitation achieved by these methods, many studies only merge the ground-based measurements with a single precipitation product (e.g., Li and Shao 2010; Rozante et al. 2010; Verdin et al. 2016; Shi et al. 2017; Xie et al. 2017; Yang et al. 2017). As a result, valuable information that could be better captured by other products is not considered. Additionally, these merging methods are generally complex and difficult to implement. For these reasons, there is a clear need to develop a merging procedure that takes advantage of combining information of multiple precipitation products and ground-based measurements in a friendly way. For this purpose, it is essential to analyse the performance of the products to determine the origin of the sources of mismatches and errors.

Although hydrological model calibration can partly compensate for errors included in these precipitation products (Elsner et al. 2014; Maggioni and Massari 2018), it may lead to unrealistic model behaviour (Nikolopoulos et al. 2013; Xue et al. 2013; Ciabatta et al. 2016) that affects the quality of hydrological modelling and parameter regionalisation results. To date, most of the studies have been conducted over regions with a dense network of meteorological stations (e.g., McIntyre et al. 2005; Parajka et al. 2005; Oudin et al. 2008; Bao et al. 2012; Singh et al. 2012; Zelelew and Alfredsen 2014; Garambois et al. 2015; Athira et al. 2016; Rakovec et al. 2016; Swain and Patra 2017; Saadi et al. 2019; Neri et al. 2020), while just a few

CHAPTER 1. INTRODUCTION

have used gridded precipitation products at the daily time scale, by selecting a best performing product during calibration (Beck et al. 2016; Rakovec et al. 2016; Beck et al. 2021b). Therefore, there is a need to understand how selecting a particular precipitation product may affect hydrological modelling and parameter regionalisation results.

In light of these research demands, this cumulative dissertation aims to predict daily streamflow data over ungauged catchments following a three-step approach as shown in Figure 1.1: *i*) evaluation, where selected state-of-the-art precipitation products are evaluated over different catchments to understand the sources of their mismatches; *ii*) merging, where the knowledge obtained in step *i* was used to propose a novel method to merge precipitation products, topography-related variables, and ground-based measurements to improve the spatio-temporal characterisation of precipitation; and *iii*) regionalisation, where the performance of the merged product, altogether with other state-of-the-art products was evaluated during a regionalisation exercise to predict daily streamflow over ungauged catchments. To address the main objective of this cumulative dissertation, six sub-objectives are proposed and enumerated as follows:

1. to assess the spatio-temporal performance of different precipitation products over areas with different climatological characteristics;
2. to analyse the influence of the spatial resolution of the precipitation products when compared to ground-based measurements;
3. to improve the characterisation of the spatio-temporal distribution of precipitation over data-scarce regions;
4. to evaluate methods to predict daily streamflow values in ungauged catchments over data-scarce settings;
5. to evaluate to what extent the choice of precipitation forcings affects the relative performance of hydrological modelling and regionalisation procedures; and
6. to analyse the relative performance of different regionalisation techniques across catchments with different hydrological regimes.

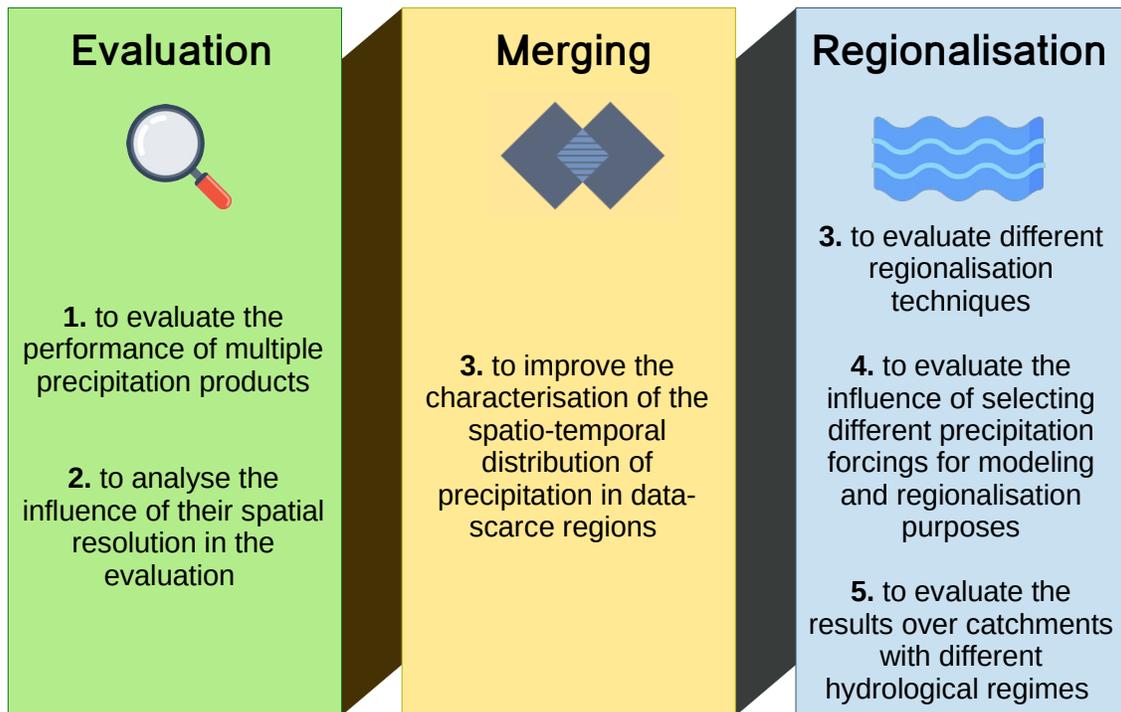


Figure 1.1: Conceptualisation of the three different components of the methodology followed in this cumulative dissertation: *i*) the evaluation step (Baez-Villanueva et al. 2018); *ii*) the merging step (Baez-Villanueva et al. 2020); and *iii*) the regionalisation step (Baez-Villanueva et al. 2021).

Chapter 2

Study Areas

To fulfil the specific objectives proposed in this cumulative dissertation, the methods were divided into three main components as described in Section 1.4. Three catchments were selected for the evaluation component and are presented in Section 2.1, while the merging and regionalisation sections were performed over continental Chile, which is described in Section 2.2.

2.1 Selected catchments for the evaluation of precipitation products

Three catchments were selected for the performance evaluation of multiple precipitation products and are presented in Figure 2.1: *i*) the Magdalena in Colombia; *ii*) the Paraiba do Sul in Brazil; and *iii*) the Imperial in Chile. The different study areas were selected to compare the performance of the chosen precipitation products over different Latin American environments. The selected catchments present differences mainly in climate, topography, and location. According to the Köppen-Geiger climate classification of Beck et al. (2018), and as observed in Figure 2.2, the Imperial presents a temperate climate with dry and warm summer (Csb) in the northwestern region of the catchment and a temperate climate without a dry season and warm summer (Cfb) in the south-eastern area. Paraiba do Sul presents a temperate climate with dry winter and hot summer (Cfa) and without a dry season and warm summer (Cfb) in the upper part of the catchment that transitions into a tropical monsoon climate (Aw) near the outlet. Finally, the Magdalena shows the greatest variety of climates, with tropical rain-forest (Af), tropical monsoon (Am) and tropical savannah (Aw) climates in the northern and southern regions of the catchment, and temperate

CHAPTER 2. STUDY AREAS

climate without a dry season and warm summer (Cfb), temperate climate with dry and warm summer (Csb), and polar tundra (ET) climates in the elevated areas of the catchment.

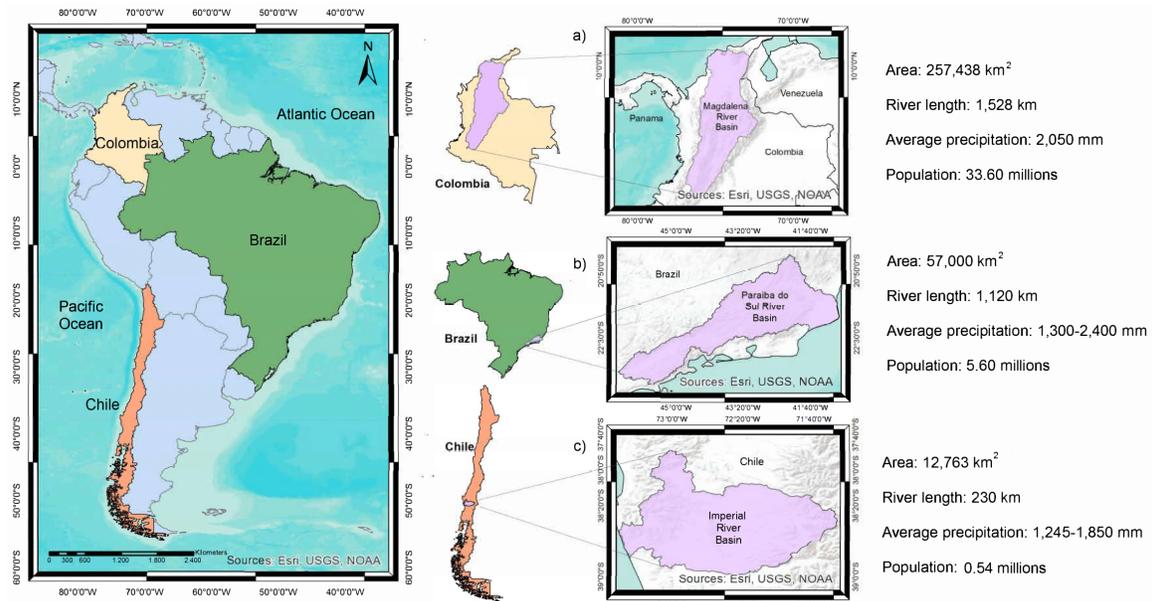


Figure 2.1: The three selected study areas; *a)* the Magdalena, located in Colombia; *b)* the Paraiba do Sul, located in Brazil; and *c)* the Imperial, located in Chile.

The Magdalena is essential in the economic development of Colombia, as it is the greatest catchment in the country. The Paraiba do Sul is the catchment that contributes the most to the Brazilian gross domestic product (GDP), as it includes the states of Rio de Janeiro, Minas Gerais, and Sao Paulo. Finally, the Imperial is important because most of the agricultural activity in Chile is performed in the southern regions.

2.1.1 The Magdalena

The Magdalena is located in Colombia, between the longitudes $76^{\circ} 58'W$ and $72^{\circ} 22'W$ and the latitudes $11^{\circ} 06'S$ and $1^{\circ} 33'S$ and has an elevation that ranges from zero to 5,000 m a.s.l. The main urban centres of Colombia are located within this catchment and have around 33.60 million inhabitants. This is reflected in the contribution of the catchment to the Colombian GDP (around 85%). The Magdalena River is the longest in Colombia with a longitude of around 1,528 km and drains the Andes mountains, which are formed in Colombia by the Western, Central, and Eastern Cordilleras. It covers an area of 257,438 km², which represents 24% of Colombia

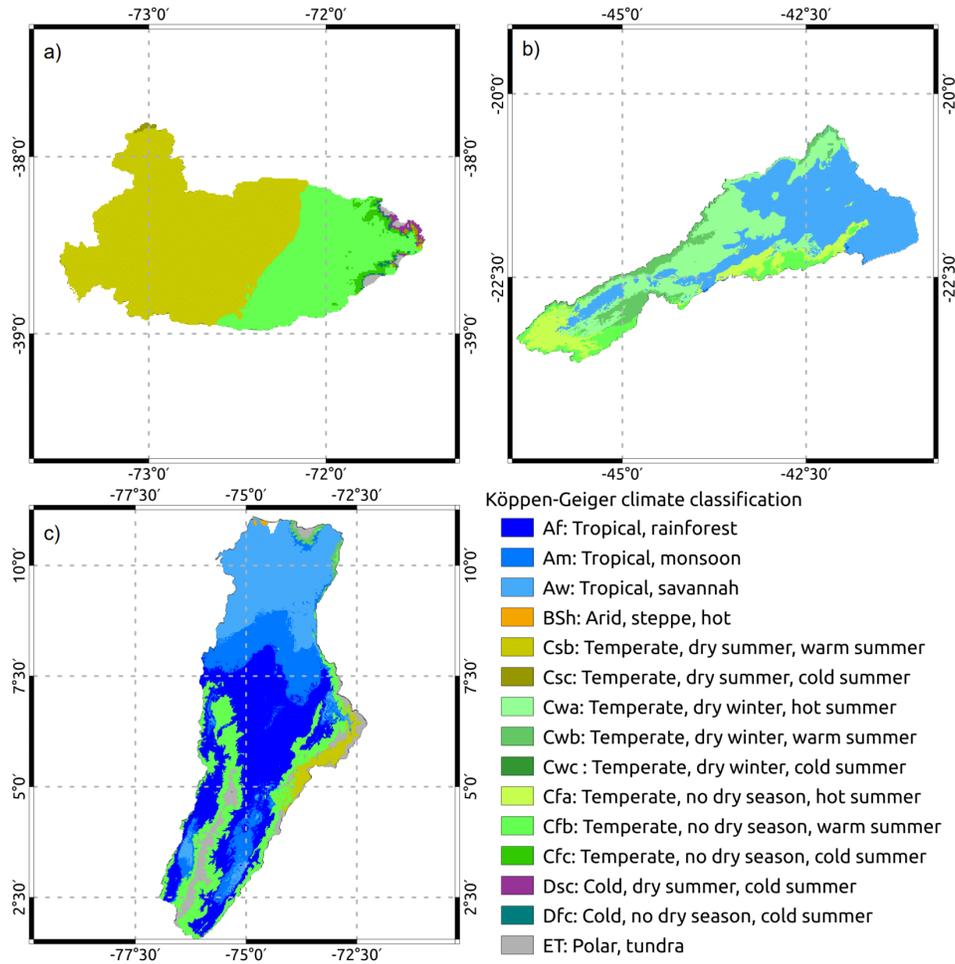


Figure 2.2: Köppen-Geiger climate classification over the three study areas according to Beck et al. (2018); the Imperial (top-left); the Paraiba do Sul (top-right); and the Magdalena (bottom-left).

(Restrepo et al. 2006). The Magdalena has two wet (March–May and October–November) and two dry (December–February and June–September) seasons (IDEAM 2001) and has a mean annual precipitation of 2,050 mm.

2.1.2 The Paraiba do Sul

The Paraiba do Sul is located in Brazil, between the longitudes 41°W and $46^{\circ} 30'\text{W}$ and the latitudes $20^{\circ} 26'\text{S}$ and $23^{\circ} 39'\text{S}$ and covers an area of around $57,000 \text{ km}^2$. It has a heterogeneous topography, hydrology, geomorphology, and soil composition (Simoes and Barros 2007). It is bordered by two mountain ranges (Serra do Mar and Serra da Mantiqueira) and its elevation ranges from 450 to 2,000 m a.s.l. (Soares

CHAPTER 2. STUDY AREAS

et al. 2012). The length of Paraíba do Sul is approximately 1,120 km and the states of Minas Gerais, Rio de Janeiro, and Sao Paulo are included in the catchment. It accounts for approximately 11% of the national GDP. About 8.70 million inhabitants live in the metropolitan area of Rio de Janeiro, which depend on the Paraíba do Sul for water supply (Silva and Simões 2014). It has a large inter-annual and spatial variability with precipitation values ranging between 1,300 and 2,400 mm (Simoes and Barros 2007; Soares et al. 2012) and has a mean annual precipitation of about 1,400 mm.

2.1.3 The Imperial

The Imperial is a rain-dominated catchment, which is located in the Araucanía region in Chile between the longitudes $73^{\circ} 30'W$ and $71^{\circ} 27'W$ and the latitudes $37^{\circ} 40'S$ and $38^{\circ} 50'S$. It has an area of around $12,763 \text{ km}^2$ and a river length of about 230 km. This catchment has low snow accumulation as a consequence of the relatively small altitude in the Andean mountains at this latitude (Rivera et al. 2004). The imperial has around 540,600 inhabitants and its elevation ranges from zero to 3,095 m a.s.l. The mean annual precipitation for the Imperial ranges from 1,245 to 1,850 mm.

Figure 2.3 shows the mean monthly precipitation values (in the case of the Imperial and Paraíba do Sul for 2001–2015, and in the case of Magdalena for 2001–2014). The Magdalena shows a bimodal distribution of precipitation caused by the double pass of the Intertropical Convergence Zone (ITCZ), having the highest values during April–May and October–November. The Paraíba do Sul presents its rainy season during October–March, while for the Imperial, the rainy season starts in April and ends in September. The Paraíba do Sul and Imperial catchments present a greater dispersion from the mean precipitation during the humid season. Contrastingly, the Magdalena presents a high dispersion throughout all seasons.

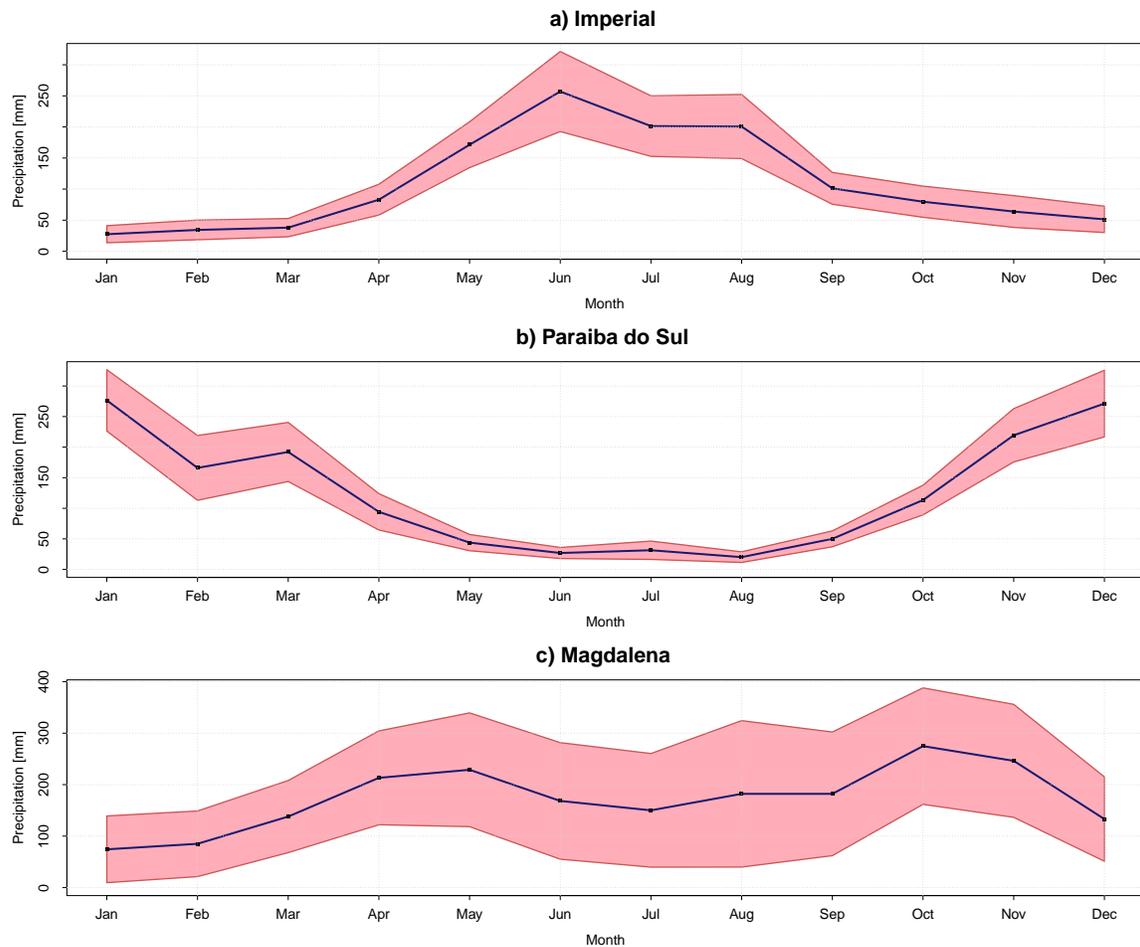


Figure 2.3: Mean annual precipitation of the three study regions: *a*) the Imperial (during 2001–2015); *b*) the Paraiba do Sul (during 2001–2015); and *c*) the Magdalena (during 2001–2014). The blue line represents the mean values for each catchment and the red contour represents one standard deviation above and below the mean value.

2.2 Continental Chile

Chile (Figure 2.4) is bounded to the north by Peru, to the east by Bolivia and Argentina, and to the west by the Pacific Ocean. It spans on average 180 km of longitudinal extension (76.0°W – 66.0°W) and 4,300 km of latitudinal extension (17.5°S – 56.0°S). The country’s geography is dominated by complex topography, with an elevation profile ranging from zero to 6,891 m a.s.l. (Jarvis et al. 2008). Chile exhibits four major geographical units: the Andes Mountains, the Coastal Mountains, the Coastal Plains, and the Intermediate Depression (Valdés-Pineda et al. 2014). The four seasons of the southern hemisphere are present: autumn (MAM), winter (JJA), spring (SON), and summer (DJF).

Figure 2.4 shows the elevation (Jarvis et al. 2008), land cover (Zhao et al. 2016), and the most updated Köppen-Geiger climate classification (Beck et al. 2018) for the five major macroclimatic zones presented in Zambrano-Bigiarini et al. (2017). Chile has a large variety of climates, transitioning from (hyper)arid and semi-arid climates in the Far North (17.50 – 26.00°S) and Near North (26.00 – 32.18°S), through temperate climates in Central Chile (32.18 – 36.40°S), to more humid and polar climates in the South (36.40 – 43.70°S) and Far South (43.70 – 56.00°S). Precipitation increases with latitude and elevation (in the southern direction; Montecinos and Aceituno 2003) and ranges from almost zero in the Atacama Desert to about $6,000\text{ mm yr}^{-1}$ in the surroundings of Puerto Cardenas ($\sim 43.2^{\circ}\text{S}$). Similar to precipitation, the mean annual streamflow and the rainfall-runoff ratio tend to increase from north to south (Alvarez-Garretón et al. 2018; Vásquez et al. 2021).

The El Niño-Southern Oscillation (ENSO) has a large impact on precipitation during winter, with positive anomalies during El Niño and negative anomalies during La Niña events (Verbist et al. 2010; Robertson et al. 2014). Although neutral ENSO conditions have prevailed since 2011, except for an intense El Niño event during 2015, an uninterrupted sequence of dry years with increased temperatures has been observed from 2010–2018, with annual precipitation deficits of about 25–45% across Chile. This long-term deficit in precipitation volume, also known as the Chilean megadrought (Boisier et al. 2016; Garreaud et al. 2017), has reduced river flows, reservoir storage, snow cover, and groundwater levels across Chile (Garreaud et al. 2017, 2020).

CHAPTER 2. STUDY AREAS

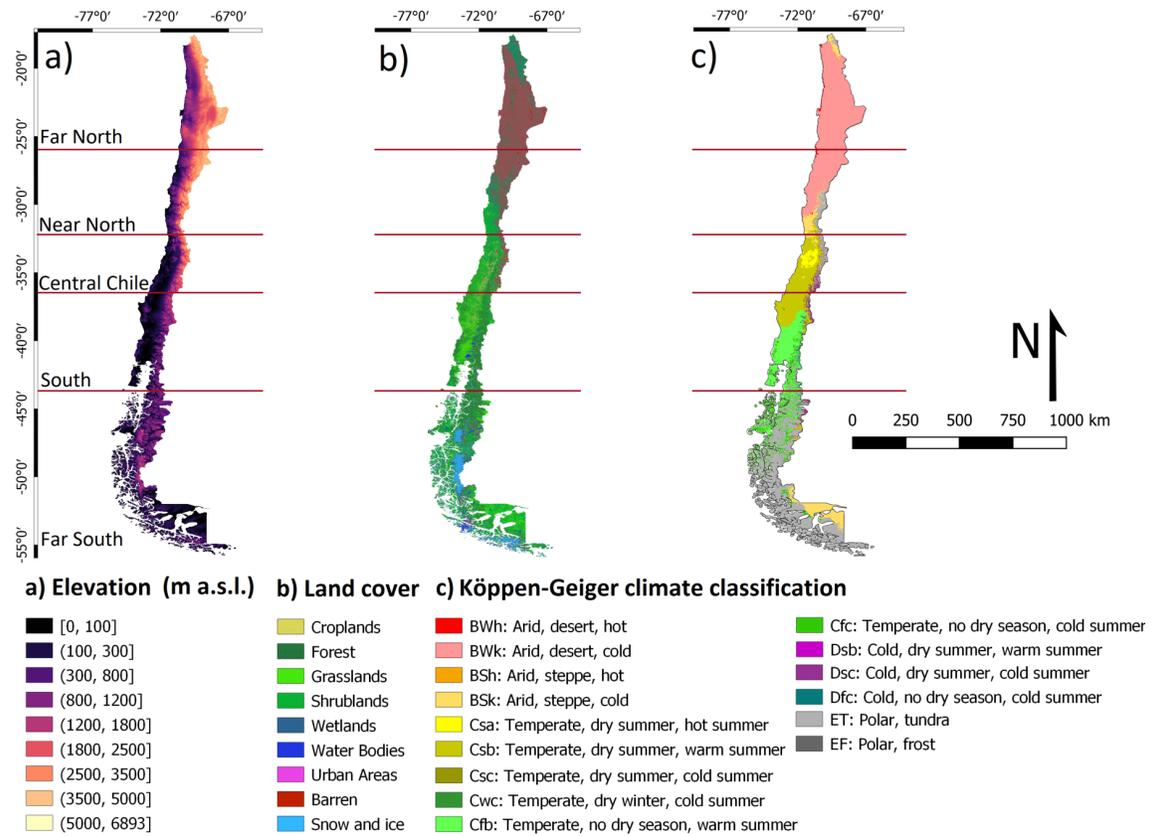


Figure 2.4: Study area: *a*) elevation (SRTMv4.1; Jarvis et al. 2008); *b*) land cover classification (Zhao et al. 2016); and *c*) Köppen-Geiger climate classification (Beck et al. 2018) over the five major macroclimatic zones described in Zambrano-Bigiarini et al. (2017).

CHAPTER 2. STUDY AREAS

Chapter 3

Data and Methods

As mentioned in Section 1.4, to fulfil the objectives of this cumulative dissertation, a three-step approach was followed: *i*) evaluation, where selected state-of-the-art precipitation products are evaluated over different catchments to understand the sources of their mismatches; *ii*) merging, where the knowledge obtained in step *i* was used to propose a novel method to merge precipitation products, topography-related variables, and ground-based measurements to improve the spatio-temporal characterisation of precipitation; and *iii*) regionalisation, where the performance of the merged product, altogether with other state-of-the-art products was evaluated during a regionalisation exercise to predict daily streamflow over ungauged catchments. A detailed flow chart of the methodology is presented in Figure 3.1. The products used in this cumulative dissertation were selected because *i*) they have shown good performance over the study areas and *ii*) they were state-of-the-art products and still under production at the time of performing the analyses.

3.1 Evaluation

In the evaluation component, the performance of six precipitation products (Table 3.1) was evaluated over the three catchments presented in Section 2.1 against ground-based measurements. For this purpose, time series of ground-based precipitation were obtained for each study area for the periods 2001–2015 for the case of Imperial and Paraíba do Sul, and 2001–2014 for the case of Magdalena. The starting date of the evaluation (2001) is related to the product’s availability, while the end date to the availability of the ground-based measurements.

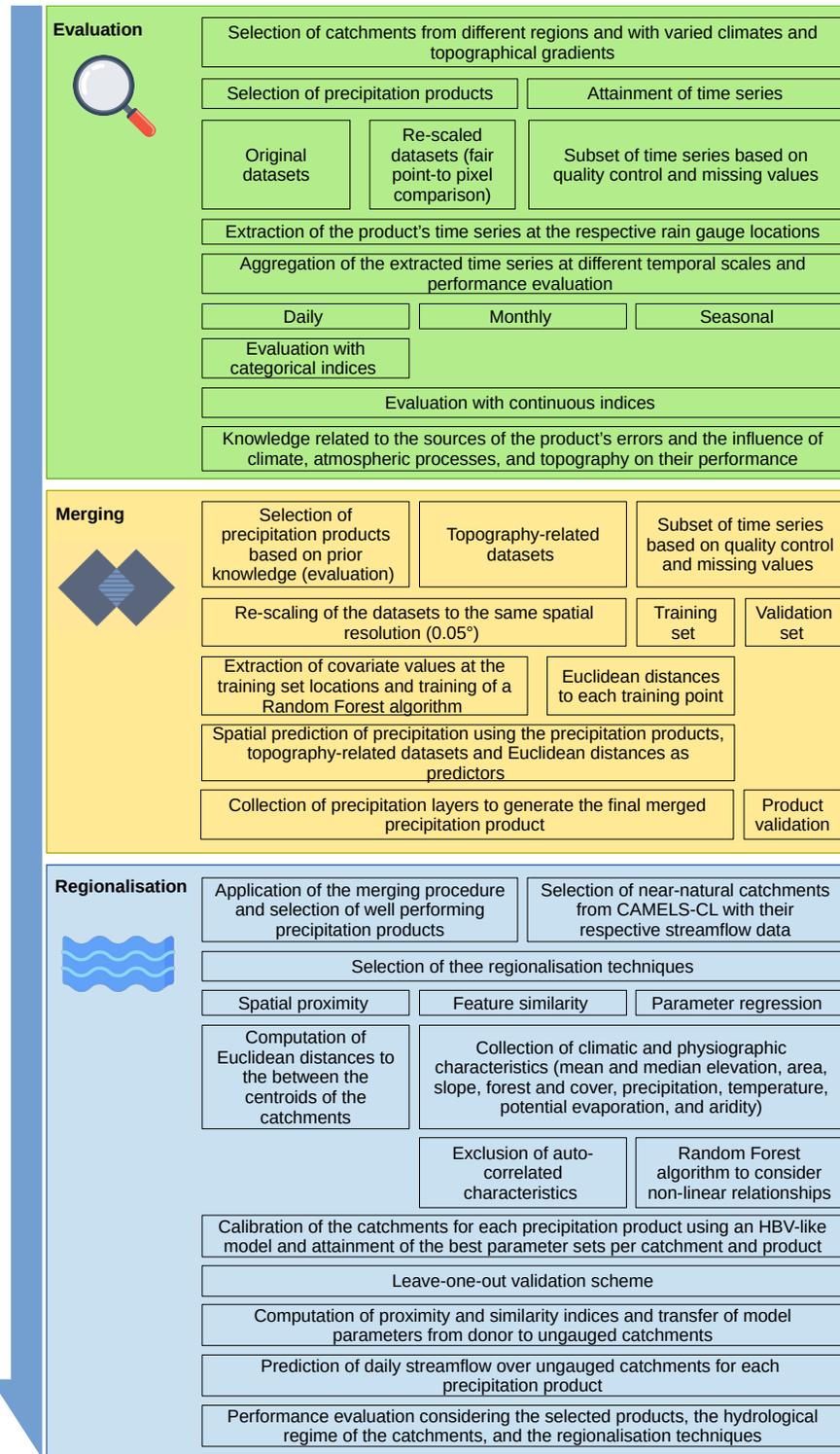


Figure 3.1: Flowchart summarising the three-step procedure followed in the methodology presented in this cumulative dissertation: *i*) evaluation; *ii*) merging; and *iii*) regionalisation. For a more detailed explanation please see Baez-Villanueva et al. (2018, 2020, 2021).

CHAPTER 3. DATA AND METHODS

For the case of Imperial, Paraiba do Sul, and Magdalena, 13, 64, and 124 quality-controlled rain gauges were selected, respectively. For a more detailed description of the products and their selection, the ground-based measurements, and the quality control criteria, the readers are referred to Baez-Villanueva et al. (2018).

Table 3.1: Evaluated precipitation products.

Product	Spatial res.	Spatial coverage	Availability	References
TRMM 3B42RT	0.25°	50°N–50°S	2000(Feb)–2020	Huffman et al. (2007)
TRMM 3B42v7	0.25°	50°N–50°S	1998–2020	Huffman et al. (2007)
CHIRPSv2	0.05°	50°N–50°S	1981–present	Funk et al. (2015)
CMORPHv1	0.25°	60°N–60°S	1998–present	Xie et al. (2017)
PERSIANN-CDR	0.25°	60°N–60°S	1983–present	Ashouri et al. (2015)
MSWEPv2	0.10°	Global	1979–2018	Beck et al. (2017a,b)

The evaluation of the selected precipitation products was performed through a point-to-pixel analysis over each study area, which compares the time series of the selected rain gauge stations against the corresponding grid-cell values of the selected products at the daily, monthly, and seasonal temporal scales. This method has been widely used for evaluating the performance of precipitation products (Thiemig et al. 2012; Dos Reis et al. 2017; Zambrano-Bigiarini et al. 2017) and assumes that the rain gauge stations are representative values of the respective grid-cells of the products.

The comparison between the selected products is not completely impartial because some products use ground-based data to correct the bias of their precipitation estimates. This is the case of TRMM 3B42v7, CMORPHv1, and PERSIANN-CDR, which use the Global Precipitation Climatology Centre dataset (GPCC; Peterson and Vose 1997); and CHIRPSv2 and MSWEPv2, which use the Global Historical Climate Network (CHCN) and the Global Surface Summary of the Day (GSOD). MSWEPv2 also uses the Latin-American Climate Assessment Dataset (LACAD) and national databases from Mexico, Brazil, Peru, and Iran. The bias-corrected datasets are expected to present a better performance than the products that do not use rain gauge stations (i.e., TRMM 3B42RT).

For the studied period, on average, the GPCC uses around 1, 6, and 12 ground-based stations over the Imperial, Paraiba do Sul, and Magdalena catchments, respectively (see Baez-Villanueva et al. (2018); their Figure 4). Also, the CHCN uses a similar amount of rain gauge stations as the GPCC dataset over the three study areas (Menne et al. 2012). As the GPCC does not provide the information of the specific stations that were used, they could not be removed. However, this number of stations are clearly not enough to capture the spatial variability of precipitation over these catchments. Consequently, this increases the estimation errors of the final

products. On the other hand, the precipitation products that use rain gauge stations from the study areas are prone to perform better than the datasets that do not.

The spatial resolution of all products except CHIRPSv2 and MSWEPv2 is 0.25° . Therefore, CHIRPSv2 and MSWEPv2 were upscaled from 0.05° and 0.1° , respectively, to a new spatial resolution of 0.25° by bilinear interpolation to enable a consistent point-to-pixel comparison between products. However, to evaluate the effect of the upscaling procedure in the performance of the products, two CHIRPSv2 and MSWEPv2 datasets were used: *i*) one with its original spatial resolution (CHIRPSv2, MSWEPv2) and *ii*) the other upscaled to 0.25° (hereafter defined as CHIRPSv2 up-scaled and MSWEPv2 up-scaled).

Six different indices of performance (three continuous and three categorical) were applied over the different regions. The continuous indices are the modified Kling-Gupta efficiency (KGE' Gupta et al. 2009; Kling et al. 2012), the Root Mean Square Error (RMSE), and the percentage bias (PBIAS). The KGE' (Eq. 3.1) compares observed data with estimations, decomposing the total performance into r , β , and γ , which are presented in Equations 3.2, 3.3, and 3.4, respectively:

$$\text{KGE}' = 1 - \sqrt{(r - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \quad (3.1)$$

$$r = \frac{\sum_{i=1}^n (O_i - \bar{O})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}} \quad (3.2)$$

$$\beta = \frac{\mu_s}{\mu_o} \quad (3.3)$$

$$\gamma = \frac{CV_s}{CV_o} = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} \quad (3.4)$$

where n is the number of observations; O_i and S_i are the ground-based measurements and estimates at day i , respectively; μ and σ are the mean and standard deviation, respectively; and \bar{O} and \bar{S} are the arithmetic means of the measurements and the estimates, respectively. r measures the temporal dynamics of the variable; β measures the total volume compared to the ground-based measurements indicating the average tendency of the estimates to underestimate ($\beta < 1$) or overestimate ($\beta > 1$); and γ measures the relative dispersion between the estimates and the ground-based measurements (Gupta et al. 2009; Kling et al. 2012). The optimal value for the KGE' and all its components is one. The KGE' is a useful index because it: *i*) does not assign disproportional weights to mismatches during high events (contrary to indices based

on squared-differences; e.g., the RMSE); *ii*) decomposes the total performance into three components, thus allowing a better understanding of the origin of mismatches (Zambrano-Bigiarini et al. 2017); and *iii*) when used to evaluate precipitation products, it allows a fair comparison of regions with different mean annual precipitation. The KGE' has been widely used in hydrological applications and to evaluate the performance of precipitation products (e.g., Dinku et al. 2007; Hirpa et al. 2010; Thiemiig et al. 2012; Gao and Liu 2013; Li et al. 2013; Thiemiig et al. 2013; Chen et al. 2014; Lievens et al. 2015; Beck et al. 2016; Zambrano-Bigiarini et al. 2017; Beck et al. 2017b).

Besides the KGE' evaluation, the RMSE (Eq. 3.5) and PBIAS (Eq. 3.6) were computed and compared to the results obtained with the KGE'. The RMSE is widely used in evaluating precipitation products, and it was included in this study to assess whether it is a useful measure of performance. The PBIAS measures the average tendency of the simulated values to be larger or smaller than the ground-based measurements. The optimal value of the PBIAS is zero, and values close to zero indicate a low bias, positive values indicate an overestimation, and negative values indicate an underestimation.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.5)$$

where y_i and \hat{y}_i are the ground-based measurements and the corresponding precipitation estimates at day i , respectively.

$$\text{PBIAS} = (100) \frac{\sum_{i=1}^n (S_i - O_i)}{\sum_{i=1}^n (O_i)} \quad (3.6)$$

The performance evaluation was implemented on a daily, monthly, and seasonal basis. In an inter-comparison between ground-based precipitation and precipitation products, the use of the KGE' makes sense because it offers a way to evaluate different components of performance separately. However, the KGE' does not allow identifying the performance of a given product at different precipitation intensities. For this reason, diverse categorical indices were also used at the daily scale to complement the evaluation of these products.

Therefore, to evaluate the performance of precipitation products in capturing different precipitation intensities, three categorical indices of performance were used to assess the different precipitation intensities described in Table 3.2 as recommended by Zambrano-Bigiarini et al. (2017). The probability of detection (POD, Eq. 3.7)

calculates how often the product correctly estimates the precipitation intensity observed at the rain gauge. The frequency bias (f_{bias} , Eq. 3.8) compares the number of events identified by the precipitation product to the number of events registered by the gauge station. If $f_{\text{bias}} > 1$, the number of occurrences of the respective precipitation intensity is overestimated by the product, while $f_{\text{bias}} < 1$ indicates underestimation. The false alarm ratio (FAR, Eq. 3.9) measures the fraction of events that were not correctly identified by the precipitation product. The POD and f_{bias} present their optimal value at one, while FAR presents it at zero.

$$\text{POD} = \frac{H}{H + M} \quad (3.7)$$

$$f_{\text{bias}} = \frac{H + F}{H + M} \quad (3.8)$$

$$\text{FAR} = \frac{F}{H + F} \quad (3.9)$$

where H indicates a hit (an event recorded by both the rain gauge and the precipitation product); M indicates a miss (an event only identified by the rain gauge); and F indicates a false alarm (an event recorded only by the precipitation product).

Table 3.2: Classification of precipitation events based on daily intensity (i) according to Zambrano-Bigiarini et al. (2017).

P event	Intensity (i), [mm d ⁻¹]
No-rain	[0, 1)
Light rain	[1, 5)
Moderate rain	[5, 20)
Heavy rain	[20, 40)
Violent rain	≥ 40

The aforementioned methodology was applied over the study areas using the R environment 3.3.1 (Team 2011) and the raster (Hijmans 2018), hydroGOF (Zambrano-Bigiarini 2017a), and hydroTSM (Zambrano-Bigiarini 2017b) R packages.

3.2 Merging

Based on the knowledge gained during the evaluation of precipitation products over different regions of Latin America, a novel method to improve the spatio-temporal characterisation of precipitation in data-scarce regions at any temporal scale was proposed. Section 3.2.1 summarises the proposed Random Forest Merging Procedure (RF-MEP; Baez-Villanueva et al. 2020), while Section 3.2.2 describes its application to the Chilean case study.

3.2.1 RF-MEP

RF-MEP is based on three key assumptions: *i*) precipitation measurements from rain gauge stations are accurate at the point scale; *ii*) precipitation products are generally biased but contain helpful information about the spatio-temporal patterns of precipitation; and *iii*) the combination of different precipitation products and rain gauge data can better represent the spatio-temporal variability of precipitation than any single product.

RF-MEP uses the Random Forest (RF) machine learning technique to predict the spatial distribution of precipitation by merging information from different gridded products (known as covariates) and quality-controlled ground-based information at a selected temporal scale (e.g., daily, monthly, or annual). Individual predictions are generated from a user-defined number of decision trees based on bootstrap samples using the covariates as predictors. The final prediction is calculated as the average of the individual predictions (Breiman 2001; Prasad et al. 2006; Roy and Larocque 2012; Biau and Scornet 2016; Hengl et al. 2018). For a detailed description of the merging procedure, the readers are referred to Baez-Villanueva et al. (2020, their Figure 1).

The first step to apply the RF-MEP procedure is to acquire the ground-based measurements and the selected covariates. These covariates are *i*) the selected precipitation products and *ii*) topography-related datasets such as a digital surface model (DSM), aspect, rate of elevation change, or slope, which are used to account for the precipitation gradient related to elevation (not to be mistaken with altitude, see McVicar and Körner 2013). Then, the ground-based measurements are quality controlled and checked for homogeneity.

The selected rain gauge stations are divided into two groups: a training set (that will be used to train the RF model) and a validation set (to assess the performance of the merged product). The selected precipitation products and the topography-related datasets are resampled to a desired spatial resolution to ensure identical raster

geometry (identical spatial resolution, spatial extent, and origin). The traditional RF algorithm ignores sampling locations, which could lead to sub-optimal predictions (Hengl et al. 2018); therefore, it is important to include covariates that account for geographical proximity. As the use of only geographical coordinates as spatial predictors can cause unnatural surfaces in the merged product (Behrens et al. 2018; Hengl et al. 2018), RF-MEP uses gridded layers of Euclidean distances from each rain gauge in the training set to the centroid of all the grid-cells in the selected study area.

Afterwards, a single RF regression model is derived for each time step to compute a single precipitation prediction at the desired temporal resolution. The RF model is trained using the ground-based observations in the training set as the dependent variable, while the grid-cell values of the selected covariates at the corresponding locations are used as predictors. To improve the accuracy and stability, and to reduce the variance and overfitting of the RF predictions, they are generated as an ensemble estimate from the numerous decision trees (Díaz-Uriarte and Alvarez de Andrés 2006; Hengl et al. 2018) as observed in Eq. 3.10:

$$\hat{\theta}^B(x) = \frac{1}{B} \sum_{b=1}^B t_b^*(x) \quad (3.10)$$

where $\hat{\theta}^B$ is the final prediction; b is the individual bootstrap sample; B is the total number of trees; and t_b^* is the individual decision tree. This process is repeated for each time step, implying that the RF model will vary temporally (see Figure 2 of Baez-Villanueva et al. (2020) as an example).

Once the final merged product is generated, the validation set of rain gauge stations is used to assess the performance of the merged product through a point-to-pixel analysis (see Section 3.1). However, this method assumes that the ground-based measurements are representative values at their respective grid-cells, which may introduce bias in the comparison because *i*) during winter, some rain gauges located at high elevations are not able to incorporate snow into their measurements and *ii*) during summer, a more dense network of rain gauges is required to capture the spatial patterns of small-scale convective events. To evaluate the performance of the merged product the KGE' (Eq. 3.1) and its components (Eqs. 3.2, 3.3, and 3.4) were selected over the traditional RMSE because it gives disproportional weights to different precipitation intensities at the daily scale (Baez-Villanueva et al. 2018). This is due to the high skewness of the precipitation distribution at the daily scale and the prevalence of temporal mismatches between estimated and observed precipitation peaks.

Additionally, to evaluate the performance of precipitation products in capturing different precipitation intensities, four categorical indices of performance were used. The POD (Eq. 3.7), f_{bias} (Eq. 3.8), FAR (Eq. 3.9), and CSI (Eq. 3.11), which combines the POD and FAR to describe the overall ability of the products to correctly detect different precipitation intensities and has its optimal value at one.

$$\text{CSI} = [(\text{POD})^{-1} + (1 - \text{FAR})^{-1} - 1]^{-1} \quad (3.11)$$

3.2.2 Application of RF-MEP to continental Chile

RF-MEP was applied to the Chilean territory (see Section 2.2) from 17.5° to 46.0°S for 2000–2016. The southern boundary was not considered due to the sparse network of gauge stations in the Far South. Chile was selected as the case study to test the performance of the proposed RF-MEP due to the notable heterogeneity in topography, climate, and land cover. To apply RF-MEP to this case study the R environment 3.5.0 (R Core Team 2020) and the raster (Hijmans 2018), hydroGOF (Zambrano-Bigiarini 2017a), hydroTSM (Zambrano-Bigiarini 2017b), GSIF (Hengl 2019), and randomForest (Liaw and Wiener 2002) R packages were used.

As mentioned in Section 3.2.1, the first step is to acquire the ground-based measurements and the covariates. Therefore, time series of ground-based daily precipitation for 1900–2018 were downloaded from a database of 816 rain gauges from the Center of Climate and Resilience Research (CR2 2022). Daily precipitation is recorded at 08:00 local time (11:00–10:59 UTC) in Chile. Five global or quasi-global state-of-the-art precipitation products with at least 15 years of daily precipitation estimates were selected (Table 3.3) as covariates. These products were selected because *i*) RF-MEP can be transferred to any selected study area using the same products (or others) if ground-based data are available, and *ii*) the selected products perform well in the study area (Zambrano-Bigiarini et al. 2017; Zambrano-Bigiarini 2018).

Table 3.3: Precipitation products used in the case study, where S stands for satellite, R for reanalysis, and G for gauge data.

Product	Spatio-temporal res.	Source(s)	References
ERA-Interim	0.75°; 3-hourly	R	Dee et al. (2011)
CHIRPSv2*	0.05°; daily	S , G , and R	Funk et al. (2015)
TRMM 3B42v7*	0.25°; 3-hourly	S and G	Huffman et al. (2007, 2010)
PERSIANN-CDR*	0.25°; 6-hourly	S and G	Sorooshian et al. (2000); Ashouri et al. (2015)
CMORPHv1*	0.25°; 30 min	S and G	Joyce et al. (2004); Xie et al. (2017)
MSWEPv2.2*	0.10°; 3-hourly	S , G , and R	Beck et al. (2017a, 2019b)

* Products that use GPCC data.

CHAPTER 3. DATA AND METHODS

The selected precipitation products to be used in RF-MEP were: ERA-Interim (Dee et al. 2011); the Climate Hazards InfraRed Precipitation with Stations data version 2.0 (CHIRPSv2; Funk et al. 2015); the TRMM Multi-satellite Precipitation Analysis (TRMM 3B42v7; Huffman et al. 2007, 2010); the Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks - Climate Data Record (PERSIANN-CDR; Sorooshian et al. 2000; Ashouri et al. 2015); and the Climate Prediction Center (CPC) Morphing technique version 1.0-BLD, gauge-satellite blended precipitation product (CMORPHv1; Joyce et al. 2004; Xie et al. 2017). The Multi-Source Weighted-Ensemble Precipitation (MSWEPv2.2; Beck et al. 2017a, 2019b) was only used in the validation step as a benchmark product because *i*) it is the first fully global precipitation dataset derived by optimally merging a range of gauge, satellite, and reanalysis estimates (Beck et al. 2019b); *ii*) it has shown more realistic spatial precipitation patterns in mean, magnitude, and frequency than other state-of-the-art global precipitation products at the global scale (Beck et al. 2017b, 2019b); *iii*) it uses the same rain gauge dataset within Chile; and *iv*) it recently outperformed other state-of-the-art precipitation products over Chile (Zambrano-Bigiarini 2018). Detailed descriptions of the algorithms used by each precipitation product can be found in their corresponding literature (see Table 3.3).

It is important to note that several precipitation products use ground-based precipitation data from the Global Precipitation Climatology Centre (GPCC; Schneider et al. 2008) to reduce their bias (see Table 3.3). The number of operational GPCC rain gauge stations in Chile has fluctuated between seven and twenty over 1986–2018. This low density of GPCC stations within Chile is insufficient to represent the spatio-temporal variability of precipitation over the country adequately.

Additionally, the Shuttle Radar Topography Mission version 4 (SRTM-v4; Jarvis et al. 2008) digital surface model (DSM) was selected as a covariate. The SRTM-v4 has offsets due to vegetation height (Gallant et al. 2012), and a reported vertical error of less than 16 m (Jarvis et al. 2008). The gap-filled SRTM-v4 product was used at a spatial resolution of 250 m. Besides the selected precipitation products and the DSM, other spatial covariates (slope, aspect, Köppen-Geiger climate classification, and land cover type) were exhaustively evaluated using the KGE' and its components to ascertain whether an improvement could be obtained. Only the DSM was selected because the inclusion of the other covariates did not improve the final product's performance.

All selected precipitation products that are sub-daily (Table 3.3) were aggregated to the daily scale. MSWEPv2.2 was obtained at the daily temporal scale because the 3-hourly version was not freely available. PERSIANN-CDR, ERA-Interim, CMORPHv1, and TRMM 3B42v7 were downscaled to the same spatial resolution as

CHIRPSv2 (0.05°) using the nearest neighbour method (to avoid any improvements in the performance of the products before the merging procedure). At the same time, the DSM was upscaled from its original spatial resolution (250 m) to 0.05° using bilinear interpolation.

The reason for resampling all the covariates to 0.05° (the highest spatial resolution of the selected precipitation products) was to obtain a merged product that can be fairly compared to all selected precipitation products. A set of 369 rain gauge stations was selected. These stations had $< 5\%$ of missing values and showed consistency when evaluated using the double-mass curve method to identify abnormalities comparing each station with the neighbouring stations, assuming homogeneity (Weiss and Wilson 1953). The period 2000–2016 was chosen because of ground-based data availability over the period of record of the selected precipitation products. A random sample of 70% of the selected rain gauge stations (258) was used as ground truth data to train the RF model (training set), while the remaining 30% of the stations (111) were used to assess the performance of the merged products (validation set). Past studies have typically selected 80% or more stations for training purposes (e.g., Li and Shao 2010; Rozante et al. 2010; Woldemeskel et al. 2013; Yang et al. 2017; Ma et al. 2018); however, 70% of the stations was selected to be more thorough in the evaluation of the method. The 258 layers of Euclidean distances were computed using the GSIF R package (Hengl 2019).

Two merged precipitation products were computed at the daily scale for 2000–2016. The first product (hereafter, RF-MEP_{3P}) used CHIRPSv2, PERSIANN-CDR, ERA-Interim, the DSM, and the 258 layers of Euclidean distances, while the second product (hereafter, RF-MEP_{5P}) added CMORPHv1 and TRMM 3B42v7 to the aforementioned covariates. The reason for computing two different merged products was to evaluate whether the addition of CMORPHv1 and TRMM 3B42v7, both of which have a shorter period of temporal coverage, would improve the final merged product. Although RF-MEP_{3P} and RF-MEP_{5P} were produced and compared over the same period (2000–2016), RF-MEP_{3P} can be generated over a longer period of record (1983–2016), while RF-MEP_{5P} can only be generated from 1998 onwards.

First, the values of the covariates at the grid-cell locations of the training set were obtained. Second, for each day, a RF model was trained using the ground-based precipitation values as the dependent variable and the respective values from the covariates as predictors. Third, the trained RF model was used with the gridded covariates to predict daily precipitation values for each grid-cell of the study area. This process was repeated for each day for 2000–2016. RF regression models have three parameters to specify: *i*) the number of regression trees (set at 2000); *ii*) the number of variables randomly sampled at each decision split (set at one-third of the number

of covariates); and *iii*) the node size (i.e., the minimum number of observations per node; set at 5).

Finally, the performance of both merged products, MSWEPv2.2, and the individual precipitation products used as covariates was evaluated through a point-to-pixel analysis with the indices of performance described in the validation process of Section 3.2.1 using the independent stations included in the validation set. The evaluation process was performed at multiple temporal scales: 3-day, monthly, annual, DJF, MAM, JJA, and SON.

Because no sub-daily measurements are available to transform the ground-based precipitation dataset to the 0:00–23:59 UTC daily period used by all the precipitation products, a 3-day accumulations were used as a proxy for evaluating daily performance. This approach reduces likely biases in the performance of the precipitation products at this temporal scale by considering the influence of reporting times. The categorical indices were evaluated using the precipitation intensities (Table 3.2; Zambrano-Bigiarini et al. 2017), which are recommended specifically for Chile.

Because RF-MEP aims to improve the characterisation of precipitation in data-scarce regions, the influence of the amount of rain gauge stations included in the training set was investigated. The RF-MEP_{5P} product was computed with varying percentages of rain gauge stations in the training set to evaluate the performance of RF-MEP under different scenarios of data scarcity. In particular, the RF-MEP_{5P} product was computed using 50%, 30%, and 10% of the stations, representing 184, 111 and 37 rain gauges, respectively.

Also, to test the influence of the different spatial resolutions of the selected precipitation products, RF-MEP_{5P} was computed at 0.05°, 0.10°, and 0.25°. For this purpose, all covariates were resampled to these spatial resolutions before applying the merging procedure. Finally, two additional merging methods were computed to compare RF-MEP against established and proven precipitation merging procedures. Kriging with external drift (KED) was computed using ERA-Interim (the best-performing product used to derive RF-MEP_{5P}) and the one-outlier-removed (OOR) arithmetic mean described in Shen et al. (2014). For a detailed explanation of KED please refer to Ly et al. (2011), Oliver and Webster (2014), and Hengl et al. (2018). Also the RF-MEP_{5P} was compared against MSWEPv2.2 because it is a state-of-the-art merged precipitation product.

3.3 Regionalisation

3.3.1 Meteorological forcings

Four precipitation products were used to investigate how the choice of precipitation forcing affects the performance of regionalisation techniques over 100 Chilean catchments (used to predict streamflow in ungauged catchments): *i*) the Center for Climate and Resilience Research Meteorological dataset version 2.0 (CR2MET; Boisier et al. 2018); *ii*) the Random Forest MERging Procedure (RF-MEP; Baez-Villanueva et al. 2020) applied using the ERA5 reanalysis (Hersbach et al. 2020) aggregated to the Chilean time, and elevation (SRTMv4.1; Jarvis et al. 2008) as covariates; *iii*) ERA5 (Hersbach et al. 2020); and *iv*) The Multi-Source Weighted-Ensemble Precipitation (MSWEPv2.8; Beck et al. 2017b, 2019b). The precipitation products are presented in Table 3.4, and were selected because previous studies have reported good agreement when evaluated against ground-based measurements over continental Chile (Zambrano-Bigiarini et al. 2017; Boisier et al. 2018; Baez-Villanueva et al. 2018, 2020). All products were aggregated to the Chilean reporting times and a more detailed description of these products can be found in Baez-Villanueva et al. (2021).

Table 3.4: Gridded precipitation products used in this study.

Product	Spatio-temporal res.	Availability	References
CR2MET	0.05°; daily	1979–2018	Boisier et al. (2018)
RF-MEP	0.05°; daily	1983–2018	Baez-Villanueva et al. (2020)
ERA5	~0.28°; hourly	1950–present	Hersbach et al. (2020)
MSWEPv2.8	0.10°; 3-hourly	1979–present	Beck et al. (2017b, 2019b)

All precipitation products show relatively similar patterns of spatial variability across continental Chile; however, there are substantial differences in their total precipitation amounts. However, despite their similar precipitation patterns, these products show marked differences in mean annual precipitation, days without rain, frequency and intermittency of precipitation, high precipitation intensities, and seasonal distribution over the selected catchments. For a more detailed analysis on the difference of these products, the readers are referred to Baez-Villanueva et al. (2021), especially their Section 3.1.1 and their Figures 2, 3, and S2–S7 (see Appendix C).

Finally, the maximum and minimum daily air temperature at a spatial resolution of 0.05° were taken from CR2METv2. Temperature is estimated using multivariate regression from the Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature (LST) and ERA5 estimates as covariates (Alvarez-Garreton et al.

2018; Boisier et al. 2018). The Hargreaves-Samani equation (Hargreaves and Samani 1985) was used to obtain daily potential evaporation from CR2METv2 maximum and minimum daily temperature at the same spatial resolution (0.05°).

3.3.2 Hydrological model

The TUWmodel (Viglione and Parajka 2020) is a conceptual hydrological model that follows the structure of the Hydrologiska Byråns Vattenbalansavdelning (HBV) model (Bergström 1976; Bergström 1995; Lindström 1997). The model simulates the catchment-scale water balance at daily steps, including snow accumulation and melting processes, change of moisture in the soil profile, and surface flow in the drainage network. The TUWmodel was validated over 320 catchments in Austria (Parajka et al. 2007) and has subsequently been used in numerous studies (e.g., Parajka et al. 2016; Zessner et al. 2017; Melsen et al. 2018; Sleziaak et al. 2020). A HBV-like conceptual model was selected because it has shown good results in *i*) many regionalisation studies (e.g., Parajka et al. 2005; Singh et al. 2012; Beck et al. 2016; Neri et al. 2020); and *ii*) catchments with diverse hydroclimatic and geomorphological characteristics (Vetter et al. 2015; Ding et al. 2016; Unduche et al. 2018; Huang et al. 2019).

The TUWmodel requires daily time series of precipitation, temperature, and potential evaporation. The parameters used by the TUWmodel to represent the hydrological processes are listed in Table 3.5, including the ranges selected for model calibration, which were adopted from previous studies (Parajka et al. 2007; Ceola et al. 2015) that calibrated the TUWmodel over a large number of mountainous catchments with snow influence. TUWmodel was run with a semi-distributed configuration for the period 1990–2018 based on meteorological and streamflow data availability. Although using a fully distributed model could provide useful insights related to the spatial distribution of key hydrometeorological variables, the TUWmodel was selected because of its relatively easy implementation and computational requirements as 100 catchments were used and four precipitation products evaluated.

For each catchment, the number of equal-area elevation bands (EZ) was defined as $EZ = (H_{max} - H_{min})/200$, where H represents elevation. In cases where $EZ > 10$, EZ was set to 10 to reduce the computational demand of the simulations. Furthermore, in catchments with H_{min} below 900 m a.s.l., the upper bound of the first EZ band was set to 900 m under the assumption that there is no snow influence below this elevation for the particular case of continental Chile. For more details about the TUWmodel implementation in R and the comparison of different HBV-like models, the readers are referred to Astagneau et al. (2021), and Jansen et al. (2021), respectively.

Table 3.5: Summary of the TUWmodel parameters considered for calibration, following the conceptualisation presented in Széles et al. (2020).

N°	Parameter ID	Description	Units	Process	Range
1	SCF	Snow correction factor	–	Snow	0.9 – 1.5
2	DDF	Degree-day factor	mm °C day ⁻¹	Snow	0.0 – 5.0
3	Twb	Wet bulb temperature	°C	Snow	-3.0 – 3.0
4	Tm	Threshold temperature above which melt starts	°C	Snow	-2.0 – 2.0
5	LPrat	Parameter related to the limit for potential evaporation	–	Evaporation	0.0 – 1.0
6	FC	Field capacity	mm	Infiltration	0.0 – 600
7	Beta	Non-linear parameter for runoff production	–	Infiltration	0.0 – 20
8	cperc	Constant percolation rate	mm day ⁻¹	Infiltration	0.0 – 8.0
9	k0	Storage coefficient for very fast response	day	Runoff	0.0 – 2.0
10	k1	Storage coefficient for fast response	day	Runoff	2.0 – 30
11	k2	Storage coefficient for slow response	day	Runoff	30 – 250
12	lsuz	Threshold storage state	mm	Runoff	1.0 – 100
13	bmax	Maximum base at low flows	day	Runoff	0.0 – 30
14	croute	Free scaling parameter	day ² mm ⁻¹	Runoff	0.0 – 50

3.3.3 Independent catchment calibration and verification

The simulation period used for this study was 1990–2018. For calibration purposes, the first ten years as a conservative warm-up period to initialise the model stores was used, as in Beck et al. (2020a). The calibration period (2000–2014) includes near-normal conditions and the beginning of the Chilean megadrought. The first evaluation period (hereafter, Verification 1, 1990–1999) represents near-normal/wet hydroclimatic conditions, while the second evaluation period (hereafter, Verification 2, 2015–2018) spans the second half of the Chilean megadrought and was used to test the ability of the hydrological simulations to represent dry conditions. To initialise model stores for the Verification 1 period, A 8-year warm-up period was used due to precipitation product availability.

The KGE' (Eq. 3.1) was used to calibrate the TUWmodel, which typically provides better hydrograph simulations than other squared-error indices (Gupta et al. 2009; Kling et al. 2012; Mizukami et al. 2019) and has been used in numerous studies (e.g., Garcia et al. 2017; Beck et al. 2019b; Baez-Villanueva et al. 2020; Neri et al. 2020; Széles et al. 2020). To calibrate the model parameters, the hydroPSO global optimisation algorithm was used (Zambrano-Bigiarini and Rojas 2013), which implements a state-of-the-art version of the Particle Swarm Optimisation technique (PSO; Eberhart and Kennedy 1995; Kennedy and Eberhart 1995). For this purpose, the standard PSO 2011 algorithm (Clerc 2011a,b), defined as *spso2011* in the hydroPSO R package (Zambrano-Bigiarini and Rojas 2013) was used. The number of particles in the swarm (*npart*) was set to 80, the maximum number of iterations (*maxit*) to 100, and the relative convergence tolerance (*reltol*) to $1E - 10$, while the default values were used for all other parameters. Over the last decade, hydroPSO has been

successfully used to calibrate numerous hydrological and environmental models (e.g., Brauer et al. 2014; Silal et al. 2015; Bisselink et al. 2016; Kundu et al. 2017; Kearney and Maino 2018; Abdelaziz et al. 2019; Ollivier et al. 2020; Hann et al. 2021). For more details on the use of the hydroPSO package to calibrate the TUWmodel, readers are referred to Zambrano-Bigiarini and Baez-Villanueva (2020).

3.3.4 Regionalisation techniques

After obtaining catchment-specific model parameters through independent catchment calibration (Section 3.3.3), three parameter regionalisation techniques were compared: *i*) spatial proximity; *ii*) feature similarity; and *iii*) parameter regression. The performance was assessed through a leave-one-out cross-validation exercise, which consisted of leaving out each of the 100 catchments, transferring model parameters, conducting streamflow simulations, and computing performance evaluation metrics.

Spatial proximity

The spatial proximity method assumes that climatic and physical characteristics are relatively homogeneous over a region (Oudin et al. 2008). The spatial proximity between the target pseudo-ungauged and the remaining catchments was quantified using the Euclidean distance between catchment centroids, computed with geographic coordinates (i.e., latitude and longitude):

$$ED_{ij} = \sqrt{\sum_{k=1}^n (x_{k,i} - x_{k,j})^2} \quad (3.12)$$

For each pseudo-ungauged catchment, the donor was chosen according to the minimum Euclidean distance, and the full parameter set obtained during the independent calibration of the donor catchment was transferred to the pseudo-ungauged catchment.

Feature similarity

In the feature similarity method, the calibrated parameter sets from 10 donor catchments were transferred to the pseudo-ungauged catchment based on the similarity between climatic and geomorphological features, quantified using the catchment characteristics presented in Table 3.6. To exclude redundant information, correlation analyses were performed between catchment descriptors using the Pearson and Spearman rank correlation coefficients (to account for linear and monotonic correlation, respectively). After these analyses, three descriptors with high correlations were discarded (mean elevation, mean annual potential evaporation, and SDII; see Appendix C). Also, the snow cover was discarded because it was found to be unreliable, leaving nine catchment features for this method. To assign equal weight to each catchment characteristic, they were normalised into the range $[0, 1]$ using Eq. 3.13:

$$Z_f = \frac{x_f - x_{min}}{x_{max} - x_{min}} \quad (3.13)$$

where x_f is the value of the characteristic for catchment f , while x_{max} and x_{min} are the maximum and minimum values of the characteristic x over all catchments. After normalising all catchment characteristics, the dissimilarity was calculated as follows:

$$S_{i,j} = \sum_{m=1}^n |Z_{i,m} - Z_{j,m}| \quad (3.14)$$

where $S_{i,j}$ is the dissimilarity index between catchments i and j ; $Z_{i,m}$ and $Z_{j,m}$ are the normalised values of the m catchment characteristic for catchments i and j , respectively; and n is the total number of characteristics.

For each pseudo-ungauged catchment i , the 10 catchments j with the lowest dissimilarity indices ($S_{i,j}$) were selected as donors (Oudin et al. 2008; Zhang and Chiew 2009; Zhang et al. 2015; Beck et al. 2016). The full parameter sets obtained during the independent calibration of each donor catchment were used to run the TUWmodel in the pseudo-ungauged catchment, thus producing an ensemble of 10 streamflow simulations, as in previous studies (McIntyre et al. 2005; Zelelew and Alfredsen 2014; Beck et al. 2016). The ten streamflow time series were then averaged to produce a single streamflow time series.

Table 3.6: Selected climatic and physiographic characteristics to quantify feature similarity between catchments. All variables related to precipitation were computed using the corresponding precipitation product used as an input to the TUWmodel for 1990–2018.

N°	Variable	Data source	Importance
1	Mean elevation	CAMELS-CL	Composite indicator that influences a range of processes such as long-term precipitation and temperature, and hence soil moisture availability. In some environments, it is also related to aridity and snow processes.
2	Median elevation	SRTMv4.1	Same as mean elevation but provides a more robust representation of elevation over mountainous catchments.
3	Catchment area	CAMELS-CL	Related to the degree of aggregation of catchment processes related to scale effects. Additionally, it is an indicator of total catchment storage capacity.
4	Slope	CAMELS-CL	Related to the response of the catchment, routing, and infiltration processes.
5	Forest cover	CAMELS-CL	Forested catchments are associated with a trade-off between high water consumption rates and enhanced soil.
6	Snow cover	CAMELS-CL	Related to the influence of snow processes within the catchment.
7	Mean annual precipitation	Precipitation product	Related to the generation of runoff and precipitation related to orographic gradients (e.g., coastal areas).
8	Mean annual air temperature	CR2MET	Indicator of snow processes in cold environments. It is also related to aridity and, consequently, to the evaporative demand.
9	Mean annual potential evaporation	Computed from CR2MET	A measure of the atmospheric water demand (especially at the annual temporal scale).
10	Aridity index	CR2MET and precipitation product	Represents the competition between energy and water availability.
11	Daily temperature range	CR2MET	Monthly mean difference between daily maximum and minimum temperature. Related to variations in the diurnal cycle and evaporative demands.
12	Simple precipitation intensity index	Precipitation product	Relation of annual precipitation to the number of wet days (precipitation > 1 mm). Serves as a proxy for seasonality and intensity of precipitation events.
13	Maximum consecutive 5-day precipitation	Precipitation product	Related to extreme precipitation events.

Parameter regression

The parameter regression technique aims to detect statistical relationships between parameter values and catchment characteristics and uses these relationships to estimate model parameters for ungauged catchments (Parajka et al. 2005; Oudin et al. 2008; Swain and Patra 2017). To account for non-linear relationships between model parameters and catchment characteristics, the random forest machine learning algorithm (RF; Breiman 2001; Prasad et al. 2006; Biau and Scornet 2016), which is provided in the RandomForest R package (Liaw and Wiener 2002) was implemented. RF uses an ensemble of decision trees between predictand and predictor values (also known as covariates) for regression and supervised classification and has the capability to deal with high-dimensional feature spaces and small sample sizes (Biau and Scornet 2016).

Previous studies have shown that RF can deal with several covariates as well as non-informative predictors because it does not lead to overfitting or biased estimates (Díaz-Uriarte and Alvarez de Andrés 2006; Biau and Scornet 2016; Hengl et al. 2018), which is why it has been used for numerous hydrological applications (Saadi et al. 2019; Baez-Villanueva et al. 2020; Beck et al. 2020b; Zhang et al. 2021). For a more detailed description of RF, the reader is referred to Prasad et al. (2006), Biau and Scornet (2016), and Addor et al. (2018).

For this study, one RF model for each TUWmodel parameter was developed, using all thirteen independent catchment characteristics listed in Table 3.6 as covariates. The experimental setup used an ensemble of 2,000 regression trees, a minimum of five terminal nodes for each model, and $p/3$ variables randomly sampled as candidates at each split, where p represents the number of predictors. The trained RF models were then used to predict parameter values in the pseudo-ungauged catchments.

3.3.5 Influence of nested catchments

To evaluate the influence of nested catchments on the performance of the three regionalisation methods, they were repeated for each target catchment, with catchments considered to be nested (in relation to the pseudo-ungauged catchment) excluded from the set of potential donors. Following Neri et al. (2020), a cutoff point of 10% of drainage area was used, meaning that only catchments that cover more than 10% of the area of the parent catchment were considered to be nested.

3.3.6 Influence of donor catchments for feature similarity

To evaluate the influence of the number of donor catchments used in feature similarity, the process followed in Section 3.3.4 was repeated to assess the performance of this regionalisation method when 1, 2, 4, 6, 8, and 10 donor catchments are selected. This analysis evaluates the impact of averaging varying numbers of simulations compared to the results that are based on only the most similar catchment. All analyses were performed using the R Project of Statistical Computing (R Core Team 2020). In addition to the R packages described in the methodology, the hydroGOF (Zambrano-Bigiarini 2020a), hydroTSM (Zambrano-Bigiarini 2020b), lfstat (Koffler et al. 2016), raster (Hijmans 2020), rasterVis (Perpiñán and Hijmans 2020), rgdal (Bivand et al. 2020), and rgeos (Bivand and Rundel 2020) packages were used.

Chapter 4

Results

4.1 Evaluation

The aim of the evaluation step was: *i*) to evaluate for the very first time the spatio-temporal performance of six state-of-the-art precipitation products (TRMM 3B42RT, TRMM 3B42v7, CMORPHv1, CHIRPSv2, PERSIANN-CDR, and MSWEPv2) over different areas in Latin America and at different temporal scales (daily, monthly, and seasonal) using 201 rain gauge stations in total and *ii*) to assess whether the upscaling procedure used to enable a consistent point-to-pixel comparison affects the evaluation of the upscaled products (in the case of CHIRPSv2 and MSWEPv2).

4.1.1 Performance of the evaluated precipitation products at different temporal scales

The evaluation of the selected precipitation products was performed at the daily, monthly, and seasonal temporal scales. The seasonal temporal scales were divided as follows: DJF (December-January-February), MAM (March-April-May), JJA (June-July-August), and SON (September-October-November). As expected, the precipitation products performed differently over each study area and results using the KGE' and RMSE differ. Table 4.1 summarises the results obtained in the evaluation with the KGE' and RMSE.

Table 4.1: Results of the evaluation of performance using the KGE' and the RMSE over the study areas.

Study area	Temporal scale		Seasonal
	Daily	Monthly	
Imperial	<p>The evaluation at the daily scale showed that MSWEPv2 performed the best over the Imperial followed by PERSIANN-CDR, MSWEPv2 upscaled, and CHIRPSv2 [0.4 < KGE' < 0.5] when evaluated with the KGE', while TRMM 3B47RT had the lowest performance followed by CMORPHv1 [0.1 < KGE' < 0.2]. PERSIANN-CDR performed the best with the RMSE (7.66 mm) followed by TRMM 3B42RT (7.98 mm). In general, all products performed similarly at different elevations (Baez-Villanueva et al. 2018, their Figure 5).</p>	<p>CHIRPSv2 performed the best with the KGE', followed by TRMM 3B42v7 [0.75 < KGE' < 0.80]. Almost all the products scored KGE' values higher than 0.7 except for TRMM 3B42RT, which scored a median value of 0.3. For the RMSE, MSWEPv2 and MSWEPv2 upscaled were the best-evaluated products, presenting the lowest precipitation values (38.90 mm and 42.17 mm, respectively). MSWEPv2 and MSWEPv2 upscaled occupied the fifth and sixth performance position when evaluated with the KGE'.</p>	<p>CMORPHv1 and TRMM 3B42v7 presented a good performance for DJF and SON when the KGE' was used, while TRMM 3B42v7 and MSWEPv2 performed better for MAM and PERSIANN-CDR for JJA. TRMM 3B42RT presented the lowest performance over this catchment (for the KGE' and RMSE) during all seasons. All products had a lower performance during the rainy season (JJA; winter).</p>
Paraiba do Sul	<p>MSWEPv2 performed the best with the KGE', followed by CMORPHv1 and MSWEPv2 upscaled [0.55 < KGE' < 0.65]. CHIRPSv2 and CHIRPSv2 upscaled obtained a median KGE' of 0.37. PERSIANN-CDR performed the worst, followed by TRMM 3B42v7, and TRMM 3B42RT, with KGE' values < 0.3. CMORPHv1 yielded the best performance using the RMSE, followed closely by MSWEPv2. However, CHIRPSv2 performed worse than PERSIANN-CDR with the RMSE. The performance of the products was slightly lower in elevated areas (Baez-Villanueva et al. 2018, their Figure 6).</p>	<p>Paraiba do Sul presented the highest KGE' values compared to the Imperial and Magdalena. The best performing product was MSWEPv2 upscaled, followed by CHIRPSv2 with KGE' values of 0.83 and 0.82, respectively. MSWEPv2, CHIRPSv2 upscaled, and CMORPHv1 presented higher KGE' values than 0.80. For both indices (KGE' and RMSE), TRMM 3B42RT, TRMM 3B42v7, and PERSIANN-CDR presented the lowest performance.</p>	<p>Paraiba do Sul presented the highest KGE' values for the selected products compared to the Imperial and Magdalena. MSWEPv2 and MSWEPv2 upscaled performed the best for all seasons for the KGE' and RMSE, followed by CHIRPSv2 and CMORPHv1. The products with the lowest performance for all seasons were PERSIANN-CDR and TRMM 3B42RT. All the datasets performed similarly over the four different seasons.</p>
Magdalena	<p>CHIRPSv2 performed the best with the KGE', followed by CMORPHv1 and CHIRPSv2 upscaled [0.25 < KGE' < 0.30]. The daily performance of all products was lower over this study area compared to the Imperial and Paraiba do Sul. PERSIANN-CDR and MSWEPv2 presented the lowest performance [0.0 < KGE' < 0.1], which is in agreement with Dinku et al. (2010), although they evaluated PERSIANN instead. All products presented low KGE' values in the lower regions, located in the northern part of the catchment.</p>	<p>The products presented KGE' values ranging from 0.3 to 0.7 (Baez-Villanueva et al. 2018, their Figure 7). CHIRPSv2 and CHIRPSv2 upscaled presented the highest KGE' values (0.70 and 0.66, respectively). PERSIANN-CDR showed the lowest performance for both indices. However, MSWEPv2, MSWEPv2 upscaled, and TRMM 3B42RT presented a higher performance when evaluated with the KGE' compared to the RMSE.</p>	<p>CHIRPSv2, CHIRPSv2 upscaled, and TRMM 3B42RT performed the best for DJF and SON using the KGE' and RMSE. These products also performed the best for MAM when the KGE' was used, while CHIRPSv2, CHIRPSv2 upscaled, and TRMM 3B42v7 performed the best when the RMSE was used. For SON, CHIRPSv2 and CHIRPSv2 upscaled showed a good performance using both indices. The products presented a lower performance for MAM and JJA (spring and summer, respectively) over this catchment.</p>

CHAPTER 4. RESULTS

The total performance of the KGE' can be decomposed into a linear correlation (r), a bias (β), and a variability (γ) term. It is a useful index because r can evaluate the temporal dynamics of precipitation, while β and γ can evaluate the volume and variability of rainfall, respectively.

Over the Imperial, the r component showed the lowest values in the rainy season (JJA). MSWEPv2 presented the highest r for all seasons, while CHIRPSv2 and CHIRPSv2 upscaled were the most unbiased products for all seasons. CMORPHv1 was almost unbiased in DJF and SON and presented the best performance for γ in DJF and MAM, while CHIRPSv2 and CHIRPSv2 upscaled in JJA and SON. For Paraiba do Sul, the components of the KGE' performed similarly for the different seasons. MSWEPv2 had the best r and γ for all the seasons, while CHIRPSv2, CHIRPSv2 upscaled, CMORPHv1, MSWEPv2 and MSWEPv2 upscaled presented good β values. Over the Magdalena, r showed the lowest values in JJA, β performed worse in DJF and SON, and γ showed the lowest values during MAM and JJA (Baez-Villanueva et al. 2018, their Figure 8). CHIRPSv2 and CHIRPSv2 upscaled performed the best in all seasons, followed by TRMM 3B42RT during DJF, MAM, and SON, and TRMM 3B42v7 during JJA. CHIRPSv2 and CHIRPSv2 upscaled presented higher values of r and β during all the seasons. In contrast, MSWEPv2, MSWEPv2 upscaled, and CMORPHv1 presented median values near 1 for their component for seasons, meaning that these products are able to capture the distribution of precipitation over this catchment. However, MSWEPv2 present a low performance over the Magdalena.

The results of the PBIAS are in total agreement with those of the β component (Baez-Villanueva et al. 2018, their Figure 9) despite PBIAS being represented as a percentage and β as the ratio between the correspondent precipitation product mean and the ground-based mean (Eq. 3.3). For Paraiba do Sul and Magdalena catchments, almost all products overestimated the precipitation compared to the ground-based measurements, except for CMORPHv1, which underestimated the total precipitation amount. On the other hand, for the Imperial, the products tended to underestimate precipitation, except for CHIRPSv2 and CHIRPSv2 upscaled.

Additionally of the evaluation with continuous indices, it is important to assess the ability of the precipitation products to detect single rain events (see Table 3.2). For this reason, three categorical indices were applied as described in Section 3.1. All products showed a relatively high POD in no rain events ($[0, 1)$ mm d⁻¹) with values higher than 0.6, except for PERSIANN-CDR, which showed a low POD over the Magdalena (Baez-Villanueva et al. 2018, their Figure 12). All products performed better at capturing the moderate precipitation intensities (between 5 and 20 mm d⁻¹) for all cases. For light precipitation intensities ($[1, 5)$ mm d⁻¹) almost all products

presented a decrease in the POD, with values lower than 0.4 over all study areas (except for MSWEPv2 and CMORPHv1 in Paraiba do Sul), showing that light rain events are still difficult to capture by the current products.

For rain intensities higher or equal to 1, PERSIANN-CDR, CHIRPSv2, and MSWEPv2 presented a better performance than the other products. The FAR results were consistent with those of the POD. The no rain events showed low values over the different catchments, while the events with precipitation presented values between 0.5 and 0.95 in all cases. This shows that for the analysed products, the days with precipitation were more difficult to identify. The violent precipitation events presented relatively low f_{bias} in all regions and the f_{bias} showed an excellent agreement with the POD for the no rain events, with values of $f_{bias} \sim 1$. The specific results for each catchment are included in Table 4.2.

Our results confirm that there is no single best performing product for all regions. Therefore, a site-specific evaluation is still recommended to identify the product that best represents precipitation's spatio-temporal characteristics over a specific study area. These results show that gauge-adjusted algorithms tend to perform better, in general, compared to those without ground-based adjustments. In some cases, an independent validation dataset is difficult to obtain. Therefore, the independence of the ground-based dataset should be analysed before any comparison among products due that this may affect the results of the evaluation. Moreover, the performance of the products increases when the daily values are aggregated into monthly or seasonal values.

4.1.2 Does the upscaling process affect the performance of the products?

Upscaled versions of CHIRPSv2 and MSWEPv2 were computed from 0.5° and 0.1° , respectively, to ensure a consistent point-to-pixel comparison with the other products that have a spatial resolution of 0.25° . However, these datasets were also analysed using their original relatively high spatial resolution to evaluate if the applied upscaling procedure interfered with the performance evaluation.

This is the first time that CHIRPSv2 and MSWEPv2 are evaluated at different spatial resolutions over the same study areas. For the Imperial, the upscaling procedure improved the performance of CHIRPSv2 and decreased the performance of MSWEPv2 at the daily scale (Baez-Villanueva et al. 2018, their Figure 13). The opposite was observed at the monthly temporal scale, where CHIRPSv2 upscaled presented a lower performance than CHIRPSv2, and MSWEPv2 upscaled performed

CHAPTER 4. RESULTS

Table 4.2: Results of the evaluation of performance over the study areas using the categorical indices.

Study Area	Main Results
Imperial	There was no single product that performed the best for all precipitation intensities. CHIRPSv2 upscaled, MSWEPv2, and MSWEPv2 upscaled performed relatively better at capturing high precipitation intensities, which is in agreement with Zambrano-Bigiarini et al. (2017). MSWEPv2, MSWEPv2 upscaled, and TRMM 3B42RT showed low FAR values. All products except CHIRPSv2 and CHIRPSv2 upscaled overestimated the light rain events when evaluated with the f_{bias} . CMORPHv1, PERSIANN-CDR, TRMM 3B42RT, and MSWEPv2 underestimated violent rain events. Only PERSIANN-CDR showed a slight overestimation of moderate events ($[5, 20)$ mm d^{-1}). Finally, the Imperial presented the second-lowest f_{bias} values.
Paraíba do Sul	MSWEPv2 performed the best for all precipitation intensities when evaluated with all categorical indices. All products except CHIRPSv2 and CHIRPSv2 upscaled overestimated the light rain events. In general, the moderate events ($[5, 20)$ mm d^{-1}) were overestimated. TRMM 3B42RT, and TRMM 3B42v7 overestimated the violent rain events, which is in agreement with Melo et al. (2015). Finally, the Paraíba do Sul presented the lowest f_{bias} values.
Magdalena	All products presented almost no skill at capturing high precipitation intensities. There was not a single product that performed the best for all intensities. All products except PERSIANN-CDR presented a high POD for the no rain events. CMORPHv1, CHIRPSv2, and CHIRPSv2 upscaled performed better for high rain intensities. The moderate events ($[5, 20)$ mm d^{-1}) were in general overestimated. All products showed a high overestimation of light rain events. Almost all products underestimated violent rain events except for MSWEPv2 and MSWEPv2 upscaled. Finally, the Magdalena presented the highest f_{bias} values for light intensities.

better than the original MSWEPv2. In the case of Paraíba do Sul, CHIRPSv2 performed better after the upscaling procedure, while MSWEPv2 was not affected when evaluated at the daily scale. At the monthly scale, CHIRPSv2 upscaled performed slightly better than CHIRPSv2, while MSWEPv2 upscaled had lower performance. Finally, in the Magdalena, the upscaling procedure did not impact the performance at the daily scale for the two products; however, at the monthly scale, CHIRPSv2 upscaled performed worse than the original product, while no changes were observed

for MSWEPv2.

Our results show that there is not a general conclusion about the impact of the upscaling procedure on the performance of a given dataset, with different results regarding products, temporal scales, and regions (see Baez-Villanueva et al. 2018, their Figure 13). However, the topography plays an important role when an upscaling procedure is applied to a particular product. The stations with higher differences in performance when compared to ground-based measurements before and after the upscaling procedure were located closer to pronounced elevation gradients. For this reason, if an evaluation of the performance of precipitation products is required in a mountainous area, the use of an upscaling procedure to enable a consistent point-to-pixel comparison between products may affect the evaluation. On the other hand, if there is not a pronounced elevation gradient, an upscaling procedure can enable a fair point-to-pixel comparison.

4.1.3 Lessons learned from the evaluation of precipitation products

The following points summarise the lessons learned from the results of the evaluation of precipitation products over three Latin American catchments (see Sections 4.1.1 and 4.1.2):

1. the KGE' has proved to be a useful measure of performance as it decomposes the performance of the products into r , β , and γ . For this reason, it is easier to understand the source(s) of the mismatches between the precipitation products and their corresponding ground-based measurements;
2. the POD, FAR, and f bias are important indices to evaluate the accuracy of the products at identifying different precipitation intensities;
3. as observed, the RMSE and KGE' can present different results when used to evaluate the spatio-temporal performance of different precipitation products. However, the KGE' has proven to be an improved performance index due to its ability to not penalise mismatches of heavy precipitation events and its ability to decompose the total performance into r , β , and γ .
4. each evaluated product performed differently for each catchment and each temporal scale. On the daily scale, MSWEPv2 performed the best over the Imperial and Paraiba do Sul, while CHIRPSv2 performed better over the Magdalena. CHIRPSv2 presented the highest performance on the monthly scale

over the Imperial and Magdalena, while MSWEPv2 performed the best over the Paraiba do Sul. When the catchments were evaluated at the seasonal scale, CMORPHv1 performed the best for DJF and SON, TRMM 3B42v7 for MAM, and PERSIANN-CDR for JJA over the Imperial. MSWEPv2 performed the best over Paraiba do Sul during all seasons, while CHIRPSv2 showed the best performance over the Magdalena. Finally, it is worth mentioning that the gauge-adjusted products generally tend to perform better than those with no ground-based adjustments. For this reason, the independence of the validation dataset must be taken into account before any comparison among products due that it may affect the results of the evaluation;

5. the highest probability of detection ($POD \sim 1$) was obtained for the no rain intensities for all products over all the study areas. In general, all products presented a low POD for high precipitation events. The moderate rain events (from 5 to 20 mm d^{-1}) were the best captured in all catchments when precipitation was larger than 1 mm d^{-1} . In particular, MSWEPv2 performed considerably better for all precipitation intensities in Paraiba do Sul than the rest of the products. The f_{bias} presented higher variations for the violent (higher than 40 mm) and the light rain events ([1, 5) mm d^{-1}), which is in agreement with the results of the POD. The FAR results showed consistency with those obtained with the POD, showing low values for no rain events;
6. despite the evolution of the precipitation products, our results confirm that a catchment-specific validation is still required to select a suitable product for hydrological purposes. This evaluation results cannot be extrapolated to other catchments and regions; this comparison shows that the same products present different behaviour over different areas. The readers are invited to evaluate the performance of any precipitation product before any application;
7. the RMSE gives more weight to the mismatches in two cases: *i*) when a low precipitation intensity is constantly not detected, and *ii*) in high precipitation events because it gives more weight to the mismatches. Therefore, the RMSE is not recommended for evaluating the performance of precipitation products at the daily scale. Additionally, its results are not comparable between areas with different precipitation regimes; and
8. the upscaling procedure can affect the performance of the precipitation products, and it may vary between products, temporal scales, and regions. If an upscaling procedure is performed while evaluating the performance of precipitation products over an area, the selection of the best performing product can be affected by it. Evaluating the products' performance is recommended before

and after an upscaling procedure is computed to select the most representative product of the spatio-temporal precipitation patterns of an area. Also, topography plays an important role when upscaling a product. The stations with higher performance differences between the original products and the upscaled ones were close to pronounced elevation gradients. For this reason, if an evaluation of different precipitation products is required over a mountainous area, the upscaling procedure may affect the results. However, if the topography is not rugged, an upscaling procedure can enable a fair point-to-pixel comparison among products.

4.2 Merging

RF-MEP was applied to derive two merged products: *i*) RF-MEP_{3P}, using CHIRPSv2, PERSIANN-CDR, ERA-Interim, the DSM, and the 258 layers of Euclidean distances; and *ii*) RF-MEP_{5P}, which included CMORPHv1 and TRMM 3B42v7 to the aforementioned covariates.

4.2.1 Spatio-temporal performance of the merged products

Together with the products used in their computation, the two merged products and the benchmark MSWEPv2.2 were evaluated across Chile at seven temporal scales (3-day, monthly, annual and the four seasons). The summarised results that appear in Baez-Villanueva et al. (2020) are presented in Table 4.3.

RF-MEP was able to improve the spatio-temporal representation of precipitation (Baez-Villanueva et al. 2020, their Figures 4–8) by combining multiple sources of information. Both merged products showed increased r , β , and γ values at all temporal scales, which indicates that this method is able to represent the total volume and distribution of precipitation by providing a better representation of daily precipitation patterns. Comparable improvements in β were obtained by Manz et al. (2016) and Yang et al. (2017), although Ma et al. (2018) reported a higher bias in their merged product. Also, the reduction in the dispersion of the KGE' and its components demonstrates that the merged products show good performance over most of the study area. The KGE' has proven to be a useful performance index because of its ability to decompose the performance into r , β , and γ , which can be used to understand the different sources of mismatches.

CHAPTER 4. RESULTS

The evaluated precipitation products showed higher performances at the monthly, seasonal and annual scales in comparison to shorter temporal scales (Baez-Villanueva et al. 2020, their Figure 4), similar to the results reported by Jiang et al. (2012) and Zambrano-Bigiarini et al. (2017). This indicates that despite systematic, random, and detection errors present in the precipitation products at the daily scale, they can still represent precipitation patterns when aggregated at longer temporal scales. On the other hand, Maggioni and Massari (2018) concluded that spatial sampling uncertainties tend to decrease for higher temporal resolutions, which means that the point-to-pixel evaluation tends to be more reliable for increasing accumulation periods.

All products showed the lowest performance in summer (DJF), which is consistent with the results obtained by Rabiei and Haberlandt (2015) and Zambrano-Bigiarini et al. (2017). This could be because *i*) small-scale convective precipitation events dominate in summer in the Far North region (Prein and Gobiet 2017); *ii*) in warm months, the evaporation of hydrometeors before they reach the ground leads to overestimation and false alarms (Maggioni and Massari 2018); and *iii*) passive microwave radiometers overestimate and underestimate precipitation during summer and winter, respectively (Tang et al. 2014).

Both merged products presented their lowest performance over the arid Far North region due to the low performance of all products used as covariates. This is in agreement with Manz et al. (2016), where the merged products presented high uncertainty and low performances predominantly over regions with low and intermittent precipitation regimes. The mismatches of the precipitation products are more evident in arid and semi-arid climates because any overestimation or underestimation will have a greater impact on the performance evaluation over low precipitation regimes. Despite this, the RF-MEP_{5P} and RF-MEP_{3P} products were able to adequately represent the precipitation patterns of the higher elevations of the Far North, showing that RF-MEP is able to improve the spatio-temporal estimation of precipitation through the inclusion of complementary information, even in regions where the selected products exhibit low performance.

The time difference that the products have with respect to the rain gauge stations (~ 11 hour difference; for discussion, see Beck et al. 2019b) must be considered for the evaluation of the precipitation products at the daily temporal scale. Among the evaluated products, only MSWEPv2.2 incorporates daily gauge data and applies corrections to account for the reporting times of the rain gauges.

Table 4.3: Results of the assessment of the temporal, spatial, and categorical performance of the merged products.

Type of assessment	Main Results
Temporal	<p>Both merged products performed similarly well, with median KGE' values of 0.83, 0.84 and 0.78 at the 3-day, monthly, and annual scale, respectively. The precipitation products used in the merging method presented median KGE' values ≤ 0.60 at the 3-day scale and ≤ 0.70 at the monthly and annual scales. Both merged products outperformed the precipitation products used in the merging procedure at all temporal scales. Similarly, the merged products performed better than MSWEPv2.2 at all temporal scales except DJF (summer), where all products showed a reduced performance and a greater dispersion in the KGE' values. This low performance in summer is the reason why the products exhibit lower KGE' values at the annual scale compared to the monthly scale (Baez-Villanueva et al. 2020, their Figure 4).</p> <p>Both merged products present a median r value of 0.94 at the 3-day temporal scale (Baez-Villanueva et al. 2020, their Figure 5), which is consistent with the improvements in r obtained by Xie and Xiong (2011) and Yang et al. (2017). Of the existing products, MSWEPv2.2 performed best with a median value of 0.89, highlighting the advantage of merging gauge, satellite, and reanalysis products. At all time scales, RF-MEP_{3P} and RF-MEP_{5P} performed considerably better than the products used in their computation. This demonstrates that the method is able to substantially improve the correlation of the precipitation products for the Chilean case study.</p> <p>RF-MEP_{5P}, RF-MEP_{3P}, MSWEPv2.2, and CHIRPSv2 were almost unbiased. Both merged datasets present lower dispersion than MSWEPv2.2 and CHIRPSv2 for all temporal scales except DJF. In DJF, both merged products presented a $\beta > 1$ and were outperformed by MSWEPv2.2. All datasets underestimated the variability of precipitation at all temporal scales. MSWEPv2.2 best represented the variability of the ground-based measurements, followed closely by both merged datasets. The high values of γ obtained for MSWEPv2.2 were expected because this product uses the same daily ground-based Chilean dataset in its computation and accounts for the difference in reporting times. Both merged products showed a reduced dispersion of the KGE' components at the 3-day, monthly, MAM, JJA, and SON scales; however, the dispersion at the annual scale increases due to the reduced performance in DJF.</p>
Spatial	<p>All products show median KGE' values lower than 0.5 and high dispersion in the Far North (Baez-Villanueva et al. 2020, their Figure 6). These regions are classified as arid according to the Köppen-Geiger climate classification, demonstrating that the performance of the evaluated products over the arid regions of Chile remains low. MSWEPv2.2 and both merged products perform considerably better than the products used as covariates, highlighting the benefit of combining data from precipitation products and ground-based measurements. The performance of all the products increased over Central Chile and South, where annual precipitation volumes are much higher than in the Far North and Near North.</p> <p>For both merged products, more than 80% of the stations in the validation set yielded KGE' values higher than 0.60 (Baez-Villanueva et al. 2020, their Figure 7). Both merged products performed best in the Near North, Central, and Southern Chile, with median KGE' values of 0.84, 0.86, and 0.81, respectively. However, in the Far North, MSWEPv2.2 performed the best (0.61), followed by RF-MEP_{3P} (0.35) and RF-MEP_{5P} (0.28). The results in the Far North show that the inclusion of more precipitation products does not necessarily improve the median performance of the merged product; however, the inclusion of the additional two products reduced the dispersion in the KGE' values of RF-MEP_{5P}. Despite the poor performance of the products used as covariates in the Far North, RF-MEP_{3P} and RF-MEP_{5P} were able to extract useful information from these products to obtain a better performance. RF-MEP_{5P} and RF-MEP_{3P} performed better in the high elevations of the Far North region compared to the low elevations. These high elevations correspond to the alpine tundra climate (ET). In contrast, the cold and arid desert climate (BWk) dominates the lower areas of the Far North, where the precipitation datasets presented their worst performance. This suggests that arid climates present a great challenge for the existing products.</p>
Categorical	<p>All datasets, except for RF-MEP_{3P} and RF-MEP_{5P}, obtained POD values lower than 0.45 for precipitation events > 1 mm, while the no rain events were well captured by all products (Baez-Villanueva et al. 2020, their Figure 8). Similar results were observed for the FAR and CSI, where RF-MEP_{3P} and RF-MEP_{5P} presented the best performance. FAR values were consistently the worst for the light rain intensities (1,5) mm d⁻¹, highlighting that the products remain unable to capture low precipitation values adequately. The CSI presents the best performance for no rain events followed by extreme events (≥ 40 mm d⁻¹), as a result of the decreased FAR compared to the other precipitation intensities.</p> <p>Finally, the median values of the f bias showed that all products overestimated the number of light rain (1,5) mm d⁻¹ and moderate rain events (5, 20) mm d⁻¹. RF-MEP_{3P} and RF-MEP_{5P} performed the best in terms of f bias for the heavy rain events (20, 40) mm d⁻¹, while MSWEPv2.2 performed the best for the other intensities, followed by the merged products. All products underestimated the occurrence of violent rain events (≥ 40 mm d⁻¹).</p>

Both merged products performed similarly well with a median KGE' of 0.83 because they use the Chilean rain gauges; however, the five products used in their computation performed slightly worse in the 1-day evaluation due to the 11 hour difference in the reporting times. The 3-day temporal scale was considered sufficient to render the difference in reporting times negligible.

4.2.2 Correction of mismatches from the original precipitation products

The results showed that the blending of multiple precipitation estimates, topography-related information, and ground-based measurements can improve the spatio-temporal characterisation of precipitation (Baez-Villanueva et al. 2020, their Figures 8–10), which is consistent with the results obtained by Verdin et al. (2016) and Manz et al. (2016). The r , β , and γ components also improved at all temporal scales. The γ of both merged products showed a systematic underestimation ($\gamma \sim 0.9$; Baez-Villanueva et al. 2020, their Figure 5) at all temporal scales as a consequence of averaging the predictions of the different trees from the RF model. Despite this, the γ values of the merged products are higher than those shown by the products used as covariates.

Recently, Alvarez-Garreton et al. (2018) derived runoff coefficients larger than 1, mainly over Central Chile and in the Far South, with increasing coefficient values towards the Andes. These findings are consistent with those of Beck et al. (2017a), indicating that more water is leaving the catchments than the total amount entering as precipitation. This suggests that the products systematically underestimate precipitation at high elevations throughout Chile, which may be due to the inability of satellite-based products to estimate precipitation and ice-covered surfaces over snow accurately (Beck et al. 2017a). Also, during winter, most Chilean rain gauges located at high elevations cannot correctly incorporate snow into the precipitation measurement, leading to an underestimation of precipitation. Therefore, even considering the good performance of the two merged products at different temporal scales, it is likely that the real amount of precipitation is underestimated at high elevations due to the absence of ground-based information. To reduce the possible underestimation of precipitation over high elevated and snow-driven catchments, the incorporation of rain gauges able to measure both liquid and solid precipitation at high elevations is recommended, along with the use of precipitation products that account for both solid precipitation (such as MSWEPv2.2 and reanalysis-based products).

CHAPTER 4. RESULTS

The inclusion of different precipitation products improved the detection of different intensities at the daily scale, as observed in the improved categorical performance of the merged products compared to that of the covariates (Baez-Villanueva et al. 2020, their Figure 8 and Table 5). The categorical performance of both merged products showed improved detection of the selected precipitation intensities and a reduction in the number of incorrectly classified days. These results, in combination with the improved values of r and β , show that RF-MEP is capable of improving the spatial representation of precipitation patterns at the daily scale by assigning more accurate precipitation amounts to each day, while preserving the total volume of precipitation at larger scales.

The analysis of the precipitation products at different intensities is affected by the difference in reporting times between the products and the ground-based measurements. All the products used as covariates, except for CHIRPSv2 and TRMM 3B42v7, presented statistically significant differences at the 95% confidence interval between the daily and 3-day values. Unfortunately, this issue is ignored in most precipitation evaluation studies and constitutes a major limitation of most evaluations carried out in time zones far from 0:00 UTC.

Precipitation is overestimated in the Far North by CMORPHv1, PERSIANN-CDR, TRMM 3B42v7, CHIRPSv2, and ERA-Interim (Baez-Villanueva et al. 2020, their Figure 13). Consequently, both merged products overestimate precipitation over this region (except for the high elevated areas). These results are in agreement with Dinku et al. (2011), and Zambrano-Bigiarini et al. (2017), where the products overestimated precipitation over the arid regions of Africa and Chile, respectively. MSWEPv2.2 and the merged products were able to capture the precipitation volume over the mountainous areas in the Far North, despite the challenge presented by climate variability caused by extreme topography and lack of ground-based measurements (Maggioni and Massari 2018).

The merged products show a lower relative difference, i.e., good performance, for almost all the stations in the Near North, Central Chile, South, and elevated areas in the Far North. The improved performance of the merged products can be observed in the lower panel of Figure 13 of Baez-Villanueva et al. (2020), which highlights that the majority of the precipitation products presented relative differences between -0.2 and 0.2 compared to rain gauges. This suggests that RF-MEP is capable of representing the mean annual precipitation patterns when applied at the daily temporal scale. The overestimation over the Far North is expected because all products used to derive both merged products tend to overestimate precipitation over this region.

4.2.3 Impact of density, spatial resolution, and limitations of RF-MEP

A high number of rain gauge stations in the training set leads to higher performance and higher detection of precipitation intensities (Baez-Villanueva et al. 2020, their Figure 9). When the training sample to 10% (37) of the total available stations was reduced, RF-MEP_{5P} was still able to outperform the products used as covariates, showing the effectiveness of the proposed RF-MEP method.

The products RF-MEP_{5P} and RF-MEP_{3P} performed similarly (Baez-Villanueva et al. 2020, their Figures 4–8). The median values and the interquartile ranges of the KGE', r , β , and γ are similar for both merged products, except over the Far North, where RF-MEP_{5P} shows less dispersion in the KGE' and its components than RF-MEP_{3P}, despite the slight decrease in the median performance. This indicates that the inclusion of more precipitation products could reduce the dispersion in areas where the selected products show low performance. The similar performance of RF-MEP_{5P} and RF-MEP_{3P} indicates that the method is able to extract useful information from the precipitation products. Similar results were obtained when RF-MEP_{3P} used ERA-Interim, CMORPHv1, and TRMM 3B42v7 instead of ERA-Interim, CHIRPSv2, and PERSIANN-CDR (see Appendix B.1), demonstrating that RF-MEP is a robust merging method. Although the precipitation products must be resampled to the same spatial resolution to generate the merged product, the effect of including products generated at different spatial resolutions is negligible (see Baez-Villanueva et al. 2020, their Figure 10).

RF-MEP_{5P} includes CMORPHv1 and TRMM 3B42v7, which reduces the potential temporal coverage by 15 years (RF-MEP_{3P} can be generated from 1983 onwards, while RF-MEP_{5P} can only be generated from 1998). Therefore, based on the similar strong performances of both merged products, RF-MEP_{3P} is preferred for the Chilean case study, as the benefits of including CMORPHv1 and TRMM 3B42v7 to generate RF-MEP_{5P} are outweighed by the loss of 15 years of record.

Although RF-MEP was only applied over Chile, it could be successfully applied over other areas due to its outstanding performance in a region with notable heterogeneity in topography and climate even when the training set was largely reduced. However, some limitations apply to this method: *i*) since ground-based data are necessary, it would be difficult to apply the proposed method globally and in near-real time; *ii*) it can be computationally intensive when applied to large areas; and *iii*) it has problems predicting values that are completely out from the training range.

4.2.4 Comparison with other merging methods

RF-MEP_{5P} was evaluated against the other two merging methods (KED, OOR arithmetic mean) and against MSWEPv2.2 (see Baez-Villanueva et al. 2020, their Figure 11). The performance of ERA-Interim was also included because it was the best-performing product used in the merging procedure. RF-MEP_{5P} showed the best performance at the 3-day temporal scale, followed by KED and MSWEPv2.2. The OOR arithmetic mean product shows the lowest KGE', γ , and r ; however, it can accurately represent the total precipitation volume at the 3-day scale. This product also shows the lowest performance when evaluated at different intensities.

Shen et al. (2014) concluded that the categorical performance of the OOR arithmetic mean product improved compared to the selected products; however, they evaluated the categorical performance only for rain and no rain events. The distribution of daily precipitation is heavily skewed; therefore, the product's performance over different intensities can be masked by the no rain events. As observed in the lower panel of Figure 11 of Baez-Villanueva et al. (2020), averaging different products reduces the performance at all precipitation intensities because all these products have errors in detection (i.e., the products may estimate different precipitation intensities for a particular day). This analysis suggests that precipitation products should not be averaged to attempt to improve daily precipitation patterns.

KED performed similarly to RF-MEP_{5P}; however, RF-MEP_{5P} showed less dispersion in the KGE' and its components, suggesting that RF-MEP is a robust method to merge precipitation products and ground-based data. Ly et al. (2011) obtained poor results when using KED with few sample points, which indicates that the number of ground stations highly influences the performance of KED. Conversely, RF-MEP performed relatively well when the training set was dramatically reduced. The performance of RF-MEP_{5P} is also the highest at monthly, annual and seasonal temporal scales, except in DJF where MSWEPv2.2 performs the best (see Figure S1 from Appendix B.1).

4.3 Regionalisation

One independent calibration (2000–2014) and two verification periods i.e., a near-normal/wet period (Verification 1; 1990–1999) and a dry period (Verification 2; 2015–2018) were used to calibrate the 100 catchments to obtain the best model parameter set for each catchment. Then, both verification periods were used to evaluate the performance of the parameter sets over data that were not used during calibration. Finally, three regionalisation techniques were used to predict daily streamflow over ungauged catchments through a leave-one-out exercise (see Section 3.3).

4.3.1 Performance of precipitation products

The summary of the results of the independent calibration and both verification periods, as well as the results of the regionalisation exercise from Baez-Villanueva et al. (2021) are presented in Table 4.4.

During the independent catchment calibration and two verification periods, good performances were obtained with all products (Baez-Villanueva et al. 2020, their Figure 4). When decomposing the results of the KGE' into its three components, r exhibited the lowest performance. At the same time, β and γ values were generally closer to their optimal values, particularly for calibration and Verification 1. The results obtained with ERA5, which is a reanalysis product, were as good or even better than those obtained with the gauge-corrected products CR2MET, RF-MEP, and MSWEPv2.8 (e.g., see results for the pluvio-nival catchments in Baez-Villanueva et al. 2020, their Figure 4). This is in agreement with Tarek et al. (2020), who concluded that ERA5 should be considered a high-potential dataset for hydrological modelling in data-scarce regions. The good performance of ERA5 suggests that, for the particular case of Chile, merging precipitation products with ground-based measurements does not necessarily translate into improved hydrological model performance, which may be attributed to *i*) the lack of precipitation rain gauges in the Andes Mountains; *ii*) the ability of the rainfall-runoff model to compensate for the precipitation forcing (visible in the performances of the β and γ components; Appendix C.1); and *iii*) the fact that precipitation products still have errors in detecting precipitation events that could impact the representation of the modelled streamflow dynamics (as suggested by the relative lower performance of the r component of the KGE').

Furthermore, the similar performances obtained with uncorrected (ERA5) and gauge-corrected (CR2MET, RF-MEP, and MSWEPv2.8) products, both in wet and dry periods, highlight that there was no single precipitation dataset outperforming the

CHAPTER 4. RESULTS

others in all periods. These results demonstrate that the calibration of hydrological model parameters smooths out, to some extent, the spatio-temporal differences between precipitation products (Baez-Villanueva et al. 2020, their Figures 2, 3, 6, and 9). This is in agreement with previous studies that have demonstrated that model calibration with each product improves the performance of streamflow simulations (e.g., Artan et al. 2007; Stisen and Sandholt 2010; Bitew et al. 2012; Thiemig et al. 2013). The decomposition of the KGE' into its components also demonstrated the ability of the TUWmodel to compensate for the total volume of precipitation, as the β component was close to the optimum value, particularly for calibration and Verification 1 (see Appendix C.1). This can be attributed to the improved detection of precipitation events of the merged products (regarding RF-MEP, see Baez-Villanueva et al. 2020), which can also be observed for MSWEPv2.8, as it produced the best performance over snow-dominated catchments under dry conditions (Verification 2).

Regarding the suitability of precipitation products for parameter regionalisation, RF-MEP provided slightly better results in the Far North for the calibration period using both spatial proximity and feature similarity, suggesting that precipitation products merged with ground-based information over arid climates can improve regionalisation performance. The lower performance obtained in regionalisation with ERA5 in the Far North compared to the other products can be attributed to its high precipitation values, which are likely due to the lack of ground-based precipitation stations over Chile in the development of the product. The incorporation of ground-based measurements has the potential to *i*) compensate for overestimations caused by the evaporation of hydrometeors before they reach the ground (Maggioni and Massari 2018); and *ii*) improve event-based detection skills (Baez-Villanueva et al. 2020; Zhang et al. 2021). The latter is evident in CR2MET and MSWEPv2.8, which are both based on ERA5 but include several rain gauges in the Far North, and have a higher performance than ERA5 (see Baez-Villanueva et al. 2020, their Figures 2 and 3).

Despite the low performance of all precipitation products in the Far North and Near North (median KGE' values <0.58 , see Baez-Villanueva et al. 2020, their Figure 7), the TUWmodel was flexible enough to compensate, to some extent, for differences between products. A similar conclusion was obtained by Elsner et al. (2014), who examined differences between four meteorological forcing datasets and their implications in hydrological model calibration in the western USA using the Variable Infiltration Capacity model (VIC; Liang et al. 1994). These results are also in agreement with Bisselink et al. (2016), who concluded that parameter sets obtained during calibration partially compensated the bias of seven precipitation products used to force the fully-distributed LISFLOOD model in four catchments in southern Africa.

Table 4.4: Calibration, verification, and regionalisation results over 100 near-natural catchments.

Period of assessment	Main Results
Calibration and Verification	<p>CR2MET provided the best performance for all evaluated periods, followed closely by RF-MEP (Baez-Villanueva et al. 2020, their Figure 4). MSWEPv2.8 provided the poorest performance for calibration and Verification 1. The lowest performances were obtained during the Verification 2 period, emphasising the challenges of estimating streamflow in dry conditions, as discussed by Maggioni et al. (2013) and Beck et al. (2016). TUWmodel performed well for all precipitation products for all periods. The selected parameter ranges were wide enough so that calibrated parameter values were not concentrated at their lower or upper limits (see Appendix C.1). The TUWmodel performed better over the pluvio-nival catchments (see Baez-Villanueva et al. 2020, their Figure 5). During the calibration period, there was no clear second-best regime. The snow-dominated catchments presented slightly higher median KGE' values but a more pronounced dispersion, while the pluvio-nival and rain-dominated catchments presented lower dispersion but reduced median values. The snow-dominated catchments presented a more pronounced decrease from calibration to both verification periods. The rain-dominated catchments presented the highest dispersion increases in both verification periods compared to calibration. Over the snow-dominated catchments, ERA5 performed the worst and presented the highest dispersion and the lowest median KGE', despite having the highest median KGE' during calibration (0.87). RF-MEP performed the best during Verification 1 (0.68), while MSWEPv2.8 performed the best during the dry Verification 2 period (median KGE' of 0.60). CR2MET performed the best over the nivo-pluvial catchments. In contrast, RF-MEP performed relatively worse for both verification periods with a larger dispersion, despite having a similar median KGE' (0.62) in Verification 1 to ERA5 and MSWEPv2.8. Over the pluvio-nival catchments, all products showed a relatively good performance, with CR2MET being the best product in calibration and Verification 1, while ERA5 performed the best during Verification 2. RF-MEP performed the best over the rain-dominated catchments in calibration and Verification 1, while ERA5 performed the worst. Finally, CR2MET performed the best in Verification 2, followed by MSWEPv2.8.</p>
Regionalisation	<p>Feature similarity had the best median performance for all products and over all periods, followed by spatial proximity and parameter regression. In addition to exhibiting a considerably lower overall performance, parameter regression returned a larger spread in the KGE's for all periods (see Baez-Villanueva et al. 2020, their Figure 6). The overall performances obtained for feature similarity and spatial proximity are relatively close for different products over each period (see Baez-Villanueva et al. 2020, their Figure 6). For feature similarity, all products generated acceptable KGE' results during the calibration and Verification 1 periods, while the median KGE' values during the dry Verification 2 were lower. CR2MET obtained the best model performance for feature similarity, followed closely by RF-MEP for calibration, ERA5 for Verification 1, and MSWEPv2.8 for Verification 2. In the case of spatial proximity, MSWEPv2.8 yielded the best performance in the calibration period (0.55), followed closely by RF-MEP (0.56, but with a higher dispersion) and CR2MET (0.53). For Verification 1, RF-MEP provided the best performance (0.54), while MSWEPv2.8 produced the best results over Verification 2 (0.48). For spatial proximity, ERA5 performed the worst over the three evaluated periods. Parameter regression yielded the lowest results, with CR2MET and ERA5 showing the highest median KGE' values. The spatial patterns obtained for all regionalisation methods were similar, independent of the product or the evaluated period, except for parameter regression, which yielded poor results over high-elevation catchments and under dry conditions (Verification 2, Baez-Villanueva et al. 2020, their Figure 7).</p> <p>All products performed better in the Central Chile and South regions than in the Far North, Near North, and Far South regions. The low performance of regionalisation in the arid north is very likely due to the convective nature of storms occurring in the highlands of the Chilean Altiplano (elevations above 4,000 m a.s.l.) and the low density of streamflow stations over this area. Despite this general low performance, RF-MEP was the best performing product over the Far North region for both spatial proximity and feature similarity in the calibration period, suggesting that merging precipitation products and ground-based observations helps to improve, to some extent, the performance of hydrological modelling across arid regions. Conversely, all products outperformed RF-MEP over the Far South. Spatial proximity provides the best performance over the Far South, for CR2MET, RF-MEP, ERA5, and MSWEPv2.8, respectively (see Baez-Villanueva et al. 2020, their Figure 6). The systematic lower performance of feature similarity compared to spatial proximity over the Far South (except for the case of ERA5) could be attributed to <i>i</i>) the lack of catchment characteristics that represent the hydrological behaviour of this complex area dominated by polar and temperate climates; and <i>ii</i>) the low amount of potential donor catchments (even for latitudes $> 49^{\circ}\text{S}$), combined with their varied hydrological regimes.</p>

CHAPTER 4. RESULTS

An unexpected result from this study is that the spatial resolution of the precipitation products did not play a major role in model performance during calibration, verification, and regionalisation; although CR2MET and RF-MEP have a higher spatial resolution (0.05° ; $\sim 25 \text{ km}^2$) than MSWEPv2.8 ($\sim 0.10^\circ$; $\sim 100 \text{ km}^2$) and ERA5 ($\sim 0.28^\circ$; $\sim 625 \text{ km}^2$), all four products performed well during the independent calibration and verification of the hydrological model. The performance of ERA5 over the 25 smallest catchments during regionalisation (area $< 353.1 \text{ km}^2$) was similar to that obtained with products with a higher spatial resolution (see Appendix C.1). This can be attributed to the fact that Chile is dominated by large-scale frontal systems (Zhang and Wang 2021); and therefore, coarse-resolution products may perform well over small catchments. Our results also align with the findings of Maggioni et al. (2013), who concluded that the loss of spatial information associated with coarser resolution (e.g., ERA5) can be compensated through model calibration.

The calibration of TUWmodel was able to compensate, to some extent, for differences in annual and intra-annual precipitation amounts, intermittency, and extremes (see Baez-Villanueva et al. 2020, their Figure 2 and 3) among the four products. The example of the nivo-pluvial catchments illustrates how TUWmodel parameters compensate for differences between the precipitation forcings used in calibration (Baez-Villanueva et al. 2020, their Figure 12), and the corresponding variations in the mean monthly water balance components (Baez-Villanueva et al. 2020, their Figure 13). Similar figures for snow-dominated, pluvio-nival, and rain-dominated catchments can be found in Appendix C.1.

In general, the calibrated parameters behave as expected for each hydrological regime. A notable exception is ERA5, which shows low values for the snow correction factor (SCF) in nivo-pluvial and snow-dominated catchments (see and Appendix C.1 and Baez-Villanueva et al. 2020, their Figure 13). These catchments are primarily located in the arid Near North region, where the estimated winter precipitation is substantially lower for CR2MET, RF-MEP, and MSWEPv2.8, and a high SCF corrects this apparent underestimation. The lower precipitation amounts presented in these products may reflect the incorporation of information from rain gauges located in drier, low-lying areas to correct their precipitation estimates.

ERA5 presented relatively low SCF values over nivo-pluvial catchments compared to the other products (Baez-Villanueva et al. 2020, their Figure 13), which is expected because it exhibits the highest precipitation values. Conversely, because RF-MEP has the lowest mean monthly precipitation over the nivo-pluvial catchments, the model adjusts the evaporation, snow water equivalent, and soil moisture components, thus increasing the simulated streamflow (to match the observed streamflow). Substantial differences were obtained for LPrat and field capacity (FC), affecting

evaporation and soil moisture. For example, over the nivo-pluvial catchments, the LPrat and FC values for RF-MEP are similar to those of ERA5, despite RF-MEP having substantially lower precipitation amounts, which in turn is reflected in the reduced soil moisture and evaporation amounts. The differences between LPrat and FC according to precipitation product are even more pronounced for snow-dominated catchments (see Appendix C.1).

Finally, higher values of the nonlinear parameter for runoff production Beta reduce the amount of water that leaves the catchment as runoff (Széles et al. 2020, their Eq. 7). For all hydrological regimes except pluvio-nival, the median Beta parameter is substantially higher for ERA5 than for the other precipitation products. The larger Beta values obtained with ERA5 are expected to attenuate the runoff generation from extreme precipitation events (Baez-Villanueva et al. 2020, their Figure 13c–d). Interestingly, the Beta parameter is zero in some pluvio-nival catchments, which means that all liquid precipitation and snowmelt was used to generate runoff. This behaviour was more pronounced with RF-MEP and MSWEPv2.8, which exhibited the lowest precipitation amounts and longer dry spells over these catchments. In general, the storage components obtained from each precipitation product (computed as the sum of the two deepest reservoirs of the model (see Széles et al. 2020, their Figure 3) are similar for all four precipitation products.

4.3.2 Evaluation of regionalisation techniques

The main results of the evaluation of the regionalisation are presented in Table 4.5. The compensation due to the flexibility of the TUWmodel observed during the independent calibration and verification also influences the regionalisation performance. Feature similarity provided the best performance when the TUWmodel was forced with all precipitation products (Baez-Villanueva et al. 2020, their Figure 8), while spatial proximity provided similar performance to feature similarity over the Central Chile and South regions, where there is a high density of streamflow stations. These results are in agreement with Parajka et al. (2005), Oudin et al. (2008), and Neri et al. (2020), who demonstrated that spatial proximity performs well over densely gauged regions.

The inclusion of donor catchments with low model performance introduces a diversity that has the potential to benefit streamflow prediction in ungauged catchments, as discussed by Oudin et al. (2008). These catchments were incorporated in the regionalisation process because of the diversity of climates and physiographic characteristics across continental Chile, with the potential downside that this may lead to errors in the transferred model parameters. Additionally, the similarity between the

CHAPTER 4. RESULTS

performance of spatial proximity and feature similarity can be partially attributed to the fact that six of the nine selected catchment characteristics are directly or indirectly related to climate, which in Chile is highly related to the geographical locations of the catchments. Parameter regression was the regionalisation method that provided the worst results (Baez-Villanueva et al. 2020, their Figures 6 and 9); however, Figure 7 of Baez-Villanueva et al. (2021), shows that this method generated good results over low-elevated areas of the Central Chile and South regions, where there are many potential donor catchments located nearby.

The compensation for precipitation differences obtained through model calibration also affected the relative performance of regionalisation techniques, producing unrealistic parameter sets in some donor catchments. In particular, such compensation may have impacted the spatial transferability of model parameters with the parameter regression method. The main reason is that, unlike techniques that transfer the entire parameter sets, the regression process denatures the already uncertain model parameters by applying independent regression procedures using climate and physiographic characteristics (Arsenault and Brissette 2014). This challenge can be overcome by simultaneously optimising both the model parameters and the regression equations (e.g., Samaniego et al. 2010; Rakovec et al. 2016; Beck et al. 2020a), but such an exercise is outside of the scope of this study.

Spatial proximity and feature similarity performed better over pluvio-nival catchments and worst over rain-dominated catchments. The good performance over pluvio-nival catchments could be in part attributed to *i*) the ability of the model to reproduce streamflow in this regime; and *ii*) the increased likelihood of transferring model parameters from a catchment with the same hydrological regime, as they are grouped close together and form 40% of the total number of catchments. However, the rain-dominated catchments (19 in total) performed slightly worse than the snow-dominated catchments (16 in total), which performed the worst during both verification periods.

Table 4.5: Summary of the results of the overall performance of the regionalisation techniques and the impact of hydrological regimes.

Type of assessment	Main Results
Overall performance	<p>The independent calibration of each catchment represents the highest model performance that can be obtained for a specific combination of hydrological model, objective function, and catchment (i.e., an absolute benchmark), whereas the two verification periods were used to evaluate the performance of the regionalisation techniques over independent periods (i.e., as verification benchmarks, Baez-Villanueva et al. 2020, their Figure 8). There are marked differences in performance according to the precipitation product used to force the TUWmodel, regardless of the regionalisation method and the evaluated period. For example, ERA5 has more dispersion in the KGE' values than other products for feature similarity and spatial proximity, while it tends to perform the best for parameter regression. For all products and evaluation periods, feature similarity performed the best, followed by spatial proximity and parameter regression, which is consistent with results from multiple studies (e.g., Parajka et al. 2005; Ondin et al. 2008; Bao et al. 2012; Garambois et al. 2015; Neri et al. 2020). Parameter regression had both the lowest median KGE's and the largest spread. Comparing the two verification periods, results obtained during the (near-normal/wet) Verification 1 period were close to those obtained during calibration, while those obtained during the (dry) Verification 2 were substantially lower, especially for spatial proximity and parameter regression.</p> <p>These results are in agreement with the lower panels located below each map in Baez-Villanueva et al. (2021); their Figure 7, which shows the empirical cumulative distribution functions (ECDFs) of the performance of each regionalisation technique during the entire period of analysis (1990–2018). These ECDFs compare the relative performance of each regionalisation method against those obtained from the independent calibration and verification of each catchment (used as benchmarks). As expected, all regionalisation methods presented a lower performance than the independent calibration and verification, with this reduction more pronounced for parameter regression.</p>
Impact of hydrological regimes	<p>Feature similarity provided the best median performance for all hydrological regimes and precipitation products except for snow-dominated catchments, where spatial proximity performed the best for MSWEPv2.8 for calibration and Verification 2 (see Appendix C.1 and Baez-Villanueva et al. 2020, their Figure 9). These results demonstrate that no single precipitation product outperformed the others for all regionalisation techniques and hydrological regimes. In other words, for this case study, the best performing precipitation product depends on the hydrological regime and chosen regionalisation method. For feature similarity in snow-dominated catchments, RF-MEP performed the best for calibration and Verification 1, while CR2MET performed the best during Verification 2. For nivo-pluvial catchments, CR2MET provided the best performance during calibration and Verification 1, while MSWEPv2.8 performed the best during Verification 2. CR2MET and ERA5 performed the best in pluvio-nival catchments for the case of feature similarity, while all products performed similarly for spatial proximity. Finally, ERA5 performed the best for feature similarity in all periods across the rain-dominated catchments.</p> <p>For both spatial proximity and feature similarity, the best and worst results were obtained for pluvio-nival catchments and rain-dominated catchments, respectively. Comparing Figures 5 and 9 from Baez-Villanueva et al. (2021), it is evident that the snow-dominated catchments performed substantially worse than in the independent performance during the same period. On the other hand, the pluvio-nival catchments performed systematically better in the independent calibration, verification, and regionalisation. This could be attributed to <i>i)</i> the ability of the model to reproduce streamflow in this regime and <i>ii)</i> the increased likelihood of transferring model parameters from a catchment with the same hydrological regime, as they are grouped close together and conform 40% of the total number of catchments.</p>

4.3.3 Impact of nested catchments

The influence of the nested catchments on the regionalisation results was evaluated. The order of performance of the regionalisation methods and precipitation products did not vary when the nested catchments were excluded, as feature similarity and CR2MET remained the best performing method and product, respectively (Baez-Villanueva et al. 2020, their Figure 10). As expected, the regionalisation technique with the largest reduction in performance when excluding nested catchments was spatial proximity, followed closely by feature similarity. All products showed a slight performance reduction and increased dispersion for spatial proximity, except for MSWEPv2.8, which showed a slight increase in the KGE' median value. Feature similarity slightly reduced performance when the nested catchments were excluded; however, the median values remained almost the same. The change in parameter regression performance was negligible after excluding nested catchments because, in the particular case of Chile, excluding only a few catchments had a negligible effect on the non-linear relationships between model parameters and the selected climatic and physiographic characteristics.

Nested catchments play an important role in the performance of regionalisation methods as they are more likely to have a strong climatological and physiological similarity to each other. The regionalisation method that was most impacted by the exclusion of nested catchments was spatial proximity, followed by feature similarity. These results are in agreement with previous studies where the exclusion of nested catchments reduced the performance of regionalisation techniques (Merz and Blöschl 2004; Oudin et al. 2008; Neri et al. 2020). Feature similarity only presented a slight decrease when the nested catchments were neglected, which can be attributed to the low degree of nestedness (i.e., the number of catchments that are nested in a larger one). As expected, the exclusion of nested catchments had a negligible effect on parameter regression, as the removal of relatively few catchments had a negligible impact on the non-linear relationships between the climatic and physiographic characteristics and the model parameters that were determined using all potential donor catchments. The reduction of regionalisation performance when the nested catchments were removed was lower than the reduction reported in a case study over Austria (Neri et al. 2020, their Figure 9a), which could be attributed to *i*) the degree of nestedness, as the unique geography of Chile limits, to some extent, the number of nested catchments within any larger catchment (only 10 of the 100 selected catchments contained more than three nested catchments); and *ii*) the percentage of catchments that are nested (42% in this study, compared to 65% in the Austrian case study).

4.3.4 Impact of donors in feature similarity

In general, the highest median performance was obtained when using four or more donor catchments (Baez-Villanueva et al. 2020, their Figure 11). However, a t-test demonstrated that the improvement in the KGE' values obtained when increasing to more than one donor was not statistically significant. The results show that the performance varies according to the precipitation product and selected analysis period. For the calibration period, feature similarity produced similar median values to those obtained with spatial proximity when one donor was used, while the performance improved as more donors were included. For both verification periods, feature similarity (median KGE' values ranging from 0.44 to 0.64) outperformed spatial proximity (median KGE' values ranging from 0.39 to 0.54). For all three periods, feature similarity provided better performance considering the distribution of the KGE' values.

Increasing the number of donor catchments in feature similarity improved the regionalisation performance. This is in agreement with several studies that have demonstrated that using an ensemble of multiple donor catchments improves regionalisation results (McIntyre et al. 2005; Zelelew and Alfredsen 2014; Garambois et al. 2015; Beck et al. 2016; Neri et al. 2020). These results are similar to those of Neri et al. (2020), who determined that three donors were optimal for the TUWmodel over Austrian catchments. Feature similarity still outperformed spatial proximity when only one catchment was used to transfer the model parameters to the ungauged catchments, which is in agreement with multiple studies that have shown the ability of this method to produce good regionalisation results (Parajka et al. 2005; Oudin et al. 2008; Bao et al. 2012; Garambois et al. 2015; Neri et al. 2020).

CHAPTER 4. RESULTS

Chapter 5

Discussion

5.1 Evaluation of precipitation products and its influence on water resources management

Precipitation is the water cycle's major component; therefore, its accurate estimation is crucial for diverse hydrological, agricultural, and ecological applications; and, in general, for water resources management. However, in many regions, the rain gauge networks are often too sparse to account for the spatio-temporal variability of precipitation. In this sense, in developing countries, where the need for information is the greatest, data collection and monitoring infrastructure is generally low. This problem is exacerbated with the ongoing lack of investment in infrastructure and human capital (Fay et al. 2017; Sheffield et al. 2018) and the ongoing decline in agricultural and hydrometeorological networks (Lorenz and Kunstmann 2012).

Precipitation products can be used to account for the spatio-temporal distribution of precipitation over data-scarce settings (McCabe et al. 2017). Therefore, these products can be used to provide valuable information for supporting water management, plan and design infrastructure, support management and operational decisions, developing early warning systems, and for disaster management purposes (e.g., Rossa et al. 2010; Brown et al. 2014; Beck et al. 2017b; Eggimann et al. 2017; Herman et al. 2018; Sheffield et al. 2018; Baez-Villanueva et al. 2021). These products have different *i*) data sources (they can be derived with satellite data, reanalysis models, ground-based data, or a combination of them); *ii*) spatial resolution (generally from 0.05° to 0.50°); *iii*) temporal resolution (from 30-minutes to the annual scale); and *iv*) latency (from hours to several years).

CHAPTER 5. DISCUSSION

Despite the opportunities that these products provide, a plethora of studies have shown that they are still subject to uncertainties and mismatches (Dinku et al. 2010; Melo et al. 2015; Beck et al. 2017b; Zambrano-Bigiarini et al. 2017). Additionally, during the performance evaluation of six precipitation products over three different catchments of Latin America at different temporal scales (Section 4.1), different performances were achieved by the selected products depending on region, climate, temporal scale, and the selected performance index. It is worth mentioning that these products were selected based on their performance reported in previous studies (Dinku et al. 2010; Melo et al. 2015; Salio et al. 2015; Zambrano-Bigiarini et al. 2017). All products had a higher probability of detecting no rain events than rainy days. The evaluated products represented better the moderate precipitation events [5, 20 mm) in all regions in comparison to light [1, 5 mm), heavy [20, 40 mm), and violent rain events (≥ 40 mm) despite that the probability of detecting rainy days was relatively low for all precipitation products, which is in agreement with (Zambrano-Bigiarini et al. 2017). Paraiba do Sul presented the best daily performance followed closely by the Imperial, while the Magdalena had systematically the lowest performance. This could be attributed to the fact that Colombia has more convective precipitation compared to the other study areas (Taszarek et al. 2021).

These results are of paramount importance for water resources management. They indicate that the evaluation of the performance of precipitation products must be tailored to the specific application that will follow the evaluation. Taking the Imperial as an example, MSWEPv2 performed the best at the daily scale, while CHIRPSv2 performed the best at the monthly scale. These products could be used for diverse water management applications, such as the design of flood controls, irrigation systems, prediction of streamflow in ungauged catchments, maximisation of hydropower production, propagation of river flow waves, and flood early warning systems, as the evaluation of the products should be centred on their performance at daily and monthly temporal scales. On the other hand, MSWEPv2 besides CHIRPSv2 could be used for transboundary water agreements, design of long-term management strategies, and water availability and consumption, as these products had the best β and r components, which are related to the total volume of precipitation and the precipitation dynamics, respectively.

It is essential to mention that despite that the evaluation of precipitation products result in the identification of a best performing product, it may still have detection errors that may hinder their application for operational purposes as shown by Zambrano-Bigiarini et al. (2017); Beck et al. (2017b); and Figures 5–12 and 14 from Baez-Villanueva et al. (2018). This is in agreement with the conclusions of Maggioni and Massari (2018) who stated that the intrinsic quality of the precipitation

CHAPTER 5. DISCUSSION

products have shown to impact the error propagation during hydrological modelling exercises. However, there have been considerable advances in measurement techniques in the past few years. This is clearly observed in the clear improvements of the relatively newly released Global Precipitation Measurement (GPM; Huffman et al. 2015) IMERG products when compared to TRMM 3B42v7 and TRMM 3B42RT (Huffman et al. 2015; Zhang et al. 2019; Arshad et al. 2021) as well as ERA5 in comparison to ERA-Interim (Hoffmann et al. 2019; Wang et al. 2019; Hersbach et al. 2020). The improvements of these products show that the upcoming products will surely have great accuracy improvements, which will be related to *i*) the development of novel sensors and measurement devices; *ii*) the fact that measurement devices are available at increasingly lower costs; and *iii*) the fact that sensors can be mounted easier on mobile platforms due to their increasingly reduced size (Eggimann et al. 2017).

In developing countries, the use of these products can help strive towards informed-based decision-making. However, these regions often lack the infrastructure and human capacity due to a systematic lack of infrastructure and training (Fay et al. 2017; Sheffield et al. 2018). Therefore, informed-based urban water management requires a change in practices ranging from network operation and data collection to the processing of data to produce information in a timely manner (Eggimann et al. 2017). In many places, water managers and professionals are not used to acquiring, processing, and producing information using large amounts of data. In this sense, there is a need to move towards data-driven water resources management by implementing strategic approaches that systematically build the capacities and infrastructure of such regions. The introduction of these kinds of practices will require institutional transitions as their implementation is complex and time-consuming (Eggimann et al. 2017).

These institutional changes are pivotal as there is a projected increase in *i*) the frequency and severity of droughts and heatwaves due to climate change (Vörösmarty et al. 2000; Brauman et al. 2016; Konapala et al. 2020; Woolway et al. 2020); *ii*) water resources stress placed by the overexploitation of surface water and groundwater; and *iii*) the demand of agricultural and energy production, and population growth (Kiguchi et al. 2015; Van Ginkel et al. 2018; Qin et al. 2019). For instance, urban floods are expected to be more critical in line with climate change and urbanisation (Whitfield 2012), which introduce the challenge of producing precipitation estimates in an accurate and timely manner. Additionally, Viviroli et al. (2020) states that around 1.5 billion people, which represents 24% of the world's lowland population, are projected to depend substantially on runoff contributions from mountainous catchments by the mid-twenty-first century. They concluded that these areas should be specially considered in water resources management by highlighting their importance towards sustainable development. However, some products have problems in

representing the gradient of precipitation related to elevation (Beck et al. 2020b) and precipitation is often under-represented at higher elevations because most rain gauges are located in lowlands due to accessibility reasons (Derin and Yilmaz 2014). Therefore, the accuracy of these products must be enhanced for local applications by combining information from precipitation products, in-situ measurements, and topography-related data. This is evident in the evaluation results of Section 4.1.1, which show that generally gauge-adjusted products showed better performance compared to those that did not include ground-based measurements. However, it is crucial to consider the independence of the selected rain gauge stations when evaluating the performance of precipitation products to avoid introducing bias in the assesment.

Finally, the evaluation of precipitation products can be affected by the upscaling procedure that enables a fair point-to-pixel comparison. These differences in performance are related to the specific precipitation product, the temporal scale in the evaluation (see Section 3 of Maggioni and Massari 2018), and the physiographic and climatic characteristics of the region in analysis. In this sense, the topography plays an important role when evaluating the performance of precipitation products as the performance of the re-scaled products was more impacted over areas with more pronounced elevation gradients. For this reason, it is essential to consider the effects that the spatial resolution may have over mountainous regions when evaluating the performance of precipitation products over regions with complex topography. On the other hand, it is essential mentioning that despite that rain gauges provide precipitation measurements with a high degree of accuracy at specific locations (Villarini and Krajewski 2008), the evaluation of precipitation products might have errors related to *i*) the underestimation of the performance of precipitation products related to the use of a scarce network of rain gauges (Tang et al. 2018) and *ii*) the fact that the gauges are still subject to errors related to wind-induced under-catch, wetting and evaporation loss, and trace amount (Adam and Lettenmaier 2003; Ma et al. 2015), which may impact the evaluation of precipitation products.

5.2 Implications of merging precipitation datasets

As noted before, the inherent errors in the detection of precipitation events, hit biases, and random errors hinder their use for operational applications (Zambrano-Bigiarini et al. 2017; Baez-Villanueva et al. 2018, 2020). Additionally, the performance of these products is affected by seasonal patterns, storm type, and the topographical features of the study area (Oliveira et al. 2016; Maggioni and Massari 2018; Sheffield et al. 2018). Therefore, there is a need to improve the spatio-temporal distribution of precipitation, which can be achieved by merging information from precipitation products, ground-based measurements, and topographical features (Xie and Xiong 2011). The development of these techniques are based on two main factors: *i*) the increased data availability from new and conventional data sources, and *ii*) the fact that computational power and data storage are becoming cheaper and more accessible, which enables a faster data processing and distribution (Eggimann et al. 2017).

Despite the improvement of precipitation patterns that past merging techniques have achieved (e.g., Li and Shao 2010; Xie and Xiong 2011; Woldemeskel et al. 2013; Shen et al. 2014; Nie et al. 2015; Fu et al. 2016; Manz et al. 2016; Verdin et al. 2016; Shi et al. 2017; Ma et al. 2018), only a single precipitation product is often selected; therefore, valuable information that has the potential to improve the characterisation of precipitation is not considered. In line with this, the two merged products (RF-MEP_{3P} and RF-MEP_{5P}) obtained with the novel RF-MEP merging method (see Section 3.2) showed improved r , β , and γ values at all temporal scales compared to all the individual precipitation products used as covariates. Additionally, both merged datasets exhibited improved POD, FAR, f_{bias} , and CSI for different precipitation intensities. RF-MEP showed better performance than the results obtained using Kriging with external drift and one-outlier-removed arithmetic mean. The good performance in the detection of single precipitation events of the one-outlier-removed arithmetic mean shown in Shen et al. (2014) compared to the results of Section 4.2 are related to the fact that Shen et al. (2014) evaluated the probability of detection of rain/no-rain events, while in this thesis the detection was evaluated over different precipitation intensities (see Table 3.2).

The performance of the merged products increased when more rain gauge stations were used to train the model; however, they were still able to improve precipitation characteristics even with few stations in the training set. Additionally, the difference in reporting times between the precipitation products and the ground-based measurements must be taken into account when assessing the performance of precipitation products at the daily temporal scale so that their performance is not underestimated. This issue constitutes a major limitation of most precipitation evaluation studies

carried out far from 0:00 UTC, which is also the case of Baez-Villanueva et al. (2018).

As observed, merging multiple precipitation estimates with ground-based observations is a promising way to derive useful information for informed-based water resources management. More reliable precipitation products can improve the results of multiple studies related to droughts characterisation (Zambrano-Bigiarini and Baez-Villanueva 2019; Quesada-Montano et al. 2019), reservoir operations (Zhao et al. 2011; Yang et al. 2017), early warning systems (Hossain 2006; Maggioni and Massari 2018), risk assessment (Brunetti et al. 2018; Rajulapati et al. 2020), and water availability (Serrat-Capdevila et al. 2014; McNamara et al. 2021). Following this line, the integration of merged products and hydrological models is a promising way to leverage remote sensing data for water resources management (Sheffield et al. 2018).

The results shown in Section 4.3.1 show the performance of the TUWmodel when forced with diverse precipitation products during calibration (2000–2014) and a near-normal and a dry verification periods (Verification 1 from 1990–1999 and Verification 2 from 2015–2018). The precipitation products showed marked differences in mean annual precipitation and different extreme indices (Baez-Villanueva et al. 2021, their Figures 2 and 3). Despite these differences, good performances were obtained with all products, which demonstrates that the calibration of the hydrological model parameters smooths out, to some extent, the evident differences in annual and intra-annual precipitation amounts, intermittency, and extremes between precipitation products. This is evident when analysing the β component of the KGE', which presented values close to 1 mainly in calibration and Verification 1. This result is in agreement with previous studies that have demonstrated that model calibration with each precipitation product improves the performance of streamflow simulations (e.g., Artan et al. 2007; Stisen and Sandholt 2010; Bitew et al. 2012; Thiemig et al. 2013).

In general, the calibrated parameters behave as expected for each hydrological regime. A notable exception is the snow correction factor (SCF) in nivo-pluvial and snow-dominated catchments, which are related to the fact that the winter precipitation estimates of ERA5 are substantially larger than those shown in CR2MET, RF-MEP, and MSWEPv2.8. Therefore, a high SCF corrects the apparent underestimation of these products compared to ERA5. Keeping the example of the nivo-pluvial catchments, and taking RF-MEP as an example, which showed the lowest monthly precipitation over these catchments, it is visible how the TUWmodel adjusts the evaporation, snow water equivalent, and soil moisture components (Baez-Villanueva et al. 2021, their Figure 13) to increase the simulated streamflow to match the streamflow measurements.

The good performance of ERA5 was surprising and somehow counter-intuitive

as this product *i*) does not use rain gauge stations over Chile and *ii*) it has the lowest spatial resolution among the evaluated products (0.25°). The good performance of ERA5 suggests that, for the particular case of Chile, merging precipitation products with ground-based measurements does not necessarily translate into improved hydrological model performance, which may be attributed to *i*) the lack of precipitation rain gauges in the Andes Mountains; *ii*) the ability of the rainfall-runoff model to compensate for the precipitation forcing; and *iii*) the fact that precipitation products still have errors in the detection of particular events that could impact the representation of the modelled streamflow dynamics (as suggested by the relative lower performance of the r component of the KGE'). These results are in agreement with Tarek et al. (2020), who concluded that ERA5 should be considered a high-potential dataset for hydrological modelling in data-scarce regions.

As discussed in Maggioni and Massari (2018), an alternative strategy for improving the performance of precipitation products in hydrological simulations is to combine them with soil moisture estimates (e.g., Ciabatta et al. 2015; Román-Cascón et al. 2017) despite that the number of dry days due to the relatively noisy soil moisture retrievals (Crow et al. 2011; Brocca et al. 2014; Beck et al. 2017b). Additionally, different strategies can be used to improve the results of hydrological modelling when using precipitation products: *i*) the use of other components of the water balance in the calibration process (Mostafaie et al. 2018; Dembélé et al. 2020a); *ii*) the application of a sensitivity analysis to calibrate the model parameters that are related to dominant physical processes (Zambrano-Bigiarini et al. 2022); *iii*) and the calibration of the hydrological model through multi-objective functions (Kollat et al. 2012; Smith et al. 2019).

5.3 On the selection of precipitation products to predict streamflow in ungauged catchments

To date, few regionalisation studies have used gridded precipitation products at the daily time scale (Samaniego et al. 2010; Beck et al. 2016; Rakovec et al. 2016; Beck et al. 2020a). However, they only selected one precipitation dataset to predict streamflow in ungauged catchments. Therefore, the effects of a precipitation dataset's choice on regionalisation results and, therefore, on water resources management remains unknown.

CHAPTER 5. DISCUSSION

The influence of the selection of different precipitation products on the relative performance of three regionalisation methods (spatial proximity, feature similarity, and parameter regression) is presented in Section 4.3. Feature similarity provided the best performance when the TUWmodel was forced with all precipitation products, followed by spatial proximity and parameter regression. These results are in agreement with Parajka et al. (2005), Oudin et al. (2008) and Neri et al. (2020), who demonstrated that spatial proximity performs well over densely gauged regions. The inclusion of donor catchments with low model performance during calibration and verification introduces a diversity that has the potential to benefit streamflow prediction in ungauged catchments, as discussed by Oudin et al. (2008). Therefore, to account for Chile’s physiographic and climatic diversity, they were included them in the evaluation with the potential downside that this may lead to errors in the transferred model parameters. The similarity in performance between spatial proximity and feature similarity can be related to the fact that six of the nine selected catchment characteristics are directly or indirectly related to climate, which is highly related to the geographical locations of the catchments for the case of Chile.

The compensation for precipitation differences obtained through model calibration also affected the relative performance of regionalisation techniques, producing unrealistic parameter sets in some donor catchments. Parameter regression was the method that was influenced the most by this compensation because the regression process denatures the already uncertain model parameters by applying independent regression procedures using climate and physiographic characteristics (Arsenault and Brissette 2014). This can be overcome by simultaneously optimising both the model parameters and the regression equations (e.g., Samaniego et al. 2010; Rakovec et al. 2016; Beck et al. 2020a). The regionalisation results were also affected by the hydrological regime of the catchments, with best results generally obtained for rain-dominated catchments with a minor snowmelt component. Although the results obtained during regionalisation yielded high performances, they indicate that the precipitation products that were corrected with daily ground-based measurements (i.e., RF-MEP, CR2METv2, and MSWEPv2.8) did not necessarily yield the best regionalisation performance, which is in line with the results of the independent calibration and verification.

Accurate daily streamflow predictions in ungauged catchments are pivotal for water resources management. However, these predictions are challenged by uncertainties arising from precipitation products. In this sense, the results of this study are very promising as they provide guidance for ongoing and future studies involving the application of gridded precipitation products for regionalising hydrological model parameters in ungauged basins.

The contributions from mountainous catchments will be more important by the mid-twenty-first century (Viviroli et al. 2020). However, the majority of these catchments remain ungauged, and there is a lack of ground-based measurements due to their complex topography and difficult access, which hinders the opportunity to shift towards proactive water management approaches. We live in an era where data is exponentially increasing; therefore, machine learning and artificial intelligence techniques will have an essential role in water management. The implementation of these techniques will enable low-latency data transmission, real-time processing, and real-time visualisation (Sun et al. 2018); and will help in *i*) the proliferation of new governance actors; *ii*) the creation of agencies related to environmental sensing; and *iii*) the implementation of transparent data collection strategies, which in turn will create conditions for significant transformations in environmental governance (Bakker and Ritts 2018).

5.4 Future research and recommendations

The methods followed in this cumulative dissertation use open-access global or quasi-global products. Therefore, they can be applied in different data-scarce regions to strive towards an information-based decision-making process. In Section 4.1 it was found that in general, CHIRPSv2 and MSWEPv2 showed the best performance at the different temporal scales over the evaluated catchments. Better results may be achieved if current state-of-the-art precipitation products are used, such as IMERG (Huffman et al. 2015), ERA5 (Hersbach et al. 2020) and MSWEPv2.8 (Beck et al. 2017a, 2019b). Additionally, as mentioned in Section 4.2, it is important to take into consideration the reporting times between the precipitation products and the ground-based measurements when assessing the performance of precipitation products at a daily scale. Furthermore, it is not recommended to use the RMSE to evaluate the performance of precipitation products at the daily scale due to the high skewness of the precipitation distribution and because it gives more weight to the mismatches between precipitation products and ground-based measurements in two cases: *i*) when there is a systematic underestimation of small precipitation events and *ii*) during high precipitation events because of its squared-error-based nature.

The proposed RF-MEP method to merge precipitation products and ground-based measurements gave outstanding results and provided the best spatio-temporal characterisation of precipitation compared to other precipitation products and merging techniques. This novel method can be applied to improve the spatio-temporal characterisation of precipitation over data-scarce regions. So far, RF-MEP has been

CHAPTER 5. DISCUSSION

applied successfully over Colombia (Rodriguez-Castiblanco et al. 2021), the precipitation product derived over Chile in Baez-Villanueva et al. (2020) was included in the precipitation monitor platform from the Center for Climate and Resilience Research (Mawün), and the method have had a large impact in the hydrological community (e.g., Chen et al. 2021a; Fan et al. 2021; Mekonnen et al. 2021; Rahman et al. 2021; Tang et al. 2021; Wong et al. 2021; Zhang et al. 2021; Sreeparvathy and Srinivas 2022). The RFmerge R package, includes the implementation of RF-MEP and can be found in the following link: <https://CRAN.R-project.org/package=RFmerge>. Future research could compare the performance of RF-MEP with multiple merging techniques (e.g., Li and Shao 2010; Rozante et al. 2010; Woldemeskel et al. 2013; Manz et al. 2016; Ma et al. 2018; Zhang et al. 2021) and analyse the influence of the geographical covariates in the precipitation patterns. Regarding the regionalisation results presented in Section 4.3, the feature similarity procedure could be used to refine the parameter regionalisation results obtained in national scale hydrological characterisations in Chile (e.g., Bambach et al. 2018; Lagos et al. 2019). Additionally, further research could address: *i*) the influence of parameter equifinality in parameter regression, which can be accounted by simultaneously optimising the model parameters and the regression equations as in Beck et al. (2020a); *ii*) the effect of the selection of a particular objective function in the simulation of streamflow-based hydrological signatures (e.g., Pool et al. 2017); *iii*) the use of additional model structures (e.g., Clark et al. 2008; Knoben et al. 2019; Neri et al. 2020); *iv*) the evaluation of other states and fluxes derived from remote sensing data (e.g., Dembélé et al. 2020b); and *v*) the sensitivity of different regionalisation methods with respect to modified climate scenarios.

Despite these recommendations, there is a lack of data governance and infrastructure for engaging a wide range of agencies and stakeholders to improve data quality and usage (Sun et al. 2018). Similarly, there is a need to build the capacities needed to download, process, and translate these datasets into information that can be used in the decision-making process. These capacities should be acquired by local and regional experts that understand the complex challenges of water resources management at the local level and not only by those who are proficient in the implementation of novel techniques and algorithms (Blumenstock 2018). One way to achieve this is to create knowledge networks to link government agencies, data scientists and domain experts, universities and research centres, diverse stakeholders, the private sector, and international development organisations (Sheffield et al. 2018; Sun et al. 2018). Inline to this, and as Sheffield et al. (2018) emphasises, there is the requirement of continuing and building existing training programmes and initiatives such as the [UNESCO Hydrology Initiatives](#); the NASA Applied Remote Sensing Training Program ([ARSET](#)); and the [Advanced Training Course on Land Remote Sensing](#) from the ESA. The new

CHAPTER 5. DISCUSSION

generation of water resources managers and stakeholders must be more proficient in using gridded products and data merging techniques, especially over data-scarce settings. This will help build sustainable solutions to address the complex challenges of water resources management and strive towards informed-based decision-making.

CHAPTER 5. DISCUSSION

Chapter 6

Conclusion

As the majority of the countries are placing unprecedented pressure on water resources, there is a need to understand the water resources dynamics at the catchment scale to achieve water security and sustainable water management. The results presented in this thesis show how precipitation products can be evaluated, corrected, and selected for operational applications over data-scarce settings to strive towards an information-based decision-making process. The main findings of this thesis are summarised in the following points. The items listed below correspond to the specific objectives described in Chapter 1.4.

1. Each precipitation product performed differently over each region and temporal scale used in the evaluation (i.e., daily, monthly, seasonal, and annual). All products presented a higher probability of detecting no rain events than days with precipitation. Despite that the probability of detection was relatively low for days with precipitation, the moderate rain events [5, 20 mm) were better represented in all regions in comparison to light [1, 5mm), heavy [20, 40 mm), and violent rain events (≥ 40 mm). Paraiba do Sul presented the best performance at the daily temporal scale, followed by the Imperial. Both catchments have tropical climates with hot and cold summers, respectively. The Magdalena was the worst-performing catchment at the daily scale. This could be attributed to the fact that Colombia presents more convective precipitation compared to the other study areas (please see Taszarek et al. 2021). The gauge-adjusted products generally showed better performance than those that did not include ground-based measurements. When evaluating the performance of precipitation products that use ground-based measurements to correct their estimates, it is crucial to consider the independence of the selected rain gauge stations to avoid introducing bias in the evaluation. The products performed similarly

CHAPTER 6. CONCLUSION

over different elevations. Additionally, it is worth mentioning that not because a precipitation product has been identified as the best performing, it can be used for operational purposes. This is the case of the Magdalena, where CHIRPSv2 performed the best at the daily temporal scale with relatively low overall performance. These results confirm that a catchment-specific evaluation is still required to select a suitable precipitation product for operational purposes. However, this evaluation should always be performed keeping in mind the final hydrological, ecological, or agricultural application, which is essential as these products perform differently regarding the detection of single precipitation events, precipitation totals, and the distribution of different precipitation intensities. Additionally, in Chapter 3, it was concluded that the difference in reporting times between the precipitation products and the rain gauges must be considered when evaluating the performance of precipitation products at the daily temporal scale. This issue still constitutes a major limitation of most precipitation evaluation studies performed in regions far from 0:00 UTC. Finally, it is worth keeping in mind that despite rain gauges provide accurate precipitation measurements at specific locations, they are still subject to errors such as wind-induced under-catch, wetting and evaporation loss, and trace amounts, which may impact the evaluation of precipitation products.

2. The evaluation of precipitation products can be affected by the upscaling procedures that aim to enable a fair point-to-pixel comparison. These performance differences related to the precipitation products' re-scaling are specifically associated with each precipitation product, the temporal scale, and the physiographic and climatic characteristics of the analysed region. The topography plays an important role in this process as the performance of the re-scaled products was more impacted over areas with a more pronounced elevation gradient. Therefore, it is crucial to consider the effects that the spatial resolution of the products may have over mountainous regions when evaluating their performance over regions with complex topography. Similarly, the application of re-scaling procedures may impact the performance of precipitation products over such settings. On the other hand, if the topography is not rugged, a re-scaling procedure can enable a fair comparison among products.
3. Despite the continuous improvements of the precipitation products, they are still subject to different types of mismatches and errors. In Chapter 3, a novel machine learning procedure was proposed to produce improved precipitation estimates at the daily temporal scale. The datasets generated through the application of the Random Forest Merging Procedure (RF-MEP) showed improved *i*) temporal precipitation dynamics; *ii*) total precipitation volumes; *iii*) relative

dispersion; and *iv*) detection skills of single precipitation events at all temporal scales compared to the individual products used in their development. The performance of this method increases when more rain gauge stations are used; however, it is still able to improve the spatio-temporal characterisation of precipitation even with few stations. RF-MEP outperformed other merging methods such as Kriging with external drift and one-outlier-removed arithmetic mean, and was validated over Chile, which exhibits notable heterogeneity in climate and topography. In this sense, RF-MEP could be successfully applied in other regions and catchments to improve precipitation and other variables.

4. The use of precipitation products to generate accurate streamflow predictions in data-scarce ungauged catchments is crucial for water resources management. In this sense, the relative performance of three common regionalisation techniques (spatial proximity, feature similarity, and parameter regression) was assessed over 100 near-natural catchments located in the topographically and climatologically diverse Chilean territory (see Chapter 4). Feature similarity yielded the best results, regardless of the choice of precipitation product or hydrological regime. The performance of feature similarity increased when four or more catchments were used as donors; however, the differences in performance were not statistically significant compared to the results obtained when only one donor catchment was used. Spatial proximity was the second-best regionalisation technique, which could be related to the fact that spatial proximity is a good proxy of climatic similarity for most neighbouring catchments in Chile. Finally, parameter regression provided the lowest performance, reinforcing the idea of transferring the complete set of model parameters to ungauged catchments. The exclusion of relatively few nested catchments impacted the results of feature similarity and spatial proximity. However, for the particular case of Chile, their exclusion had minimal impact on the non-linear relationships between the climatic and physiographic characteristics and model parameters; therefore, a negligible effect on the parameter regression results.
5. The results of Chapter 4 indicate that the precipitation product that provided the best (worst) performance during the independent calibration and verification did not necessarily yield the best (worst) results during regionalisation. Similarly, the products corrected using daily ground-based measurements did not necessarily yield the best hydrological model performances compared to other good-performing products. The hydrological model was able to compensate, to some extent, for differences between precipitation products by adjusting the model parameters during the calibration process and, therefore, the water balance components.

CHAPTER 6. CONCLUSION

6. The hydrological regime of the catchments influenced the performance of the evaluated regionalisation techniques. The best results were obtained in pluvio-nival catchments for the case of feature similarity, spatial proximity, and during the independent calibration and verification. These results could be attributed to the ability of the hydrological model to reproduce streamflow in this regime and to the increased likelihood of transferring model parameters from catchments with the same hydrological regime as the pluvio-nival catchments consisted of 40% of the total number of catchments and they were grouped close together. Feature similarity and spatial proximity provided the worst results over rain-dominated catchments, which was not the case for the independent calibration and verification. The results demonstrate that no single precipitation product outperformed the others for all regionalisation techniques and hydrological regimes.

The use of gridded datasets can help to strive towards informed-based decision-making. However, data-scarce settings often lack the infrastructure and human capacity to use this type of information efficiently. Therefore, an informed-based decision-making process requires institutional transitions and changes that help address the present and future challenges of water resources management, such as *i*) the projected increase in the frequency and severity of droughts and heatwaves due to climate change; *ii*) the increased water resources stress placed by the overexploitation of surface and groundwater; *iii*) the increased demand for agricultural and energy production; and *iv*) population growth. As observed, merging multiple sources of information is a promising way to derive useful information to strive towards informed-based decision-making. In this sense, there is a need to move towards demand-driven water resources management through the implementation and use of novel data-driven techniques and datasets, which can support the implementation of strategical approaches that systematically build such regions' capacities and infrastructure.

Appendices

Appendix A

Evaluation step

APPENDIX A.

**Temporal and spatial evaluation of satellite rainfall estimates
over different regions in Latin-America**

Oscar Manuel Baez-Villanueva; Mauricio Zambrano-Bigiarini;
Lars Ribbe; Alexandra Nauditt; Juan Diego Giraldo-Osorio; Nguyen
Xuan Thinh

Atmospheric Research, Volume 213, 2018, Pages 34-50.

<https://doi.org/10.1016/j.atmosres.2018.05.011>

Appendix B

Merging step

APPENDIX B.

**RF-MEP – A novel Random Forest method for merging
gridded precipitation products and ground-based
measurements**

Oscar Manuel Baez-Villanueva; Mauricio Zambrano-Bigiarini;
Hylke E. Beck; Ian McNamara; Lars Ribbe; Alexandra Nauditt;
Christian Birkel; Koen Verbist; Juan Diego Giraldo-Osorio; Nguyen
XuanThinh

Remote Sensing of Environment, Volume 239, 2020, Pages 111606.

<https://doi.org/10.1016/j.rse.2019.111606>



Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

RF-MEP: A novel Random Forest method for merging gridded precipitation products and ground-based measurements



Oscar M. Baez-Villanueva^{a,b,*}, Mauricio Zambrano-Bigiarini^{c,d}, Hylke E. Beck^e, Ian McNamara^a, Lars Ribbe^a, Alexandra Nauditt^a, Christian Birkel^{f,g}, Koen Verbist^h, Juan Diego Giraldo-Osorioⁱ, Nguyen Xuan Thinh^b

^a Institute for Technology and Resources Management in the Tropics and Subtropics (ITT), TH Köln, Cologne, Germany

^b Faculty of Spatial Planning, TU Dortmund University, Dortmund, Germany

^c Department of Civil Engineering, Universidad de la Frontera, Temuco, Chile

^d Center for Climate and Resilience Research, Universidad de Chile, Santiago, Chile

^e Department of Civil and Environmental Engineering, Princeton University, Princeton, USA

^f Geography Department, University of Costa Rica, San José, Costa Rica

^g Northern Rivers Institute, University of Aberdeen, Aberdeen, UK

^h UNESCO International Hydrological Programme, Paris, France

ⁱ Pontificia Universidad Javeriana, Bogotá, Colombia

ARTICLE INFO

Edited by Menghua Wang

Keywords:

Bias correction

Merging

Precipitation

Precipitation products

Random Forest

RF-MEP

ABSTRACT

The accurate representation of spatio-temporal patterns of precipitation is an essential input for numerous environmental applications. However, the estimation of precipitation patterns derived solely from rain gauges is subject to large uncertainties. We present the Random Forest based Merging Procedure (RF-MEP), which combines information from ground-based measurements, state-of-the-art precipitation products, and topography-related features to improve the representation of the spatio-temporal distribution of precipitation, especially in data-scarce regions. RF-MEP is applied over Chile for 2000–2016, using daily measurements from 258 rain gauges for model training and 111 stations for validation. Two merged datasets were computed: RF-MEP_{3P} (based on PERSIANN-CDR, ERA-Interim, and CHIRPSv2) and RF-MEP_{5P} (which additionally includes CMORPHv1 and TRMM 3B42v7). The performances of the two merged products and those used in their computation were compared against MSWEPv2.2, which is a state-of-the-art global merged product. A validation using ground-based measurements was applied at different temporal scales using both continuous and categorical indices of performance. RF-MEP_{3P} and RF-MEP_{5P} outperformed all the precipitation datasets used in their computation, the products derived using other merging techniques, and generally outperformed MSWEPv2.2. The merged *P* products showed improvements in the linear correlation, bias, and variability of precipitation at different temporal scales, as well as in the probability of detection, the false alarm ratio, the frequency bias, and the critical success index for different precipitation intensities. RF-MEP performed well even when the training dataset was reduced to 10% of the available rain gauges. Our results suggest that RF-MEP could be successfully applied to any other region and to correct other climatological variables, assuming that ground-based data are available. An R package to implement RF-MEP is freely available online at <https://github.com/hzambran/RFmerge>.

1. Introduction

Precipitation (*P*) is a key parameter in the hydrological cycle and an accurate estimation of its spatio-temporal variability is therefore crucial

for numerous hydrological, agricultural, and ecological purposes. *P* is commonly measured with rain gauge stations, with a high accuracy at specific locations (Villarini et al., 2008). If only ground-based measurements are used, the accuracy of the representation of spatial *P*

* Corresponding author.

E-mail addresses: obaevvil@th-koeln.de (O.M. Baez-Villanueva), mauricio.zambrano@ufrontera.cl (M. Zambrano-Bigiarini), hylke.beck@gmail.com (H.E. Beck), ian.mcnamara@th-koeln.de (I. McNamara), lars.ribbe@th-koeln.de (L. Ribbe), alexandra.nauditt@th-koeln.de (A. Nauditt), christian.birkel@ucr.ac.cr (C. Birkel), k.verbist@unesco.org (K. Verbist), j.giraldo@javeriana.edu.co (J.D. Giraldo-Osorio), nguyen.thinh@tu-dortmund.de (N. Xuan Thinh).

<https://doi.org/10.1016/j.rse.2019.111606>

Received 30 March 2019; Received in revised form 10 October 2019; Accepted 11 December 2019

Available online 02 January 2020

0034-4257/© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Table 1
Studies that have applied merging procedures to improve the spatio-temporal characterisation of *P* at different temporal scales. Those marked with a star have derived merged *P* estimates at daily or sub-daily temporal resolution

Study	Merging method(s)	Region	Product(s)	Number of stations(training / validation)	Main results
Li and Shao (2010)*	Nonparametric kernel smoothing	Australia	TRMM 3B42v6	5007 (90%) / 556 (10%)	TRMM 3B42v6 had an RMSE of 6.04 mm, mean error (ME) of -0.19 mm, and coefficient of efficiency of 0.29. The merged product had an RMSE of 3.43 mm, ME of 0.05 mm, and CE of 0.74
Rozante et al. (2010)*	Barnes objective analysis method	South America	TRMM 3B42RT	~1350 (90%) / ~150 (10%)	The 5-day RMSE for TRMM 3B42RT ranged from ~4 to ~22 mm, and the POD ranged from ~0.63 (no-rain) to ~0.2 (heavy rain). The RMSE of the merged product ranged from ~2 to ~18 mm, while the POD ranged from ~0.75 (no-rain) to ~0.38 (heavy rain)
Xie and Xiong (2011)*	Optimal interpolation	China (validation over South Korea)	CMORPH	2400 (China); 600 (96%) / 28 (4%) (over South Korea)	Bias corrected CMORPH correlations ranged from 0.65 to 0.75, while the derived product correlations ranged from 0.70 to 0.98
Gebregorgis and Hossain (2011)*	Linear weights based on hydrologic model predictability	Mississippi River Basin	TRMM 3B42RT, CMORPH, and PERSIANN-CCS	Calibration and validation through hydrological modelling	Validation over 6 catchments within the Mississippi River Basin. The Nash-Sutcliffe efficiency (NSE) using the products ranged from ~304 to 0.6, while the NSE using the merged product derived using runoff weights ranged from ~0.3 to 0.6
Woldemeskel et al. (2013)	Linearised weighting procedure	Australia	TRMM 3B42v6	207 (90%); 184 (80%) / 23 (10%); 46 (20%, representing 1 of 5 regions)	The RMSE of TRMM 3B42v6 was 27.8 mm and the ME - 3.42 mm. The RMSEs of the merged product were 23.73 mm and 28.24 mm, and the MEs were -2.42 mm and -2.65 mm for 10% and 20% of the stations, respectively
Shen et al. (2014)*	Arithmetic mean and inverse-error-square weighting methods	Tibetan Plateau	CMORPH, PERSIANN, NRL, TRMM 3B42v7 and 3B42RT	~330 stations / no information about stations used for validation	Summer period: merged products showed slight improvements on <i>P</i> products. Winter period: TRMM 3B42v7 outperformed all merged products
Nie et al. (2015)*	Optimal interpolation	China	CMORPH and NCEP	1920 (80%) / 480 (20%)	The RMSEs of the products ranged from 6.05 to 6.68 mm; <i>r</i> from 0.53 to 0.61, and bias from -0.86 to 0.48. The RMSE of the merged product was 4.38 mm, <i>r</i> of 0.78, and bias of -0.02. POD ranged from 0.9 (no-rain) to 0.5 (heavy-rain) and FAR from -0.27 to -0.35, respectively
Fu et al. (2016)	Bayesian model averaging	China	MERRA, Princeton, ERA-Interim, CMAP, NCEP, GPCPv2.2, and GPCCv6	~378 (50%) / ~377 (50%)	The RMSEs of the products ranged from ~105 to ~265 mm and the <i>R</i> ² from ~0.38 to ~0.81, while the merged product showed a RMSE of ~85 mm and an <i>R</i> ² of ~0.89
Manz et al. (2016)	Linear modelling, residual IDW, and Kriging-based methods	Tropical Andes	TRMM 2A25	722 / 1 (Leave one-out cross validation)	The RMSE of the product ranged from ~70 to ~95 mm and the relative bias (RB) from ~37% to ~15%. Ordinary Kriging (OK) performed best with a range of RMSE from ~45 to ~65 mm and relative bias from ~-3% to ~3%
Verdin et al. (2016)	OK and k-nearest neighbour local polynomials	Central America, Colombia and northwestern Venezuela	CHIRP	Not specified / Leave one-out cross validation	The RMSE from CHIRP ranged from 67.6 to 200.1 mm and the bias from ~-49.0% to 25.7%. The OK RMSE ranged from 60.7 to 137.0 mm and the bias from ~-0.1% to 0.8%
Shi et al. (2017)*	Merging weights based on the effective influence radius of rain gauges	Beijing, China	CMORPH	115 stations / validation through hydrological modelling	Streamflow simulation: NSE from 0.38 (CMORPH) to 0.66 (merged product)
Yang et al. (2017)*	Inverse-root-mean-square-error weighting	Chile	PERSIANN-CCS	414 (90%) / 42 (10%)	Performances reported according to 3 geographic regions. The RMSE from PERSIANN-CCS ranged from 3.08 to 7.35 mm; <i>r</i> from 0.66 to 0.70; and bias from -3.28 to 1.11. The RMSE of the merged product ranged from 0.27 to 2.69 mm; <i>r</i> from 0.95 to 0.97; and bias from -0.04 to 0.02
Mia et al. (2018)*	Bayesian model averaging (BMA)	Tibetan Plateau	TRMM 3B42RT, TRMM 3B42v7, CMORPH, and PERSIANN-CDR	200 (93%) / 15 (7%)	The RMSE of the products ranged from 7.0 to 14.1 mm; RB from 72.6 to 130.0%; and the POD ranged from 0.69 (no-rain) to 0.78 (heavy-rain). The merged product presented an RMSE of 6.77 mm; an RB of 70.83%; and an overall POD of 0.90
Beck et al. (2019)*	Weighted averaging with CDF matching	Global (evaluated over the CONUS)	CMORPH, ERA-Interim, GPCC FDR, Gridsat, GSMAp, JRA-55, TRMM 3B42RT, and WorldClim	76,747 <i>P</i> stations and 13,762 streamflow stations / Stage IV dataset	Validated using the NCEP Stage IV dataset as the ground-truth and compared to MSWEPv1, CMORPH, ERA-Interim, and MERRA-2. The modified Kling-Gupta efficiency (KGE') of the products ranged from 0.35 to 0.53. The KGE' of MSWEPv2.2 was 0.70

patterns relies on the density and configuration of the gauge network (Adhikary et al., 2015; Borga and Vizzaccaro, 1997; Chen et al., 2008; Garcia et al., 2008; Goudenhoofdt and Delobbe, 2009; Villarini and Krajewski, 2008). In particular, a high network density is of most importance to capture the spatial distribution of convective events (Garcia et al., 2008).

In many developing countries the network of rain gauges is sparsely distributed; therefore, the use of only ground-based measurements to estimate the spatial distribution of P is subject to large uncertainties (Woldemeskel et al., 2013). Elevation must be considered because of the important role it plays in the P process. In general, higher elevation causes more P (Jaagus et al., 2010), an effect that can be extremely pronounced even over small elevation changes. For example, Bergeron (1960) reported that precipitation rates over small hills were twice the value of the lower areas in a flat region of 30 km², with approximately 50 m elevation difference. In regions with complex topography, P is typically under-represented at higher elevations because most rain gauges are located in lowlands due to accessibility and economical considerations (Derin and Yilmaz, 2014).

Satellite and reanalysis-based P estimates (hereafter P products) provide an unprecedented opportunity to estimate the spatio-temporal distribution of P in regions with a sparse network of rain gauge stations. However, the evaluation of these products has shown that multiple sources of errors are still present (e.g., false detection, systematic, and random errors) and that these products tend to perform worse at shorter time scales (e.g., daily and sub-daily) than at longer time scales (e.g., monthly, seasonal, and annual), making their application difficult for hydrological modelling (Maggioni and Massari, 2018). Therefore, a need remains to improve the spatio-temporal distribution of P by combining different data sources such as P products and ground-based information (Xie and Xiong, 2011).

Several approaches have been implemented to derive gridded P and other climatological variables using point-based information and gridded products. These include optimal interpolation (OI) (Xie and Xiong, 2011), the linearised weighting procedure (Woldemeskel et al., 2013), non-parametric kernel smoothing (Li and Shao, 2010), Kriging-based methods (Seo et al., 1990; Grimes et al., 1999; 1990; Verdin et al., 2016), conditional merging (Sinclair and Pegram, 2005), partial thin plate splines (Hutchinson, 1995; McKenney et al., 2006; McVicar et al., 2007), among others. Table 1 lists merging studies used to improve the characterisation of P , with a more detailed description of the steps employed in each method included in the Table A1 from Appendix A.

Despite the improvements in the spatio-temporal representation of P achieved by these methods, many studies only merge the ground observations with a single P product (e.g., Li and Shao, 2010; Rozante et al., 2010; Shi et al., 2017; Verdin et al., 2016; Xie et al., 2017; Yang et al., 2017). Therefore, valuable information that is better captured by other products is not considered. Averaging P products (e.g., Shen et al., 2014) has negative effects in the detection of P intensities at daily temporal scale. The assumption of a Gaussian distribution is invalid for daily scales; therefore, the daily P data must be first transformed when using Bayesian model averaging (Ma et al., 2018) or Kriging-based approaches. Furthermore, these merging methods are generally complex and difficult to implement.

Random Forest (RF; Biau and Scornet, 2016; Breiman, 2001; Prasad et al., 2006) is an ensemble learning method that can be used for supervised classification and regression tasks by constructing numerous decision trees using the relationship between independent and dependent variables. This technique is recognised for being accurate and able to deal with small sample sizes and high-dimensional feature spaces (Biau and Scornet, 2016). RF also performs well even when some explanatory variables do not add information to the prediction and when several covariates are used, mainly because it does not produce biased estimates or lead to overfitting (Biau and Scornet, 2016; Díaz-Uriarte and Alvarez de Andrés, 2006; Hengl et al., 2018). Although RF is a non-

spatial technique, it can indirectly consider geographical covariates (e.g., coordinates, Euclidean distances to sampling locations, or down-slope distances) and process-based covariates (e.g., elevation, rate of elevation change, or aspect).

Recently, Hengl et al. (2018) compared RF and several Kriging-based methods to evaluate whether RF was suitable for deriving spatial predictions of daily P . Although the performances of both methods were similar, they described several advantages in applying RF: *i*) there is no need to define an initial variogram; *ii*) the trend model is built automatically; *iii*) there is no need to define a search radius; *iv*) there are built-in protections against overfitting; and *v*) the method shows which individual observations and parameters are most influential. Therefore, RF is identified as an appropriate technique for merging P products with ground-based information, especially because different P products exhibit distinct performances and errors (e.g., under/overestimation, correlation with ground-based measurements, or detection of P events) depending on the region (Baez-Villanueva et al., 2018; Maggioni and Massari, 2018; Zambrano-Bigiarini et al., 2017).

In this study, the RF-based MERging Procedure (RF-MEP) is presented with the aim of improving the characterisation of the spatio-temporal distribution of P in data-scarce regions at any temporal scale. RF-MEP takes advantage of combining information from different P products, topography-related datasets, and P time series from rain gauges.

2. RF-MEP

RF-MEP is based on three key assumptions: *i*) P measurements from rain gauge stations are accurate at the point scale; *ii*) P products are generally biased but contain useful information about the spatio-temporal patterns of P ; and *iii*) the combination of different P products and rain gauge data can provide a better representation of the spatio-temporal variability of P than any single product.

RF-MEP uses RF to predict the spatial distribution of P by merging information from different gridded products (known as covariates) and quality-controlled ground-based information at a selected temporal scale (e.g., daily, monthly, or annual). Individual predictions are generated from a user-defined number of decision trees based on bootstrap samples using the covariates as predictors. The final prediction is calculated as the average of the individual predictions (Biau and Scornet, 2016; Breiman, 2001; Hengl et al., 2018; Prasad et al., 2006; Roy and Larocque, 2012). Fig. 1 summarises the four steps involved in this method.

2.1. Data acquisition

First, the selected covariates and ground-based measurements are acquired. The spatial covariates are: *i*) the selected P products, and *ii*) topography-related datasets such as digital elevation model (DEM), aspect, rate of elevation change, or slope, which are used to account for the P gradient related to elevation (not to be mistaken with altitude, see McVicar and Körner, 2013). The ground-based measurements are quality controlled and checked for homogeneity.

2.2. Data processing

The selected rain gauge stations are divided into two groups: a training set (to train the RF model) and a validation set (to assess the performance of the merged product). The selected P products and topography-related datasets are resampled to a selected spatial resolution to ensure identical raster geometry (spatial resolution, spatial extent, and origin).

The traditional RF algorithm ignores sampling locations which could lead to sub-optimal predictions (Hengl et al., 2018); therefore, covariates that account for geographical proximity are incorporated. The use of only geographical coordinates as spatial predictors can cause

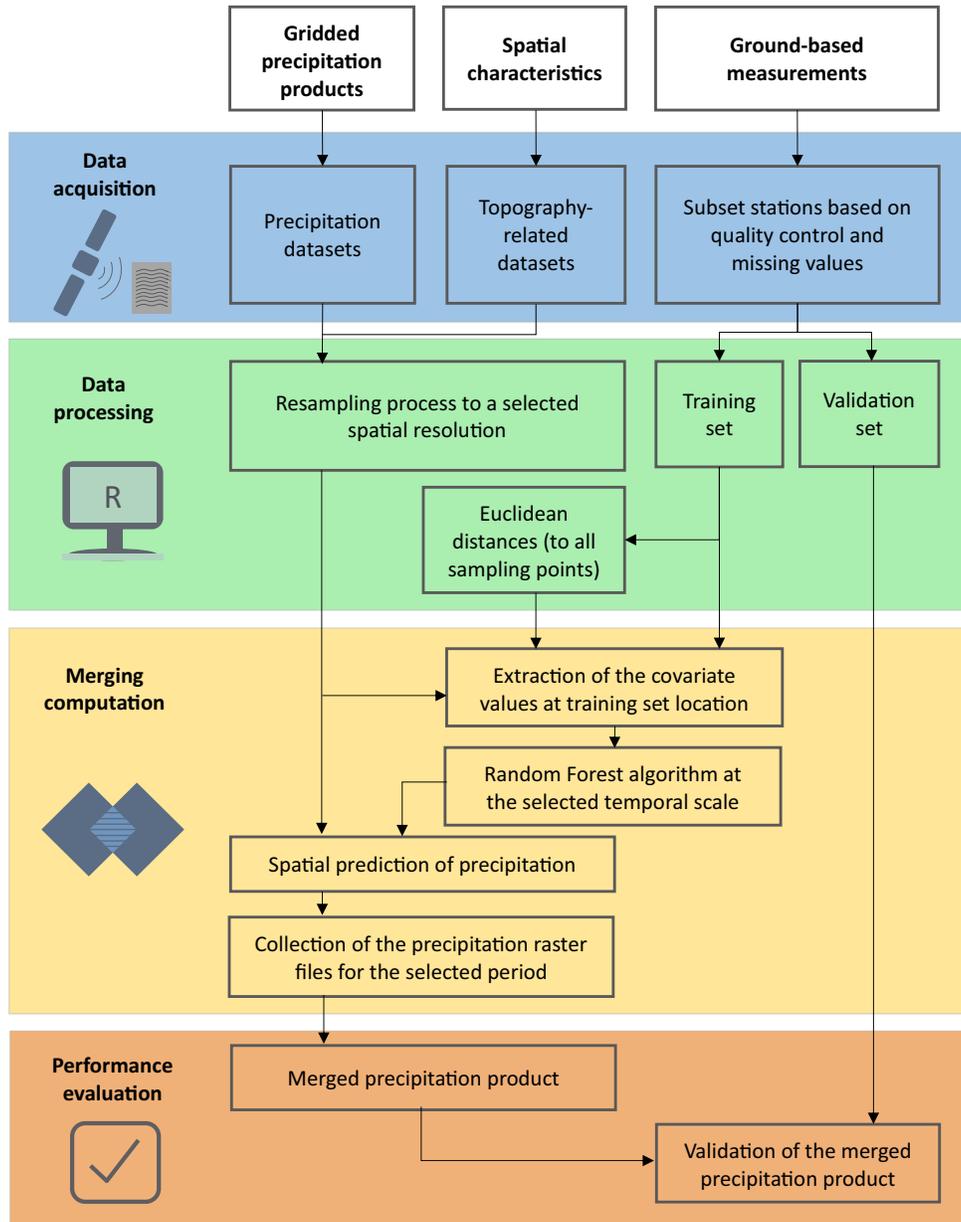


Fig. 1. Flow chart summarising RF-MEP, which is used to derive a better representation of the spatio-temporal distribution of P from the combination of P products, topography-related datasets, and ground-based data.

unnatural surfaces in the merged product (Behrens et al., 2018; Hengl et al., 2018). Instead, RF-MEP uses gridded layers of Euclidean distances from each rain gauge in the training set to the centroid of all the grid-cells in the selected study area.

2.3. Merging procedure

For each time step a single RF regression model is derived to compute a single P prediction at the desired temporal resolution. The RF model is trained using the ground-based observations in the training set as the dependent variable, while the grid-cell values of the selected covariates at the corresponding locations are used as predictors. To improve the accuracy and stability, and to reduce the variance and

overfitting of the RF predictions, they are generated as an ensemble estimate from the numerous decision trees (Díaz-Urriarte and Alvarez de Andrés, 2006; Hengl et al., 2018) as observed in Eq. (1):

$$\hat{\theta}^B(x) = \frac{1}{B} \sum_{b=1}^B t_b^*(x) \quad (1)$$

where $\hat{\theta}^B$ is the final prediction; b is the individual bootstrap sample; B is the total number of trees; and t_b^* is the individual decision tree. This process is repeated for each time step, implying that the RF model will vary temporally. Fig. 2 illustrates an example of the merging procedure using two P products, a digital surface model (DSM), three rain gauge stations, and the three correspondent Euclidean distance layers

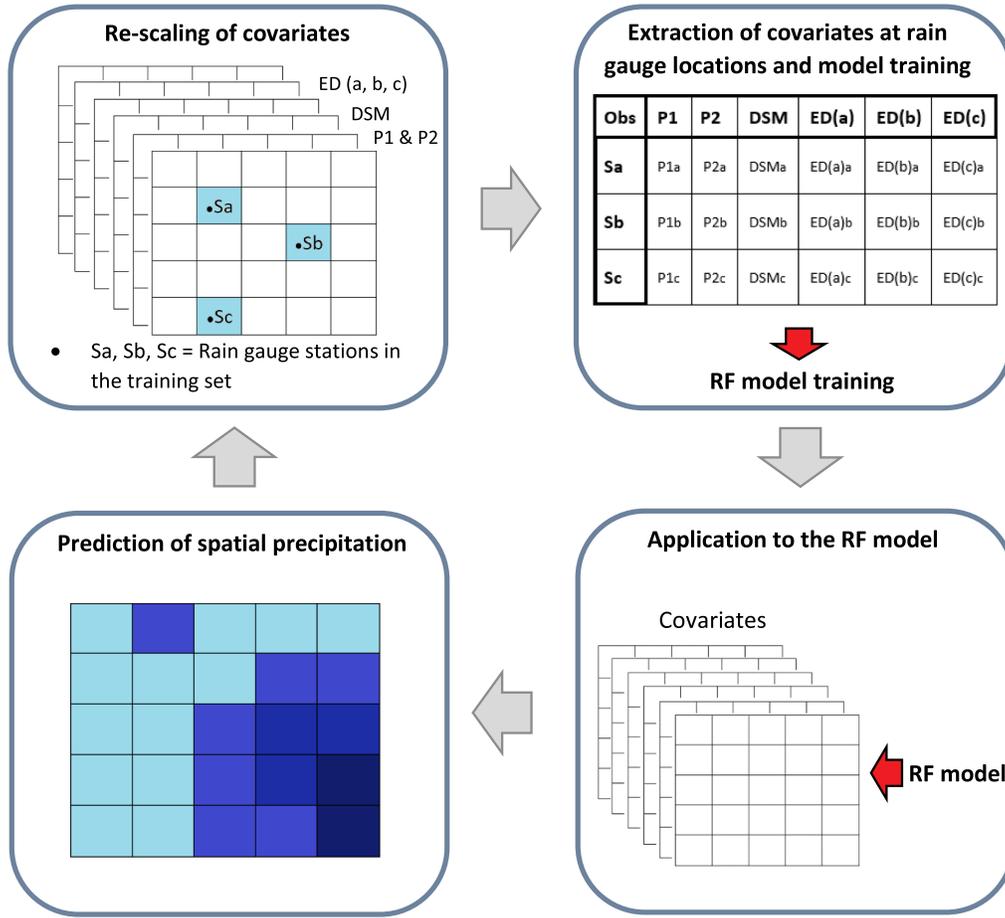


Fig. 2. Illustration of the merging procedure using two P products (P1 and P2), a DSM (to account for the topography-related datasets), three rain gauge stations (S_a , S_b , and S_c), and the three Euclidean distance layers ($ED_{(a)}$, $ED_{(b)}$, and $ED_{(c)}$).

($ED_{(a)}$, $ED_{(b)}$, and $ED_{(c)}$). An R package to implement RF-MEP is freely available online at <https://github.com/hzambran/RFmerge>.

2.4. Validation process

The validation set of rain gauge stations is used to assess the performance of the merged product using a point-to-pixel analysis, where the rain gauge station measurements are compared against the corresponding grid-cell values of the P products under the assumption that the rain gauge measurements are representative values at their respective grid-cells. However, this assumption may introduce bias in the comparison because: *i*) during winter, some rain gauges located at high elevation are not able to incorporate snow into the P measurement; and *ii*) during summer, a more dense network of rain gauges is required to capture the spatial patterns of small-scale convective events. Despite this, the point-to-pixel analysis is widely used to assess the performance of P products (e.g., Baez-Villanueva et al., 2018, Dinku et al., 2007, Gao and Liu, 2013, Hirpa et al., 2010, Li et al., 2013, Thiemig et al., 2012, Zambrano-Bigiarini et al., 2017). Among the plethora of indices available to assess the performance of P products, we selected the modified Kling-Gupta efficiency (KGE'; Gupta et al., 2009, Kling et al., 2012) over the traditional root mean squared error (RMSE) because the latter assigns disproportional weights to different P intensities at the daily scale (Baez-Villanueva et al., 2018). This is due to the high skewness of the precipitation distribution at the daily scale and the prevalence of

temporal mismatches between estimated and observed precipitation peaks. The KGE' (Eq. 2) compares observed data with estimations, decomposing the total performance into three components: the linear correlation (r), the bias ratio (β), and the variability ratio (γ), presented in Eqs. (3), (4), and (5), respectively:

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

$$r = \frac{\sum_{i=1}^n (O_i - \bar{O})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}} \quad (3)$$

$$\beta = \frac{\mu_s}{\mu_o} \quad (4)$$

$$\gamma = \frac{CV_s}{CV_o} = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} \quad (5)$$

where n is the number of observations; O_i and S_i are the observed and simulated values of the corresponding P product at day i ; and \bar{O} and \bar{S} are the arithmetic means of the observations and the P product, respectively. r measures the temporal P dynamics; β measures the total P volume compared to ground-based observations indicating the average tendency of the P products to underestimate ($\beta < 1$) or overestimate ($\beta > 1$); and γ measures the relative dispersion between the gridded product and the ground-based measurements (Gupta et al., 2009; Kling et al., 2012). The optimal value for the KGE' and all its components is

one. The KGE' is a useful evaluation index because: *i*) it does not assign disproportional weights to mismatches in high precipitation values (contrary to squared-difference indices; e.g., the RMSE); *ii*) it decomposes the total performance into three components, thus allowing a better understanding of the origin of mismatches (Baez-Villanueva et al., 2018; Zambrano-Bigiarini et al., 2017); and *iii*) it allows a fair comparison of regions with different mean annual P . The KGE' has been widely used in hydrological applications and to evaluate the performance of P products (e.g., Baez-Villanueva et al., 2018, Beck et al., 2016, 2017b, Chen et al., 2014, Lievens et al., 2015, Thiemig et al., 2013, Wang et al., 2018, Zambrano-Bigiarini et al., 2017).

To evaluate the performance of P products in capturing different P intensities we used several categorical indices of performance: the probability of detection (POD; Eq. (6)), frequency bias (f_{bias} ; Eq. (7)), false alarm ratio (FAR; Eq. (8)), and critical success index (CSI; Eq. (9)).

$$POD = \frac{H}{H + M} \quad (6)$$

$$f_{bias} = \frac{H + F}{H + M} \quad (7)$$

$$FAR = \frac{F}{H + F} \quad (8)$$

$$CSI = [(POD)^{-1} + (1 - FAR)^{-1} - 1]^{-1} \quad (9)$$

where H indicates a hit (an event recorded by both the rain gauge and the P product); M indicates a miss (an event only identified by the rain gauge); and F indicates a false alarm (an event recorded only by the P product). The POD calculates how often the product correctly estimates the precipitation intensity observed at the rain gauge. The f_{bias} compares the number of events identified by the P product to the number of events registered by the gauge station. If $f_{bias} > 1$, the number of occurrences of the respective P intensity is overestimated by the product, while $f_{bias} < 1$ indicates underestimation. The FAR measures the fraction of events that were not correctly identified by the P product. Finally, the CSI combines the POD and FAR to describe the overall ability of the products to correctly detect different P intensities. The POD, f_{bias} , and CSI present their optimal value at one, while FAR presents it at zero.

3. Case study

The Chilean territory was selected as the case study to test the performance of the proposed RF-MEP due to the notable heterogeneity in topography, climate and land cover.

3.1. Study

Chile is a South American country with nearly 4300 km of latitudinal extension (from 17.5°S to 56.0°S) and an average longitudinal extension of around 180 km (from 76.0°W to 66.0°W). Chile is bounded to the north by Peru, to the east by Bolivia and Argentina, and to the west by the Pacific Ocean. The geography of the country is dominated by mountainous terrains, with an elevation profile ranging from 0 to 6891 m a.s.l. Morphologically, Chile exhibits four major geographical units distributed from east to west: the Andes Mountains, the Intermediate Depression, the Coastal Mountains, and the Coastal Plains (Valdés-Pineda et al., 2014). The four seasons of the southern hemisphere are present: autumn (MAM), winter (JJA), spring (SON), and summer (DJF). P tends to increase with latitude (in the southern direction) and elevation (Montecinos and Aceituno, 2003). The inter-annual variability of P is mostly related to the El Niño-Southern Oscillation (ENSO), which strongly impacts winter P patterns, generating positive anomalies during El Niño events and negative anomalies during La Niña events (Robertson et al., 2014; Verbist et al., 2010).

Fig. 3 shows the elevation (Jarvis et al., 2008), the Köppen-Geiger

climate zones (Beck et al., 2018), and the most updated Chilean land cover classification (Zhao et al., 2016), dividing the country according to the five major macroclimatic zones defined in Zambrano-Bigiarini et al. (2017). A variety of climates are observed throughout Chile: arid and semi-arid climates in the north with extremely low P (≤ 50 mm yr^{-1}) and high temperatures; temperate climates in Central Chile; and humid climates in the southern regions, with P values reaching up to 5000 mm yr^{-1} . Furthermore, polar and tundra climates are observed in the highest elevations of the Andes Mountains. Land cover is characterised by barren land in the Far North, which transitions to forest in the Near North. Forest, grasslands, and croplands are present in Central Chile and the two southern regions, while grassland, forest, and snow/ice areas are predominantly observed in the Far South.

3.2. Datasets

3.2.1. Ground-based precipitation

Time series of ground-based daily P for 1900–2018 were downloaded from a database of 816 rain gauges from the Center of Climate and Resilience Research (CR2; http://www.cr2.cl/recursos_y_publicaciones/bases-de-datos/). These data are provided by Dirección General de Aguas (DGA) and Dirección Meteorológica de Chile (DMC), the Chilean water and meteorological agencies, respectively. In Chile, daily P is recorded at 08:00 local time (11:00–10:59 UTC).

3.2.2. SRTM-v4

We used the Shuttle Radar Topography Mission version 4 (SRTM-v4) DSM, which incorporates offsets due to vegetation height (Gallant et al., 2012), and has a reported vertical error of less than 16 m (Jarvis et al., 2008). We used the gap-filled SRTM-v4 product at a spatial resolution of 250 m.

3.2.3. Precipitation products

We selected six global or quasi-global state-of-the-art P products with at least 15 years of daily estimates (Table 2). These products were selected because: *i*) RF-MEP can be transferred to any selected study area using the same P products (or others) if ground-based data are available; and *ii*) the selected P products perform well in the study area (Baez-Villanueva et al., 2018; Zambrano-Bigiarini, 2018; Zambrano-Bigiarini et al., 2017).

The selected P products used in RF-MEP were: ERA-Interim (Dee et al., 2011); the Climate Hazards InfraRed Precipitation with Stations data version 2.0 (CHIRPSv2; Funk et al., 2015); the TRMM Multi-satellite Precipitation Analysis (TRMM 3B42v7; Huffman et al., 2010, 2007); the Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks - Climate Data Record (PERSI-ANN-CDR; Ashouri et al., 2015, Sorooshian et al., 2000); and the Climate Prediction Center (CPC) Morphing technique version 1.0-BLD, gauge-satellite blended precipitation product (CMORPHv1; Joyce et al., 2004, Xie et al., 2017). The Multi-Source Weighted-Ensemble Precipitation (MSWEPv2.2; Beck et al., 2017a, 2019) was only used in the validation step as a benchmark product because: *i*) it is the first fully global P dataset derived by optimally merging a range of gauge, satellite, and reanalysis estimates (Beck et al., 2019); *ii*) it has shown more realistic spatial P patterns in mean, magnitude, and frequency than other state-of-the-art global precipitation products at the global scale (Beck et al., 2017b, 2019); *iii*) it uses the same rain gauge dataset within Chile; and *iv*) it recently outperformed other state-of-the-art P products over Chile (Zambrano-Bigiarini, 2018). Detailed descriptions of the algorithms used by each P product can be found in their corresponding literature (see Table 2).

It is important to note that several P products use ground-based P data from the Global Precipitation Climatology Centre (GPCC; Schneider et al., 2008) to reduce bias (see Table 2). The number of operational GPCC rain gauge stations in Chile has fluctuated between seven and twenty over 1986–2018. This low density of GPCC stations

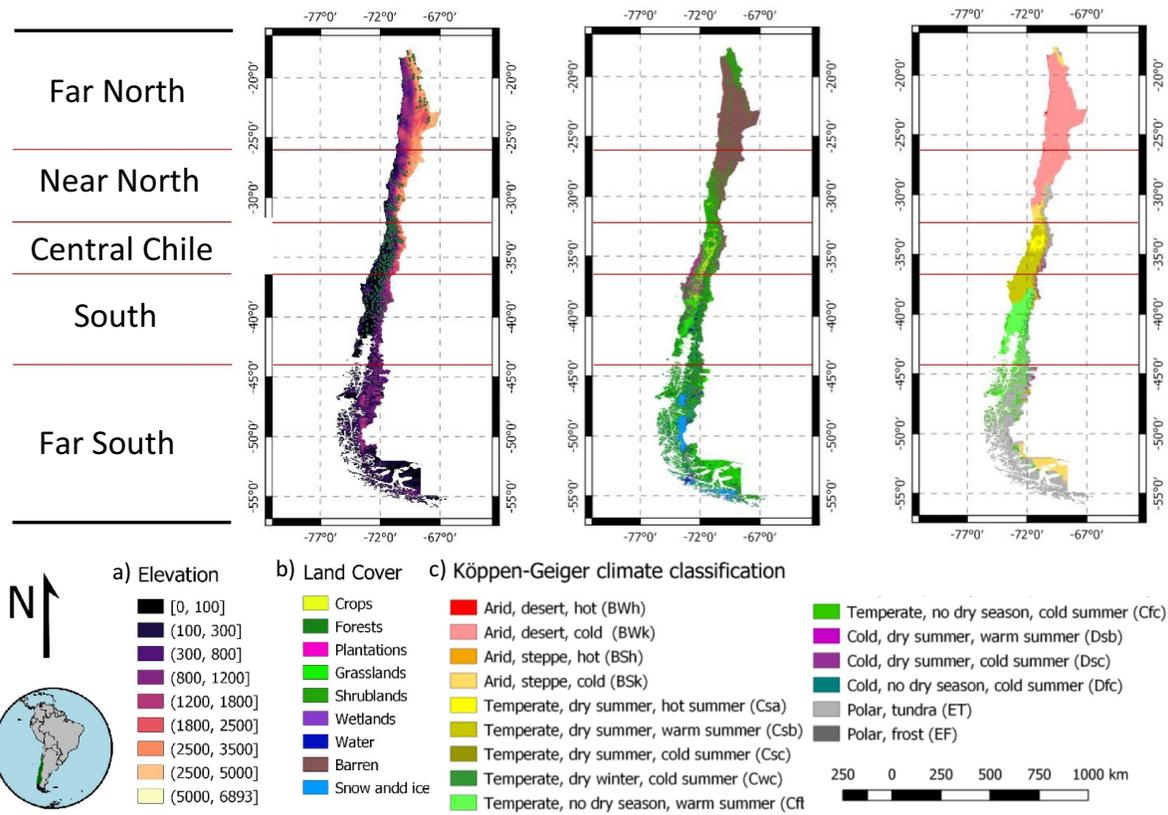


Fig. 3. Study area. (a) Elevation (Jarvis et al., 2008), including the rain gauge stations used in this case study; (b) land cover classification (Zhao et al., 2016); and (c) climate zones based on the Köppen-Geiger classification (Beck et al., 2018).

within Chile is clearly insufficient to adequately represent the spatio-temporal variability of *P* over the country.

3.2.4. Covariates

In addition to the selected *P* products and the DSM, other spatial covariates (slope, aspect, Köppen-Geiger climate classification, land cover type) were exhaustively evaluated using the KGE' and its components to ascertain whether an improvement could be obtained. Only the DSM was selected because the inclusion of the other covariates did not improve the performance of the final product.

3.3. Application of RF-MEP to the study area

RF-MEP was applied to the Chilean territory from 17.5° to 46.0°S for 2000–2016. The southern boundary was set due to the sparse network of gauge stations in the Far South. We used the R environment 3.5.0 (R Core Team, 2018) and the raster (Hijmans, 2018), hydroGOF (Zambrano-Bigiarini, 2017a), hydroTSM (Zambrano-Bigiarini, 2017b),

GSIF (Hengl, 2019), and randomForest (Liaw and Wiener, 2002) R packages.

3.3.1. Data processing

All selected *P* products that are sub-daily (Table 2) were aggregated to the daily scale. MSWEPv2.2 was obtained at daily temporal scale because the 3-hourly version is not freely available. We downscaled PERSIANN-CDR, ERA-Interim, CMORPHv1 and TRMM 3B42v7 to the same spatial resolution as CHIRPSv2 (0.05°) using the nearest neighbour method (to avoid any improvements in the products performance prior to the merging procedure), while the DSM was upscaled from its original spatial resolution (250 m) to 0.05° using bilinear interpolation. The reason for resampling all the covariates to 0.05° (the highest spatial resolution of the selected *P* products) was to obtain a merged product that can be fairly compared to all selected *P* products.

We selected the 369 rain gauge stations that had < 5% of missing values and showed consistency when evaluated using the double-mass curve method to identify abnormalities comparing each station with the

Table 2
P products used in the case study.

Product	Spatial res.	Temporal res.	Period	Spatial coverage	Source(s)	Reference(s)
ERA-Interim	0.75°	3 hourly	1979–present	Global	Reanalysis	Dee et al. (2011)
CHIRPSv2*	0.05°	Daily	1981–present	50°N – 50°S	Satellite, gauge, and reanalysis	Funk et al. (2015)
TRMM 3B42v7*	0.25°	3 hourly	1998–present	50°N – 50°S	Satellite and gauge	Huffman et al. (2010, 2007)
PERSIANN-CDR*	0.25°	6 hourly	1983–2017 (April)	60°N – 60°S	Satellite and gauge	Ashouri et al. (2015), Sorooshian et al. (2000)
CMORPHv1*	0.25°	30 min	1998–present	60°N – 60°S	Satellite and gauge	Joyce et al. (2004), Xie et al. (2017)
MSWEPv2.2*	0.10°	3 hourly	1979–present	Global	Satellite, gauge, and reanalysis	Beck et al. (2017a, 2019)

* Products that use GPCP data.

Table 3

Classification of P events in Chile based on daily intensity (i) according to Zambrano-Bigiarini et al. (2017).

Precipitation event	Intensity (i) in mm d^{-1}
No rain	[0, 1)
Light rain	[1, 5)
Moderate rain	[5, 20)
Heavy rain	[20, 40)
Violent rain	≥ 40

neighbouring stations, assuming homogeneity (Weiss and Wilson, 1953). The period 2000–2016 was chosen because of ground-based data availability over the period of record of the selected P products. A random sample of 70% of the selected rain gauge stations (258) were used as ground truth data to train the RF model (training set), while the remaining 30% of the stations (111) were used to assess the performance of the merged products (validation set). Past studies have typically selected 80% or more stations for training purposes (e.g., Li and Shao, 2010, Ma et al., 2018, Rozante et al., 2010, Woldemeskel et al., 2013, Yang et al., 2017); however, we selected 70% to be more thorough in the evaluation of the method. We computed the 258 layers of Euclidean distances using the GSIF R package (Hengl, 2019).

3.3.2. Merging procedure

Two merged P products were computed at the daily scale for 2000–2016. The first product (hereafter, RF-MEP_{3p}) used CHIRPSv2, PERSIANN-CDR, ERA-Interim, the DSM, and the 258 layers of Euclidean distances, while the second product (hereafter, RF-MEP_{5p}) added CMORPHv1 and TRMM 3B42v7 to the aforementioned covariates. The reason for computing two different merged products was to evaluate whether the addition of CMORPHv1 and TRMM 3B42v7, both of which have a shorter period of temporal coverage, would improve the final merged product. Although RF-MEP_{3p} and RF-MEP_{5p} were produced and compared over the same period (2000–2016), RF-MEP_{3p} can be generated over a longer period of record (1983–2016), while RF-MEP_{5p} can only be generated from 1998 onwards.

First, we obtained the values of the covariates at the grid-cell locations of the training set. Second, for each day, an RF model was trained using the ground-based P values as the dependent variable, and the respective values from the covariates as predictors. Third, the trained RF model was used with the gridded covariates to predict daily P values for each grid-cell of the study area. This process was repeated for each day for 2000–2016. RF regression models have three parameters to specify: *i*) the number of regression trees (set at 2000); *ii*) the number of variables randomly sampled at each decision split (set at one third of the number of covariates); and *iii*) the node size (i.e., the minimum number of observations per node; set at 5).

3.3.3. Performance evaluation

We evaluated the performance of both merged products, MSWEPv2.2, and the individual P products used as covariates, through a point-to-pixel analysis with the indices of performance described in Section 2.4, applied for the stations included in the validation set. The evaluation process was performed at multiple temporal scales: 3-day, monthly, annual, DJF, MAM, JJA, and SON.

Because no sub-daily measurements are available to transform the ground-based P dataset (see Section 3.2.1) to the 0:00–23:59 UTC daily period used by all the P products, we used 3-day accumulations as a proxy for evaluating daily performance. This approach reduces likely biases in the performance of the P products at this temporal scale by considering the influence of reporting times.

The categorical indices were evaluated using P intensities (Table 3; Zambrano-Bigiarini et al., 2017) recommended specifically for Chile.

Because the aim of RF-MEP is to improve the characterisation of P in data-scarce regions, we investigated the influence of the amount of rain

gauge stations included in the training set. We computed the RF-MEP_{5p} product with varying percentages of rain gauge stations in the training set to evaluate the performance of RF-MEP under different data-scarcity scenarios. We computed the RF-MEP_{5p} product using 50%, 30%, and 10% of the stations, representing 184, 111, and 37 rain gauges, respectively.

To test the influence of the different spatial resolutions of the selected P products, we computed RF-MEP_{5p} at 0.05°, 0.10°, and 0.25°. For this purpose, all covariates were resampled to these spatial resolutions before the application of the merging procedure. Finally, we applied two additional merging methods to compare RF-MEP against established and proven precipitation merging procedures. We computed Kriging with external drift (KED) using ERA-Interim (the best-performing product used to derive RF-MEP_{5p}) and the one-outlier-removed (OOR) arithmetic mean described in Shen et al. (2014). For a detailed explanation of KED please refer to Ly et al. (2011), Oliver and Webster (2014), and Hengl et al. (2018). We also compared the RF-MEP_{5p} against MSWEPv2.2 because it is a state-of-the-art merged P product.

4. Results

4.1. Temporal assessment of the merged products

Fig. 4 plots the KGE' values at the seven assessed temporal scales for the existing and merged P products using the ground-based validation set. Both merged products (RF-MEP_{3p} and RF-MEP_{5p}) performed similarly well, with median KGE' values of 0.83, 0.84 and 0.78 at the 3-day, monthly, and annual scale, respectively. The P products used in the merging method presented median KGE' values between 0.20 and 0.60 at the 3-day scale, which increased to between 0.35 and 0.70 at the monthly and annual scales. Both merged products outperformed the P products used in the merging procedure at all temporal scales, demonstrating that the combination of P products and ground-based measurements generates a better representation of the spatio-temporal variability of P .

The merged products performed better than MSWEPv2.2 at all temporal scales except DJF (summer), where all P products showed a reduced performance and a greater dispersion in the KGE' values. This low performance in summer is the reason why the P products exhibit lower KGE' values at the annual scale compared to the monthly scale.

Fig. 5 shows boxplots with the individual KGE' components (r , β , and γ) at all temporal scales. Both merged products present a median r value of 0.94 at the 3-day temporal scale, which is consistent with the improvements in r obtained by Xie and Xiong (2011) and Yang et al. (2017). Of the existing P products, MSWEPv2.2 performed best with a median value of 0.89, highlighting the advantage of merging gauge, satellite, and reanalysis products. At all time scales, RF-MEP_{3p} and RF-MEP_{5p} performed considerably better than the products used in their computation. This demonstrates that the method is able to substantially improve the correlation of the P products for the Chilean case study.

Fig. 5b plots the performance of the β component of the KGE', showing that RF-MEP_{5p}, RF-MEP_{3p}, MSWEPv2.2, and CHIRPSv2 were close to exhibiting no bias. Both merged datasets present lower dispersion than MSWEPv2.2 and CHIRPSv2 for all temporal scales except DJF. This result shows that the evaluated products are generally biased but contain useful information that can be combined with ground-based measurements to derive improved P estimates. In DJF, both merged products presented a $\beta > 1$ and were outperformed by MSWEPv2.2.

Fig. 5c shows the γ component of the KGE', highlighting that all datasets underestimated the variability of P at all temporal scales. MSWEPv2.2 best represented the variability of the ground-based measurements, followed closely by both merged datasets. The high values of γ obtained for MSWEPv2.2 were expected because this product uses the same daily ground-based Chilean dataset in its computation and accounts for the difference in reporting times. Both merged products

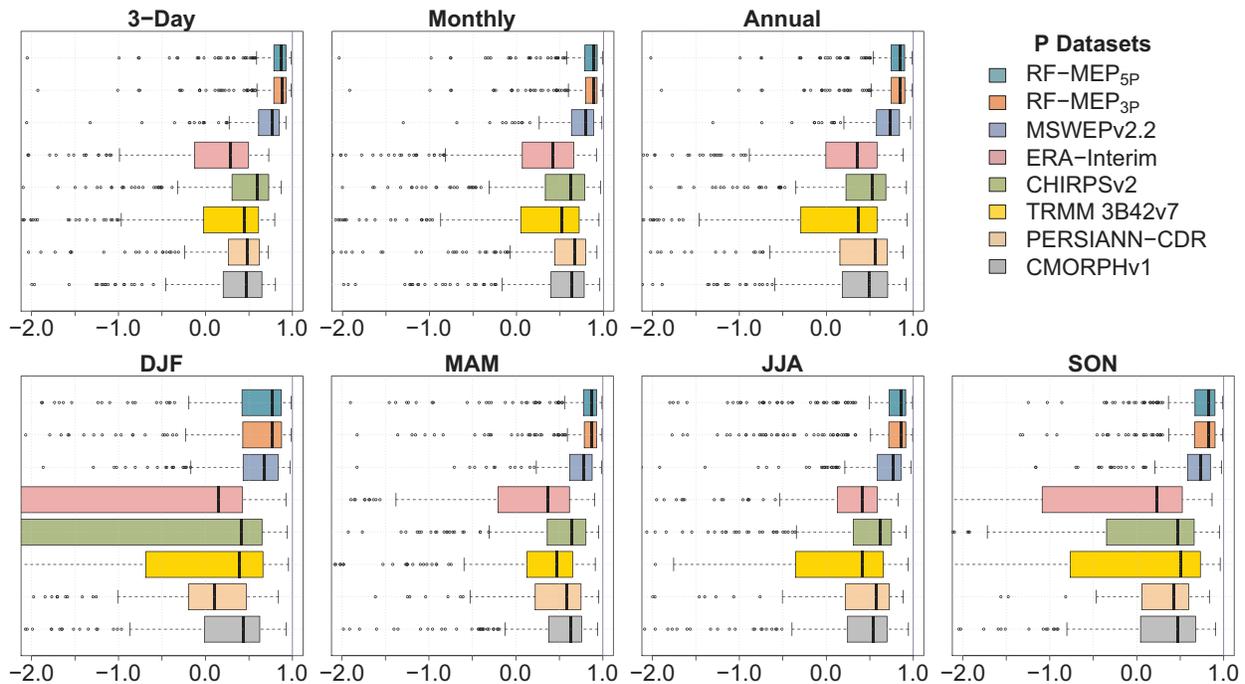


Fig. 4. KGE' values for all P products using the ground-based validation set. From left to right and top to bottom: 3-day, monthly, annual, DJF, MAM, JJA, and SON. The solid line represents the median value, the edges of the boxes represent the first and third quartiles, and the whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box. The blue line indicates the optimal value for the KGE'.

showed a reduced dispersion of the KGE' components at the 3-day, monthly, MAM, JJA, and SON scales; however, the dispersion at the annual scale increases due to the reduced performance in DJF.

4.2. Spatial assessment of the merged products

Fig. 6 summarises the KGE' of the 3-day P products over the four analysed macroclimatic zones, while Fig. 7 presents its spatial distribution. All products show median KGE' values lower than 0.5 and high dispersion in the Far North. These regions are classified as arid according to the Köppen-Geiger classification (see Fig. 3), demonstrating that the performance of the evaluated products over the arid regions of Chile remains low. MSWEPv2.2 and both merged products perform considerably better than the products used as covariates, highlighting the benefit of combining data from P products and ground-based measurements. The performance of all the products increased over Central Chile and South, where annual P volumes are much higher than in the Far North and Near North.

Fig. 7 shows that for both merged products, more than 80% of the stations in the validation set yielded KGE' values higher than 0.60. Both merged products performed best in the Near North, Central, and Southern Chile, with median KGE' values of 0.84, 0.86, and 0.81, respectively. However, in the Far North, MSWEPv2.2 performed the best (0.61), followed by RF-MEP_{3P} (0.35) and RF-MEP_{5P} (0.28). These results in the Far North show that the inclusion of more P products does not necessarily improve the median performance of the merged product; however, the inclusion of the additional two products reduced the dispersion in the KGE' values of RF-MEP_{5P}. Despite the poor performance of the P products used as covariates in the Far North, RF-MEP_{3P} and RF-MEP_{5P} were able to extract useful information from these products to obtain a better performance. RF-MEP_{5P} and RF-MEP_{3P} performed better in the high elevations of the Far North region compared to the low elevations (see Figs. 3 and 7). These high elevations

correspond to the alpine tundra climate (ET), while the cold and arid desert climate (BWk) dominates the lower areas of the Far North, where the P datasets presented their worst performance. This suggests that arid climates present a great challenge for existing P products.

4.3. Assessment of precipitation intensities

Fig. 8 plots the median values of the four categorical indices for the five classes of daily P intensity described in Table 3. All datasets, with the exception of RF-MEP_{3P} and RF-MEP_{5P}, obtained POD values lower than 0.45 for P events higher than 1 mm, while the no-rain events were well captured by all products. Similar results were observed for the FAR and CSI, where RF-MEP_{3P} and RF-MEP_{5P} presented the best performance of the evaluated products. FAR values were consistently the worst for the light rain intensities ([1, 5) mm d⁻¹), highlighting that the products remain unable to adequately capture low P values. The CSI presents the best performance for no-rain events followed by extreme events (≥ 40 mm d⁻¹), as a result of the decreased FAR compared to the other P intensities.

Finally, the median values of the fbias showed that all P products overestimated the number of light rain ([1, 5) mm d⁻¹) and moderate rain events ([5, 20) mm d⁻¹). RF-MEP_{3P} and RF-MEP_{5P} performed the best in terms of fbias for the heavy rain events ([20, 40) mm d⁻¹), while MSWEPv2.2 performed the best for the other P intensities, followed by the merged products. All products underestimated the occurrence of violent rain events (≥ 40 mm d⁻¹).

4.4. Impact of gauge density and spatial resolution of covariates

Fig. 9 shows the performance of RF-MEP_{5P} with a varying number of stations used in the training set. The red line in the bottom left panel of Fig. 9 represents the median KGE' of the best-performing product used in the computation of RF-MEP_{5P} (see Fig. 4), illustrating the

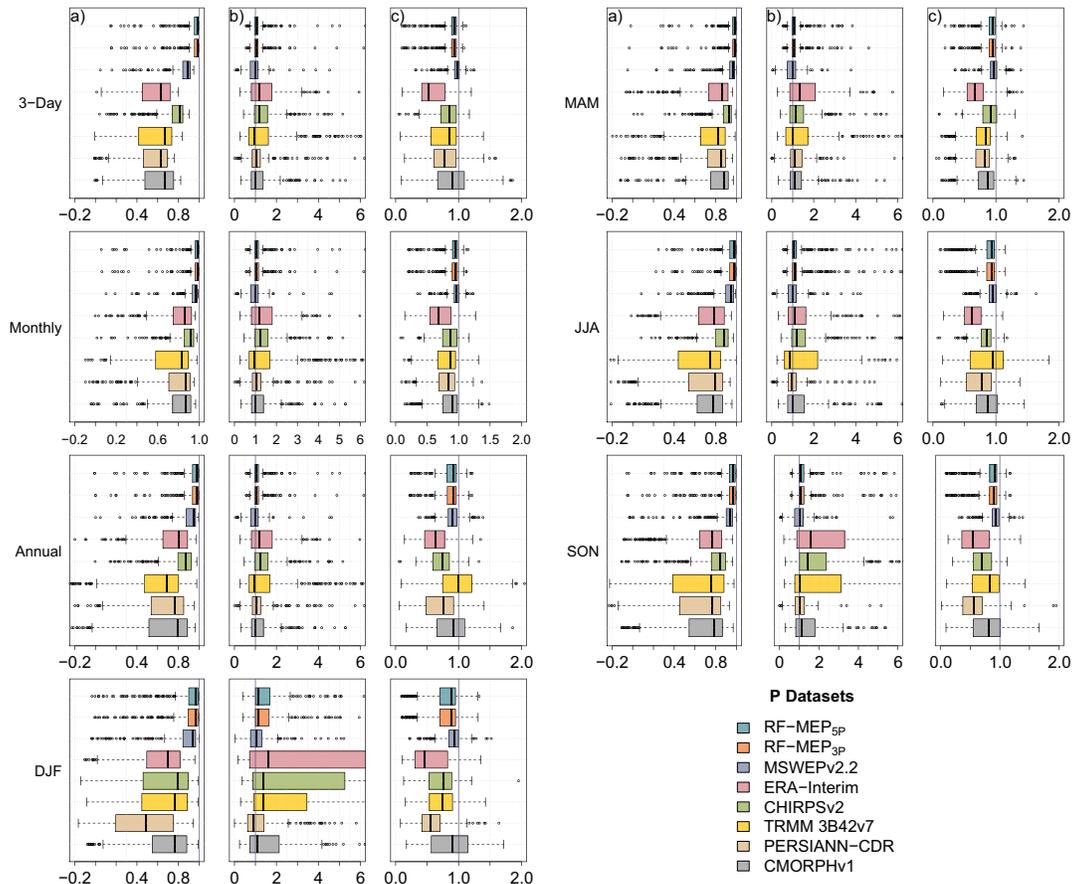


Fig. 5. The r (a), β (b), and γ (c) components of the KGE' for all P products using the ground-based validation dataset. The solid line represents the median value, the edges of the boxes represent the first and third quartiles, and the whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box. The blue line indicates the optimal value for each component.

improvement obtained even when only 10% (37) of stations are used in the training set. Also, Fig. 9 indicates that the inclusion of more stations improves the product performance in comparison to the best product available, which is consistent with other studies (Borga and Vizzaccaro, 1997; Chen et al., 2008; Goudenhoofd and Delobbe, 2009). This suggests that the application of this method in other data-scarce regions is expected to improve the representation of P . The results of the CSI and f_{bias} show that the RF-MEP_{5P} increases the detection of different P intensities in comparison to the single P products (see Fig. 8). Similar to the KGE' , there is a visible improvement in the detection of these events when more stations are used.

Fig. 10 plots the KGE' values of RF-MEP_{5P} at all evaluated timescales for varying spatial resolutions of the covariates. It shows that resampling all the P products into a unified grid has a negligible impact on the performance of the final merged product.

The SRTM-v4 contains offsets in vegetated areas because the SRTM radar signal scatters from the woody structure within the canopy (Gallant et al., 2012). Although we did not remove the impacts of vegetation height to calculate a bare-earth DEM (~40 m over the South and Far South forests of Chile), we do not expect substantial changes because these elevation offsets become negligible at such a spatial resolution (0.05°).

4.5. Comparison between RF-MEP and different merging methods

Fig. 11 shows the performance of RF-MEP_{5P} compared to KED, OOR arithmetic mean, and MSWEPv2.2. The performance of ERA-Interim is also plotted because it is the best-performing P product used in the merging procedure. RF-MEP_{5P} showed the best performance at the 3-day temporal scale, followed by KED and MSWEPv2.2. The OOR arithmetic mean product shows the lowest KGE' , γ , and r ; however, it is able to accurately represent the total P volume at the 3-day scale. This product also shows the lowest performance when evaluated at different P intensities. Shen et al. (2014) concluded that the categorical performance of the OOR arithmetic mean product improved compared to the selected P products; however, they evaluated the categorical performance only for rain and no-rain events. The distribution of daily P is heavily skewed; and therefore, the performance of the product over different intensities can be masked by the no-rain events. As observed in the lower panel of Fig. 11, averaging different P products reduces the performance at all P intensities because all these products have errors in detection (i.e., the products may estimate different P intensities for a particular day). This analysis suggests that P products should not be averaged to attempt to improve daily P patterns.

KED performed similarly to RF-MEP_{5P}; however, RF-MEP_{5P} showed

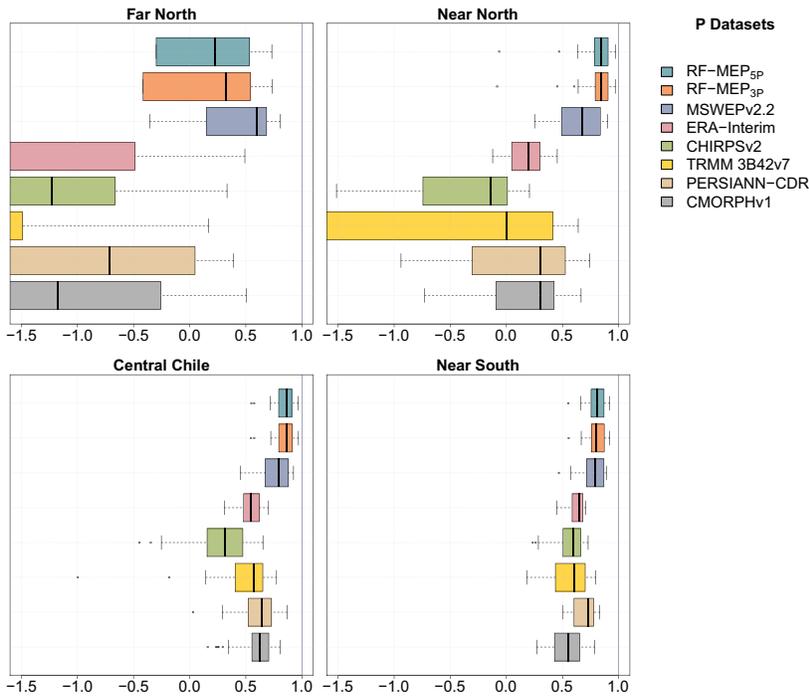


Fig. 6. 3-day KGE' values for the P products at the corresponding grid-cells of the validation set for the four analysed macroclimate zones: Far North, Near North, Central Chile, and South (see Fig. 3). The solid line represents the median value, the edges of the boxes represent the first and third quartiles, and the whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box. The vertical blue line indicates the optimal value for KGE'.

less dispersion in the KGE' and its components, suggesting that RF-MEP is a robust method to merge P products and ground-based data. Ly et al. (2011) obtained poor results when using KED with few sample points, which indicates that the performance of KED is highly influenced by the number of ground stations. Conversely, RF-MEP performed relatively well when the training set was dramatically reduced. The performance of RF-MEP_{SP} is also the highest at monthly, annual and seasonal temporal scales, except in DJF where MSWEPv2.2 performs the best (see

Fig. S1 in the supplementary material).

5. Discussion

5.1. Performance of the merged products

RF-MEP was applied at the daily temporal scale to derive two merged products (RF-MEP_{SP} and RF-MEP_{3P}), which outperformed those

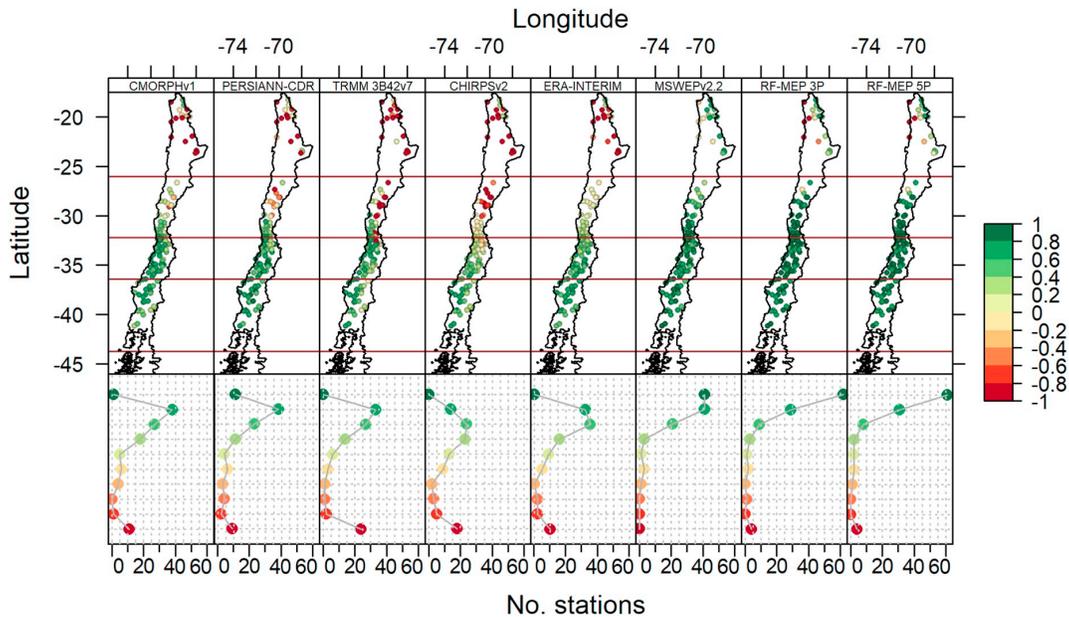


Fig. 7. Spatial distribution of the 3-day KGE' for all P products using ground-based measurements. The dotplots in the bottom of the figure show the number of stations from the validation set (111 stations in total) within each KGE' range.

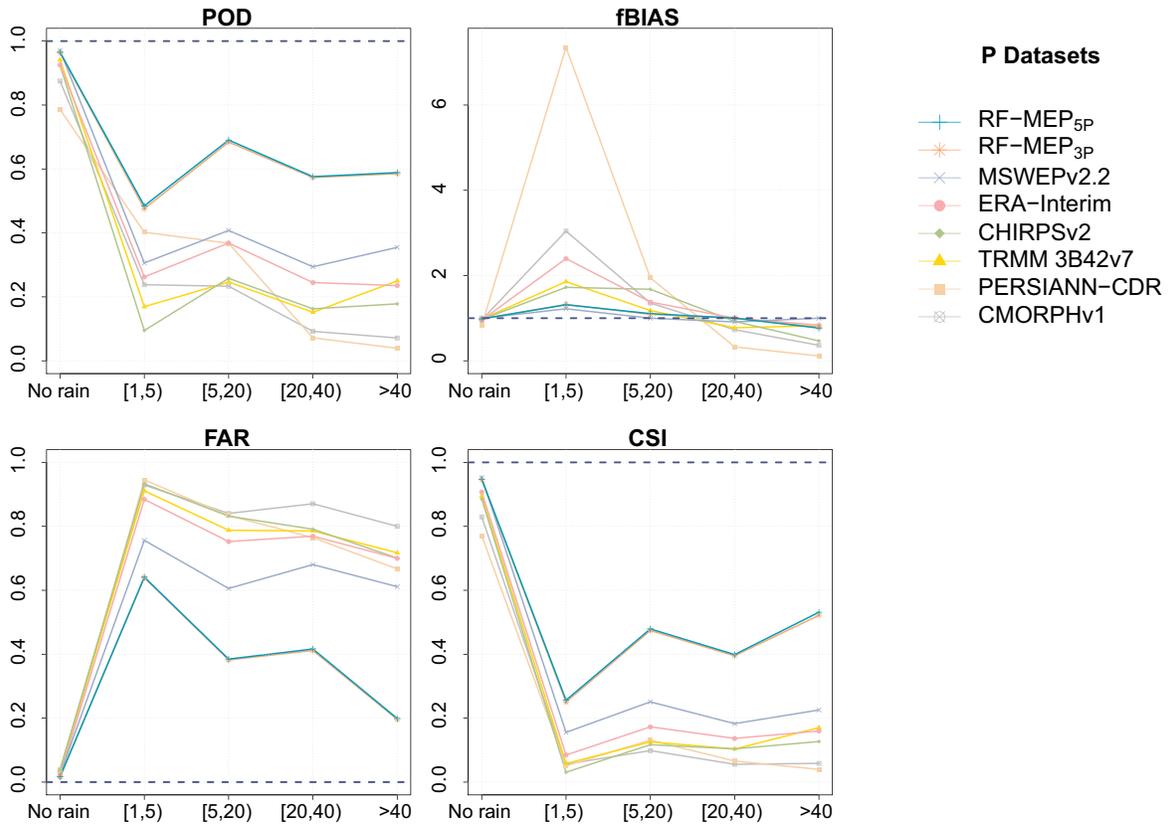


Fig. 8. Median values of the categorical indices of performance at the five P intensity classes (in mm d^{-1}) described in Table 3. From left to right and top to bottom: POD, f_{BIAS} , FAR, and CSI. This analysis is biased towards the merged products due to the difference in the reporting times between the rain gauge stations and the P products. A square bracket indicates the inclusion of the limit value, while a round bracket indicates its exclusion. The blue line represents the optimal value of each index.

used in their computation at all evaluated temporal scales (see Figs. 4, 5, and Table 4). RF-MEP was able to improve the spatio-temporal representation of P (see Figs. 4–8) by combining multiple sources of information. Both merged products showed increased r , β , and γ values at all temporal scales, which indicates that this method is able to represent the total volume and distribution of P by providing a better representation of daily P patterns. Comparable improvements in β were obtained by Manz et al. (2016) and Yang et al. (2017), although Ma et al. (2018) reported a higher bias in their merged product. Also, the reduction in the dispersion of the KGE' and its components

demonstrates that the merged products show good performance over most of the study area. The KGE' has proven to be a useful performance index because of its ability to decompose the performance into r , β , and γ , which can be used to understand the different sources of mismatches.

The evaluated P products showed higher performances at the monthly, seasonal and annual scales in comparison to shorter temporal scales (Fig. 4), similar to the results reported by Jiang et al. (2012) and Zambrano-Bigiarini et al. (2017). This indicates that despite systematic, random, and detection errors present in P products at the daily scale, they are still able to represent P patterns when aggregated at longer

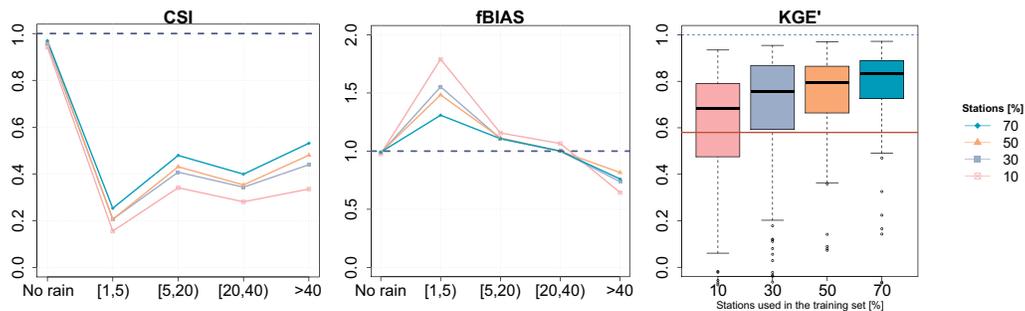


Fig. 9. Performance of the RF-MEP_{5P} using varying percentages of rain gauge stations in the training set. From left to right and top to bottom: the CSI; f_{BIAS} ; and the KGE' evaluation of the derived products (the red line indicates the median KGE' of the best performing product used in the computation of RF-MEP_{5P}). The blue line represents the optimal value for each index. A square bracket indicates the inclusion of the limit value, while a round bracket indicates its exclusion.

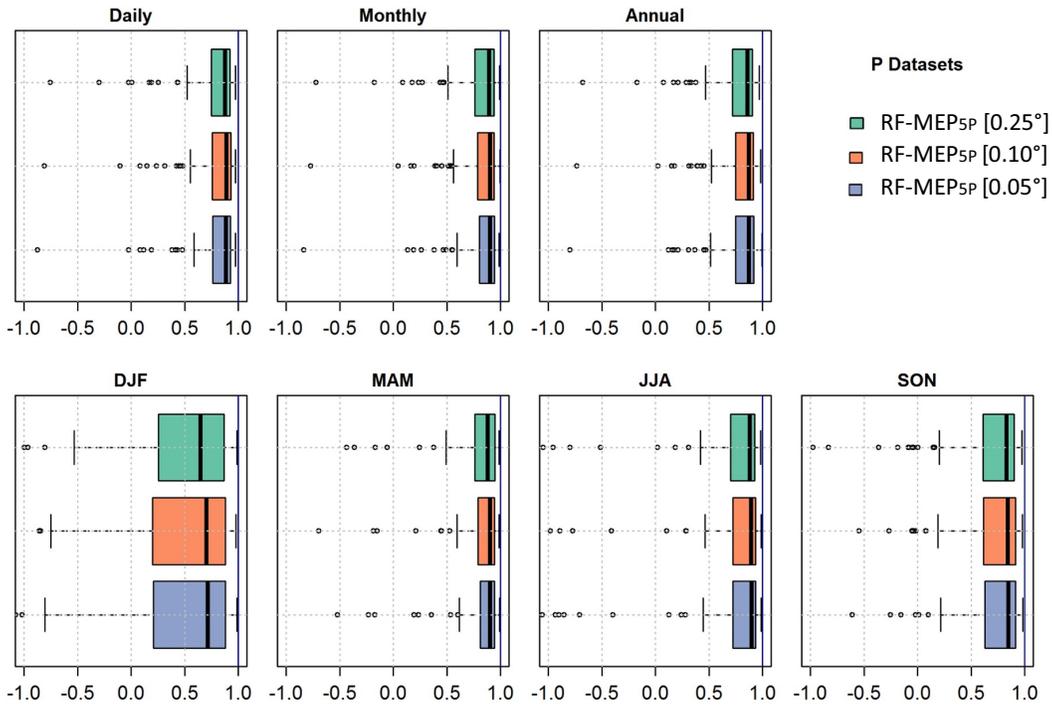


Fig. 10. KGE' values of the RF-MEP_{5P} computed at three spatial resolutions (0.25°, 0.10°, and 0.05°). The solid line represents the median value, the edges of the boxes represent the first and third quartiles, and the whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box. The blue line shows the optimal value for the KGE'.

temporal scales. On the other hand, Maggioni and Massari (2018) concluded that spatial sampling uncertainties tend to decrease for higher temporal resolutions, which means that the point-to-pixel evaluation tends to be more reliable for increasing accumulation periods.

All products showed the lowest performance in summer (DJF), which is consistent with the results obtained by Rabiei and Haberlandt (2015) and Zambrano-Bigiarini et al. (2017). This could be because: i) small-scale convective precipitation events dominate in summer in the

Far North region (Prein and Gobiet, 2017); ii) in warm months, the evaporation of hydrometeors before they reach the ground leads to overestimation and false alarms (Maggioni and Massari, 2018); and iii) passive microwave radiometers overestimate and underestimate *P* during summer and winter, respectively (Tang et al., 2014).

Both merged products presented their lowest performance over the arid Far North region as a consequence of the low performance of all *P* products used as covariates (see Fig. 7). This is in agreement with Manz

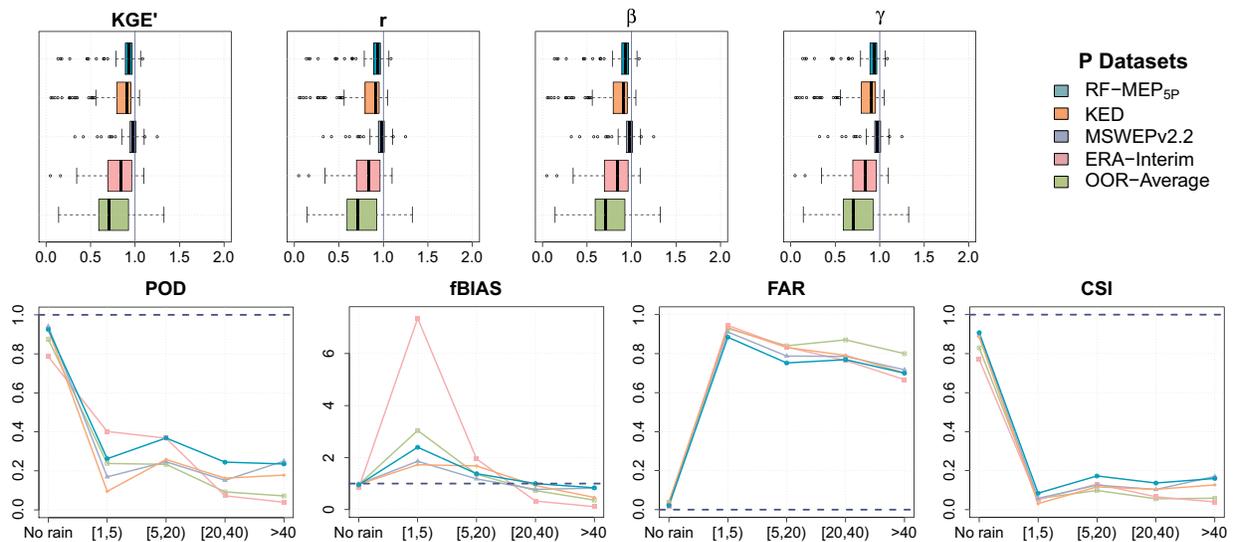


Fig. 11. Performance of different merging procedures using the KGE' and its components as continuous indices and POD, fbias, FAR, and CSI as categorical indices at the 3-day temporal scale.

Table 4
Median values of the continuous indices used in the evaluation of P products.

P product	KGE'	r	β	γ
CMORPHv1	0.43	0.67	1.03	0.82
PERSIANN-CDR	0.23	0.62	1.34	0.50
TRMM 3B42v7	0.47	0.69	1.05	0.88
CHIRPSv2	0.48	0.62	1.04	0.71
ERA-Interim	0.58	0.82	1.31	0.84
MSWEPv2.2	0.74	0.89	1.00	0.97
RF-MEP _{3P}	0.83	0.94	1.03	0.93
RF-MEP _{5P}	0.83	0.94	1.04	0.94

et al. (2016), where the merged products presented high uncertainty and low performances predominantly over regions with low and intermittent P regimes. The mismatches of the P products are more evident in arid and semi-arid climates because over low P regimes, any overestimation or underestimation will have a greater impact on the performance evaluation. Despite this, the RF-MEP_{5P} and RF-MEP_{3P} products were able to adequately represent the P patterns of the higher elevations of the Far North, showing that RF-MEP is able to improve the spatio-temporal estimation of P through the inclusion of complementary information, even in regions where the selected products exhibit low performance.

Because both merged products were computed using daily gauge data from the national water agencies they represent daily accumulations from 11:00–10:59 UTC, whereas all other selected P products represent daily P accumulations from 0:00 to 23:59 UTC (~11 h difference; for discussion, see Beck et al., 2019). This time difference must be considered for the evaluation of the P products at the daily temporal scale. Among the evaluated P products, only MSWEPv2.2 incorporates daily gauge data and applies corrections to account for the reporting times of the rain gauges. Fig. 12 shows the evaluation of the P products for 1-day and 3-day periods. Both merged products performed similarly well with a median KGE' of 0.83 because they use the Chilean rain gauges; however, the five P products used in their computation performed slightly worse in the 1-day evaluation due to the 11 h difference in the reporting times. The 3-day temporal scale was considered sufficient to render the difference in reporting times negligible.

5.2. Correction of mismatches of the original P products

Our results showed that the blending of multiple P estimates, topography-related information, and ground-based measurements, can improve the spatio-temporal characterisation of P , which is consistent with the results obtained by Verdin et al. (2016) and Manz et al. (2016). The r , β , and γ components improved at all temporal scales. The γ of both merged products showed a systematic underestimation ($\gamma \sim 0.9$, see Fig. 5) at all temporal scales as a consequence of averaging the predictions of the different trees from the RF model. Despite this, Fig. 5c demonstrates that the γ values of the merged products are higher than those shown by the products used as covariates.

Recently, Alvarez-Garreton et al. (2018) derived runoff coefficients larger than 1, mainly over Central Chile and in the Far-South, with increasing coefficient values towards the Andes. This finding is consistent with those of Beck et al. (2017a), indicating that more water is leaving the catchments than the total amount entering as P . This suggests that the P products systematically underestimate P at high elevations throughout Chile, which may be due to the inability of satellite-based products to accurately estimate P over snow and ice-covered surfaces (Beck et al., 2017a). Also, during winter, most Chilean rain gauges located at high elevations are not able to correctly incorporate snow into the P measurement, leading to an underestimation of P . Therefore, even considering the good performance of the two merged products at different temporal scales, it is likely that the real amount of P is underestimated at high elevations due to the absence of ground-based information. To reduce the possible underestimation of P over high elevation and snow-driven catchments, the incorporation of rain gauges able to measure both liquid and solid precipitation at high elevations is recommended, along with the use of P products that account for solid P (such as MSWEPv2.2 and reanalysis products).

The inclusion of different P products improved the detection of different P intensities at the daily scale, as observed in the improved categorical performance of the merged products compared to that of the covariates (see Fig. 8 and Table 5). The categorical performance of both merged products showed an improved detection of the selected P intensities and a reduction in the amount of days that are incorrectly classified. These results, in combination with the improved values of r and β , show that RF-MEP is capable of correcting P events at the daily

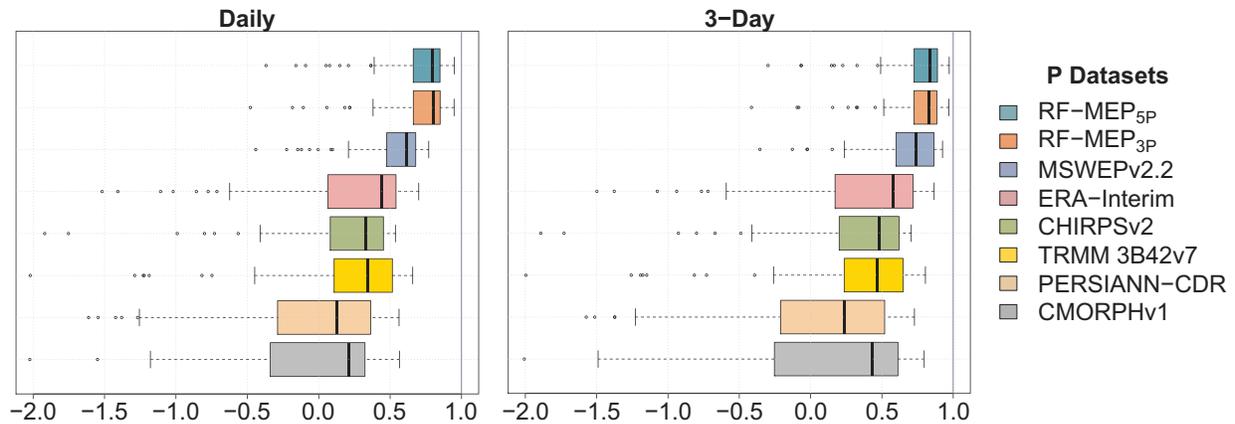


Fig. 12. KGE' values calculated using the ground-based validation dataset at the 1-day time scale (left) and the 3-day time scale (right). The solid line represents the median value, the edges of the boxes represent the first and third quartiles, and the whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box. The vertical blue line indicates the optimal value for KGE'.

Table 5

Median values of POD, FAR, f_{bias} , and CSI for the different P intensities (see Table 3) for ERA-Interim, MSWEPv2.2, and RF-MEP_{5P}.

Intensity (mm)	ERA-Interim				MSWEPv2.2				RF-MEP _{5P}			
	POD	FAR	f_{bias}	CSI	POD	FAR	f_{bias}	CSI	POD	FAR	f_{bias}	CSI
[0, 1)	0.92	0.02	0.95	0.91	0.97	0.02	1.00	0.95	0.96	0.02	0.98	0.95
[1, 5)	0.26	0.88	2.40	0.08	0.30	0.76	1.22	0.15	0.48	0.64	1.31	0.26
[5, 20)	0.37	0.75	1.38	0.17	0.40	0.60	1.00	0.25	0.69	0.38	1.10	0.48
[20, 40)	0.24	0.77	1.00	0.14	0.29	0.68	0.91	0.18	0.58	0.42	1.00	0.40
≥ 40	0.23	0.70	0.83	0.16	0.35	0.61	1.00	0.22	0.59	0.20	0.77	0.53

scale, assigning more accurate P amounts to each day, and preserving the total volume of P at larger scales; consequently improving the spatial representation of P patterns.

The analysis of the P products at different intensities is affected by the difference in reporting times between the products and the ground-based measurements (see Fig. 12). All the products used as covariates, with the exception of CHIRPSv2 and TRMM 3B42v7, presented statistically significant differences at the 95% confidence interval between the daily and 3-day values. This issue is unfortunately ignored in the majority of P evaluation studies and constitutes a major limitation of most evaluations carried out in time zones far from 0:00 UTC.

Fig. 13 shows the relative difference of mean annual P (2000–2016) between each product and the values observed at the rain gauges of the validation set. These values are in agreement with the spatial performance assessment (Fig. 7), where the P products presented the lowest performance in the Far North. The blue colours indicate overestimation of the products, while the red colours indicate underestimation. P is overestimated in the Far North by CMORPHv1, PERSIANN-CDR, TRMM 3B42v7, CHIRPSv2, and ERA-Interim; and as a consequence, both merged products overestimate P over this region (except for the high elevated areas). These results are in agreement with Dinku et al. (2011) and Zambrano-Bigiarini et al. (2017), where the products

overestimated P over the arid regions of Africa and Chile, respectively. MSWEPv2.2 and the merged products were able to capture the P volume over the mountainous area in the Far North, despite the challenges presented by climate variability caused by extreme topography and by a lack of ground-based measurements (Maggioni and Massari, 2018).

The merged products show lower relative difference, i.e. good performance, for almost all stations in the Near South, Central Chile, South, and elevated areas in the Far North. The improved performance of the merged products can be observed in the lower panel of Fig. 13, which highlights that the majority of the P products presented relative differences between -0.2 and 0.2 compared to rain gauges. This suggests that RF-MEP is capable of representing the mean annual P patterns when applied at daily temporal scale. The overestimation over the Far North is expected because all products used to derive both merged products tend to overestimate P over this region.

5.3. Impact of network density, spatial resolution, and limitations

A high number of rain gauge stations in the training set leads to higher performance and higher detection of P intensities, as observed in Fig. 9. When we reduced the training sample to 10% (37) of the total

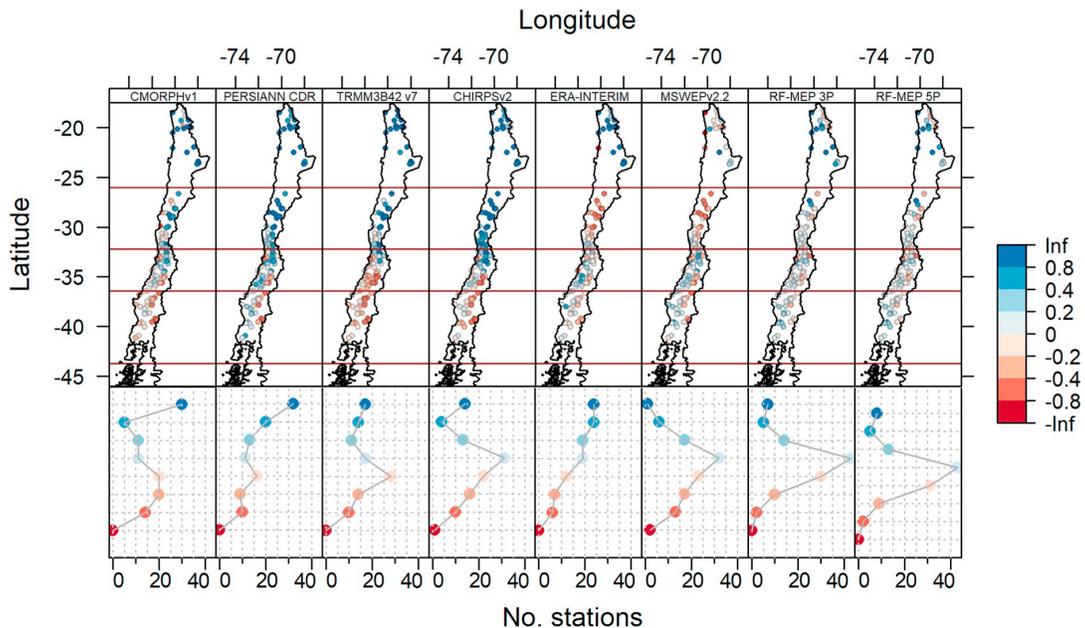


Fig. 13. Mean annual relative difference for the P products for 2000–2016. The points with negative values (red colours) are underestimated by the respective product, while the points with positive values (blue colours) are overestimated.

available stations, RF-MEP_{5P} was still able to outperform the products used as covariates, showing the effectiveness of the proposed RF-MEP method.

The products RF-MEP_{5P} and RF-MEP_{3P} performed similarly, as observed in Figs. 4–8. The median values and the interquartile ranges of the KGE', r , β , and γ are similar for both merged products, except over the Far North, where RF-MEP_{5P} shows less dispersion in the KGE' and its components than RF-MEP_{3P}, despite the slight decrease in the median performance. This indicates that the inclusion of more P products could reduce the dispersion in areas where the selected products show low performance. The similar performance of RF-MEP_{5P} and RF-MEP_{3P} indicates that the method is able to extract useful information from the P products. Similar results were obtained when RF-MEP_{3P} used ERA-Interim, CMORPHv1, and TRMM 3B42v7 instead of ERA-Interim, CHIRPSv2, and PERSIANN-CDR (please see Figs. S2 and S3 from the supplementary material), demonstrating that RF-MEP is a robust merging method. Although the P products must be resampled to the same spatial resolution to generate the merged product, the effect of including P products generated at different spatial resolutions is negligible (see Fig. 10).

RF-MEP_{5P} includes CMORPHv1 and TRMM 3B42v7, which reduces the potential temporal coverage by 15 years (RF-MEP_{3P} can be generated from 1983 onwards, while RF-MEP_{5P} can only be generated from 1998). Therefore, based on the similar strong performances of both merged products (see Section 5.1), we prefer RF-MEP_{3P} for the Chilean case study, as the benefits of including CMORPHv1 and TRMM 3B42v7 to generate RF-MEP_{5P} are outweighed by the loss of 15 years of record.

Although RF-MEP was only applied over Chile, we are confident that this method could be successfully applied over other areas, due to its outstanding performance in a region with notable heterogeneity in topography and climate, and because it was able to improve the spatio-temporal characterisation of P even when the training set was largely reduced. However, some limitations apply to this method: *i*) since ground-based data are necessary, it would be difficult to apply the proposed method globally and in near-real time; *ii*) it can be computationally intensive when applied to large areas; and *iii*) it has problems predicting values that are completely out from the training range.

6. Conclusion

Satellite and reanalysis-based P estimates provide an unprecedented opportunity for numerous hydrological, meteorological, and other environmental applications. Despite the continuous improvements of P products, different types of mismatches still exist in most of them. Here we present RF-MEP, a novel method capable of deriving improved P estimates by merging information from (near-)global and publicly available P products, rain gauge stations, and topography-related data. Two merged products (RF-MEP_{3P} and RF-MEP_{5P}) obtained with the proposed method showed improved r , β , and γ values at all temporal scales compared to all the individual P products used as covariates. Furthermore, both merged datasets exhibited improved POD, FAR, f_{bias} , and CSI for different P intensities. Finally, both merged products performed better than the benchmark dataset MSWEPv2.2, except during summer (DJF). The key findings of the application of this method to the Chilean case study are as follows:

- RF-MEP can be applied at different temporal scales (e.g., daily, monthly, or annually) to obtain an improved spatio-temporal representation of P patterns.
- The different P products used in this study performed better at longer timescales than at short timescales, while both merged products performed well at all timescales.
- RF-MEP_{3P} and RF-MEP_{5P} outperformed all the evaluated P products

at the 3-day, monthly, annual, MAM, JJA, and SON temporal scales. However, the benchmark MSWEPv2.2 outperformed the merged products during summer (DJF).

- RF-MEP_{3P} (which uses CHIRPSv2, PERSIANN-CDR, and ERA-Interim) showed a similar performance to RF-MEP_{5P} (which also included CMORPHv1 and TRMM 3B42v7). Therefore, including CMORPHv1 and TRMM 3B42v7 as covariates in the merging procedure only led to a minor increase in the overall performance of the final merged product. Consequently, for the Chilean case study, it is preferable to use RF-MEP_{3P} and gain 15 years of data (1983 as the starting date instead of 1998).
- The performance of RF-MEP increases when more rain gauge stations are used to train the model; however, it is still able to improve P characteristics even with relatively few stations in the training set.
- RF-MEP showed better performance than the results obtained using Kriging with external drift and one-outlier-removed arithmetic mean.
- The difference in reporting times between the P products and the ground-based measurements must be taken into account when assessing the performance of P products at the daily temporal scale so that their performance is not underestimated. This issue constitutes a major limitation of most P evaluation studies carried out far from 0:00 UTC.
- The KGE' proved to be a versatile performance index because of its ability to decompose the performance of the P products into r , β , and γ . Therefore, the KGE' helps us understand the sources of mismatches between the P products and ground-based observations. In addition, the use of categorical indices provides crucial information about the performance of these P datasets for capturing different P intensities.

RF-MEP was developed to improve the characterisation of the spatio-temporal variability of P by merging multiple P products, topography-related datasets, and ground-based information. The P products used in this study are publicly available and have a (quasi-)global spatial coverage. This method was validated over Chile, a country which exhibits notable heterogeneity in topography, climate, and land cover. For this reason, we are confident that RF-MEP can be successfully applied in different regions and catchments worldwide, and could also be used to improve other climatological variables when ground-based data are available.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank the Centers for Natural Resources and Development (CNRD) Ph.D. program for their financial support to the main author. The daily RF-MEP_{3P} will be made available upon request for 2000–2016 at a spatial resolution of 0.05°. In addition, Dr. Zambrano-Bigiarini thanks Conicyt-Fondecyt 11150861 for the financial support from 2016 to 2018 and the Center for Climate and Resilience Research (CR2, Conicyt-FONDAP 15110009) for providing the rain gauge data. We particularly thank Tim R. McVicar and the three anonymous reviewers, whose constructive comments helped to improve the quality of the final manuscript. We would also like to thank the RSE editorial team for their support.

Appendix A. Literature review table

Table A1

Main steps in the methodology of different studies that have applied merging algorithms to improve the spatio-temporal characterisation of P at different temporal scales.

Study	Merging method(s)	Spatio-temporal resolution of the merged product(s)	Description of the approach
Li and Shao (2010)	Nonparametric kernel smoothing	Daily (0.25°)	1. Calculation of residual values; 2. Background error estimation using a kernel smoothing method (double smoothing); 3. Removal of the estimated error from the background field
Rozante et al. (2010)	Barnes objective analysis method	Daily (0.25°)	1. Only the rain gauge observations are considered over the 5 by 5 square of cells centred around every grid-cell with a rain gauge station; 2. Interpolation using the Barnes objective analysis method for the remaining grid-cells
Xie and Xiong (2011)	Optimal Interpolation	Daily (0.25°)	1. Bias correction through a probability density function matching of satellite and rain gauge data; 2. Optimal interpolation
Gebregiorgis and Hossain (2011)	Linear weights based on hydrologic model predictability	Daily (0.125°)	1. Calculation of the mean squared error (MSE) of soil moisture and runoff using each P product to force a distributed hydrological model; 2. Inversion of MSEs to be used as weights; 3. Merging of the P products using linear weighting
Woldemeskel et al. (2013)	Linearised weighting procedure	Monthly (0.05)	1. P interpolation using thin plate smoothing splines (TPSS) with standardised rain gauge data followed by a back-transformation; 2. Merging using a linearised weighting procedure
Shen et al. (2014)	Arithmetic mean and inverse-error-square weighting methods	Daily (0.25°)	Three methods: M1. Arithmetic mean; M2. Inverse-error-square weighting; M3. One-outlier removed arithmetic mean (i.e., one product removed)
Nie et al. (2015)	Optimal interpolation	Daily (0.25°)	1. Bias correction through a cumulative distribution function matching procedure; 2. Quantification of background and observation errors; 3. Application of the optimal interpolation technique
Fu et al. (2016)	Bayesian model averaging	Annual mean (0.1°)	1. Non-linear spatial interpolation of P products; 2. Merging using the Bayesian model averaging technique
Manz et al. (2016)	Linear modelling, residual IDW, and Kriging-based methods	Monthly mean (5 km – 0.05°)	Five methods: M1. Linear Modeling; M2. Residual IDW; M3. Ordinary Kriging (only gauge-based); M4. Residual ordinary Kriging; M5. Kriging with external drift
Verdin et al. (2016)	Ordinary Kriging and k-nearest neighbour local polynomials	Monthly (0.05°)	Two methods: M1. Ordinary Kriging; M2. A local regression is fitted considering data from within a small neighbourhood, and the weighted least squares are used to fit the local polynomials
Shi et al. (2017)	Merging weights based on the effective influence radius of rain gauges	Hourly (1 km)	1. Selection of the P product; 2. Downscaling of the P product using a DEM; 3. Determination of weighted differences between the downscaled product and rain gauge data; 4. Merging the downscaled product and the weighted differences considering the number of gauges in the effective influence radius
Yang et al. (2017)	Inverse-root-mean-square-error weighting	Daily (0.04°)	1. Bias correction of the P product using a quantile mapping technique and a Gaussian weighting interpolation scheme; 2. Interpolation of rain gauge data using a Gaussian weighting function; 3. Data merging using inverse-mean-square-error weighting
Ma et al. (2018)	Bayesian Model Averaging	Daily (0.25°)	1. A BMA scheme is used to adjust the PDF of the satellite estimates with the expectation-maximisation method used for each member for each day at the gauge locations; 2. Interpolation using OK
Beck et al. (2019)	Weighted averaging with CDF matching	3-hourly (0.10°)	1. Gauge data quality control; 2. Inferring gauge reporting times; 3. Rainfall estimation using thermal infrared imagery; 4. Gauge-based assessment of satellite and reanalysis P datasets; 5. Global maps of weights and wet-day biases; 6. Determination of long-term mean P ; 7. P frequency correction and dataset harmonisation; 8. Reference P distributions; 9. Merging of satellite and reanalysis P datasets; 10. Gauge correction scheme

Appendix B. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.rse.2019.111606>. These data include the Google maps of the most important areas described in this article.

References

- Adhikary, S.K., Yilmaz, A.G., Muttil, N., 2015. Optimal design of rain gauge network in the Middle Yarra River catchment, Australia. *Hydrol. Process.* 29, 2582–2599.
- Alvarez-Garretón, C., Mendoza, P.A., Boisier, J.P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Cortes, G., Garreaud, R., McPhee, J., et al., 2018. The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies-Chile dataset. *Hydrol. Earth Syst. Sci.* 22, 5817–5846.
- Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D.K., Knapp, K.R., Cecil, L.D., Nelson, B.R., Prat, O.P., 2015. PERSIANN-CDR: daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bull. Am. Meteorol. Soc.* 96, 69–83.
- Baez-Villanueva, O.M., Zambrano-Bigiarini, M., Ribbe, L., Nauditt, A., Giraldo-Osorio, J.D., Thinh, N.X., 2018. Temporal and spatial evaluation of satellite rainfall estimates over different regions in Latin-America. *Atmos. Res.* 213, 34–50.
- Beck, H.E., van Dijk, A.I., De Roo, A., Miralles, D.G., McVicar, T.R., Schellekens, J., Bruijnzeel, L.A., 2016. Global-scale regionalization of hydrologic model parameters. *Water Resour. Res.* 52, 3599–3622.
- Beck, H.E., van Dijk, A.I.J.M., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B., de Roo, A., 2017a. MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrol. Earth Syst. Sci.* 21, 589–615.
- Beck, H.E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A.I.J.M., Weedon, G.P., Brocca, L., Pappenberger, F., Huffman, G.J., Wood, E.F., 2017b. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrol. Earth Syst. Sci.* 21, 6201–6217.
- Beck, H.E., Wood, E.F., Pan, M., Fisher, C.K., Miralles, D.G., van Dijk, A.I., McVicar, T.R., Adler, R.F., 2019. MSWEP V2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment. *Bull. Am. Meteorol. Soc.* <https://doi.org/10.1175/BAMS-D-17-0138.1>.
- Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Sci. Data* 5, 180–214.
- Behrens, T., Schmidt, K., Viscarra Rossel, R.A., Gries, P., Scholten, T., MacMillan, R.A., 2018. Spatial modelling with Euclidean distance fields and machine learning. *Eur. J. Soil Sci.*
- Bergeron, T., 1960. Preliminary results of project pluvius. *Publ. Comm. Land Erosion* 53, 226–237.

- Biau, G., Scornet, E., 2016. A random forest guided tour. *TEST* 25, 197.
- Borga, M., Vizzaccaro, A., 1997. On the interpolation of hydrologic variables: formal equivalence of multiquadratic surface fitting and kriging. *J. Hydrol.* 195, 160–171.
- Breiman, L., 2001. *Random Forests*. *Mach. Learn.* 45, 5–32.
- Chen, M., Shi, W., Xie, P., Silva, V.B., Kousky, V.E., Wayne Higgins, R., Janowiak, J.E., 2008. Assessing objective techniques for gauge-based analyses of global daily precipitation. *J. Geophys. Res.-Atmos.* 113.
- Chen, Y., Xia, J., Liang, S., Feng, J., Fisher, J.B., Li, X., Li, X., Liu, S., Ma, Z., Miyata, A., et al., 2014. Comparison of satellite-based evapotranspiration models over terrestrial ecosystems in china. *Remote Sens. Environ.* 140, 279–293.
- Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.K., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* 137, 553–597.
- Derin, Y., Yilmaz, K.K., 2014. Evaluation of multiple satellite-based precipitation products over complex topography. *Journal of Hydrometeorology* 15, 1498–1516. <https://doi.org/10.1175/JHM-D-13-0191.1>.
- Díaz-Uriarte, R., Alvarez de Andrés, S., 2006. Gene selection and classification of microarray data using random forest. *BMC Bioinf.* 7, 3.
- Dinku, T., Ceccato, P., Connor, S.J., 2011. Challenges of satellite rainfall estimation over mountainous and arid parts of East Africa. *Int. J. Remote Sens.* 32, 5965–5979.
- Dinku, T., Ceccato, P., Grover-Kopec, E., Lemma, M., Connor, S., Ropelewski, C., 2007. Validation of satellite rainfall products over East Africa's complex topography. *Int. J. Remote Sens.* 28, 1503–1526.
- Fu, Y., Xia, J., Yuan, W., Xu, B., Wu, X., Chen, Y., Zhang, H., 2016. Assessment of multiple precipitation products over major river basins of China. *Theor. Appl. Climatol.* 123, 11–22.
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci. Data* 2, 150066.
- Gallant, J., Read, A., Dowling, T., 2012. Removal of tree offsets from SRTM and other digital surface models. *Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* 39, 275–280.
- Gao, Y.C., Liu, M.F., 2013. Evaluation of high-resolution satellite precipitation products using rain gauge observations over the Tibetan Plateau. *Hydrol. Earth Syst. Sci.* 17, 837–849.
- García, M., Peters-Lidard, C.D., Goodrich, D.C., 2008. Spatial interpolation of precipitation in a dense gauge network for monsoon storm events in the southwestern united states. *Water Resour. Res.* 44.
- Gebregiorgis, A., Hossain, F., 2011. How much can a priori hydrologic model predictability help in optimal merging of satellite precipitation products? *J. Hydrometeorol.* 12, 1287–1298.
- Goudenhoofd, E., Delobbe, L., 2009. Evaluation of radar-gauge merging methods for quantitative precipitation estimates. *Hydrol. Earth Syst. Sci.* 13, 195–203.
- Grimes, D.I.F., Pardo-Igúzquiza, E., Bonifacio, R., 1999. Optimal areal rainfall estimation using rain gauges and satellite data. *J. Hydrol.* 222, 93–108.
- Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009, Oct. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *J. Hydrol.* 377, 80–91.
- Hengl, T., 2019. *GSIF: Global Soil Information Facilities*. R package version 0.5-5. .
- Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B., Gräler, B., 2018. Random Forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ* 6, e5518.
- Hijmans, R.J., 2018. *raster: Geographic Data Analysis and Modeling*. R package version 2.8-4. .
- Hirpa, F.A., Gebremichael, M., Hopson, T., 2010. Evaluation of high-resolution satellite precipitation products over very complex terrain in Ethiopia. *J. Appl. Meteorol. Climatol.* 49, 1044–1051.
- Huffman, G., Adler, R., Bolvin, D., Nelkin, E., 2010. *The TRMM Multi-satellite Precipitation Analysis (TMPA)*. Chapter 1 in *Satellite Rainfall Applications for Surface Hydrology*, f. hossain and m. gebremichael, eds.
- Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Wolff, D.B., Adler, R.F., Gu, G., Hong, Y., Bowman, K.P., Stocker, E.F., 2007. The TRMM Multisatellite Precipitation Analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.* 8, 38.
- Hutchinson, M.F., 1995. Interpolating mean rainfall using thin plate smoothing splines. *Int. J. Geogr. Inf. Syst.* 9, 385–403.
- Jaagus, J., Briede, A., Rimkus, E., Remm, K., 2010. Precipitation pattern in the baltic countries under the influence of large-scale atmospheric circulation and local landscape factors. *Int. J. Climatol.* 30, 705–720.
- Jarvis, A., Reuter, H., Nelson, A., Guevara, E., 2008. Hole-filled SRTM for the globe version 3, from the CGIAR-CSI SRTM 90m database.
- Jiang, S., Ren, L., Hong, Y., Yong, B., Yang, X., Yuan, F., Ma, M., 2012. Comprehensive evaluation of multi-satellite precipitation products with a dense rain gauge network and optimally merging their simulated hydrological flows using the Bayesian model averaging method. *J. Hydrol.* 452, 213–225.
- Joyce, R.J., Janowiak, J.E., Arkin, P.A., Xie, P., 2004. CMORPH: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrol.* 5, 487–503.
- Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.* 424, 264–277.
- Li, M., Shao, Q., 2010. An improved statistical approach to merge satellite rainfall estimates and rain gauge data. *J. Hydrol.* 385, 51–64.
- Li, Z., Yang, D., Hong, Y., 2013. Multi-scale evaluation of high-resolution multi-sensor blended global precipitation products over the yangtze river. *J. Hydrol.* 500, 157–169.
- Liaw, A., Wiener, M., 2002. Classification and regression by random forest. *R News* 2, 18–22.
- Lievens, H., Tomer, S.K., Al Bitar, A., De Lannoy, G.J., Drusch, M., Dumedah, G., Franssen, H.-J.H., Kerr, Y.H., Martens, B., Pan, M., et al., 2015. SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia. *Remote Sens. Environ.* 168, 146–162.
- Ly, S., Charles, C., Degre, A., 2011. Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium. *Hydrol. Earth Syst. Sci.* 15, 2259–2274.
- Ma, Y., Hong, Y., Chen, Y., Yang, Y., Tang, G., Yao, Y., Long, D., Li, C., Han, Z., Liu, R., 2018. Performance of optimally merged multisatellite precipitation products using the dynamic Bayesian model averaging scheme over the Tibetan Plateau. *J. Geophys. Res.-Atmos.* 123, 814–834.
- Maggioni, V., Massari, C., 2018. On the performance of satellite precipitation products in riverine flood modeling: a review. *J. Hydrol.* 558, 214–224.
- Manz, B., Buytaert, W., Zulkafli, Z., Lavado, W., Willems, B., Robles, L.A., Rodríguez-Sánchez, J.-P., 2016. High-resolution satellite-gauge merged precipitation climatologies of the Tropical Andes. *J. Geophys. Res.-Atmos.* 121, 1190–1207.
- McKenney, D.W., Pedlar, J.H., Papadopol, P., Hutchinson, M.F., 2006. The development of 1901–2000 historical monthly climate models for Canada and the United States. *Agric. For. Meteorol.* 138, 69–81.
- McVicar, T.R., Körner, C., 2013. On the use of elevation, altitude, and height in the ecological and climatological literature. *Oecologia* 171, 335–337.
- McVicar, T.R., Van Niel, T.G., Li, L., Hutchinson, M.F., Mu, X., Liu, Z., 2007. Spatially distributing monthly reference evapotranspiration and pan evaporation considering topographic influences. *J. Hydrol.* 338, 196–220.
- Montecinos, A., Aceituno, P., 2003. Seasonality of the ENSO-related rainfall variability in Central Chile and associated circulation anomalies. *J. Clim.* 16, 281–296.
- Nie, S., Luo, Y., Wu, T., Shi, X., Wang, Z., 2015. A merging scheme for constructing daily precipitation analyses based on objective bias-correction and error estimation techniques. *J. Geophys. Res.-Atmos.* 120, 8671–8692.
- Oliver, M., Webster, R., 2014. A tutorial guide to geostatistics: computing and modelling variograms and kriging. *Catena* 113, 56–69.
- Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and Random Forests for ecological prediction. *Ecosystems* 9, 181.
- Prein, A.F., Gobiet, A., 2017. Impacts of uncertainties in european gridded precipitation observations on regional climate analysis. *Int. J. Climatol.* 37, 305–327.
- R Core Team, 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rabiei, E., Haberlandt, U., 2015. Applying bias correction for merging rain gauge and radar data. *J. Hydrol.* 522, 544–557.
- Robertson, A.W., Baethgen, W., Block, P., Lall, U., Sankarasubramanian, A., de Souza Filho, F.d.A., Verbist, K.M., 2014. Climate risk management for water in semi-arid regions. *Earth Perspect.* 1, 12.
- Roy, M.-H., Larocque, D., 2012. Robustness of random forests for regression. *Journal of Nonparametric Statistics* 24, 993–1006.
- Rozante, J.R., Moreira, D.S., de Goncalves, L.G.G., Vila, D.A., 2010. Combining TRMM and surface observations of precipitation: technique and validation over south america. *Weather Forecast.* 25, 885–894.
- Schneider, U., Fuchs, T., Meyer-Christoffer, A., Rudolf, B., 2008. *Global precipitation analysis products of the GPCP*. 112 Global Precipitation Climatology Centre (GPCC), DWD: Internet Publication.
- Seo, D.-J., Krajewski, W.F., Bowles, D.S., 1990. Stochastic interpolation of rainfall data from rain gauges and radar using cokriging: 1. Design of experiments. *Water Resour. Res.* 26, 469–477.
- Shen, Y., Xiong, A., Hong, Y., Yu, J., Pan, Y., Chen, Z., Saharia, M., 2014. Uncertainty analysis of five satellite-based precipitation products and evaluation of three optimally merged multi-algorithm products over the Tibetan Plateau. *Int. J. Remote Sens.* 35, 6843–6858.
- Shi, H., Chen, J., Li, T., Wang, G., 2017. A new method for estimation of spatially distributed rainfall through merging satellite observations, rain gauge records, and terrain digital elevation model data. *J. Hydro Environ. Res.*
- Sinclair, S., Pegram, G., 2005. Combining radar and rain gauge rainfall estimates using conditional merging. *Atmos. Sci. Lett.* 6, 19–22.
- Sorooshian, S., Hsu, K.-L., Gao, X., Gupta, H.V., Imam, B., Braithwaite, D., 2000. Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Am. Meteorol. Soc.* 81, 2035–2046.
- Tang, L., Tian, Y., Lin, X., 2014. Validation of precipitation retrievals over land from satellite-based passive microwave sensors. *J. Geophys. Res.-Atmos.* 119, 4546–4567.
- Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., De Roo, A., 2013. Hydrological evaluation of satellite-based rainfall estimates over the volta and Baro-Akobo Basin. *J. Hydrol.* 499, 324–338.
- Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., Levizzani, V., De Roo, A., 2012. Validation of satellite-based precipitation products over sparsely gauged African river basins. *J. Hydrometeorol.* 13, 1760–1783.
- Valdés-Pineda, R., Pizarro, R., García-Chevesich, P., Valdés, J.B., Olivares, C., Vera, M., Balocchi, F., Pérez, F., Vallejos, C., Fuentes, R., Abarza, A., Helwig, B., 2014. Water governance in Chile: availability, management and climate change. *J. Hydrol.* 519, 2538–2567.
- Verbist, K., Robertson, A.W., Cornelis, W.M., Gabriels, D., 2010. Seasonal predictability of daily rainfall characteristics in central northern Chile for dry-land management. *J.*

APPENDIX B.

O.M. Baez-Villanueva, et al

Remote Sensing of Environment 239 (2020) 111606

- Appl. Meteorol. Climatol. 49, 1938–1955.
- Verdin, A., Funk, C., Rajagopalan, B., Kleiber, W., 2016. Kriging and local polynomial methods for blending satellite-derived and gauge precipitation estimates to support hydrologic early warning systems. *IEEE Trans. Geosci. Remote Sens.* 54, 2552–2562.
- Villarini, G., Krajewski, W.F., 2008. Empirically-based modeling of spatial sampling uncertainties associated with rainfall measurements by rain gauges. *Adv. Water Resour.* 31, 1015–1023.
- Villarini, G., Mandapaka, P.V., Krajewski, W.F., Moore, R.J., 2008. Rainfall and sampling uncertainties: a rain gauge perspective. *J. Geophys. Res.-Atmos.* 113.
- Wang, C., Tang, G., Han, Z., Guo, X., Hong, Y., 2018. Global intercomparison and regional evaluation of GPM imerg version-03, version-04 and its latest version-05 precipitation products: similarity, difference and improvements. *J. Hydrol.* 564, 342–356.
- Weiss, L.L., Wilson, W.T., 1953. Evaluation of significance of slope changes in double-mass curves. *Trans. Am. Geophys. Union* 34, 893–896.
- Woldemeskel, F.M., Sivakumar, B., Sharma, A., 2013. Merging gauge and satellite rainfall with specification of associated uncertainty across Australia. *J. Hydrol.* 499, 167–176.
- Xie, P., Joyce, R., Wu, S., Yoo, S.-H., Yarosh, Y., Sun, F., Lin, R., 2017. Reprocessed, bias-corrected CMORPH global high-resolution precipitation estimates from 1998. *J. Hydrometeorol.* 18, 1617–1641.
- Xie, P., Xiong, A.-Y., 2011. A conceptual model for constructing high-resolution gauge-satellite merged precipitation analyses. *J. Geophys. Res.-Atmos.* 116, D21106.
- Yang, Z., Hsu, K., Sorooshian, S., Xu, X., Braithwaite, D., Zhang, Y., Verbist, K.M.J., 2017. Merging high-resolution satellite-based precipitation fields and point-scale rain gauge measurements - a case study in Chile. *J. Geophys. Res.-Atmos.* 122, 5267–5284.
- Zambrano-Bigiarini, M., 2017a. hydroGOF: goodness-of-fit functions for comparison of simulated and observed hydrological time series. R package version 0.3-10. <https://doi.org/10.5281/zenodo.840087>. <http://hzambran.github.io/hydroGOF/>.
- Zambrano-Bigiarini, M., 2017b. hydroTSM: Time Series Management, Analysis and Interpolation for Hydrological Modelling. R package version 0.5-1. <https://doi.org/10.5281/zenodo.83964>. <https://github.com/hzambran/hydroTSM>.
- Zambrano-Bigiarini, M., 2018. Temporal and spatial evaluation of long-term satellite-based precipitation products across the complex topographical and climatic gradients of Chile. In: *Remote Sensing and Modeling of the Atmosphere, Oceans, and Interactions VII International Society for Optics and Photonics*, pp. 1078202. <https://doi.org/10.1117/12.2513645>.
- Zambrano-Bigiarini, M., Nauditt, A., Birkel, C., Verbist, K., Ribbe, L., 2017. Temporal and spatial evaluation of satellite-based rainfall estimates across the complex topographical and climatic gradients of Chile. *Hydrol. Earth Syst. Sci.* 21, 1295–1320. <https://doi.org/10.5194/hess-21-1295-2017>.
- Zhao, Y., Feng, D., Yu, L., Wang, X., Chen, Y., Bai, Y., Hernández, H.J., Galleguillos, M., Estades, C., Biging, G.S., Radke, J.D., Gong, P., 2016. Detailed dynamic land cover mapping of Chile: accuracy improvement by integrating multi-temporal data. *Remote Sens. Environ.* 183, 170–185.

B.1 Supplement material

APPENDIX B.

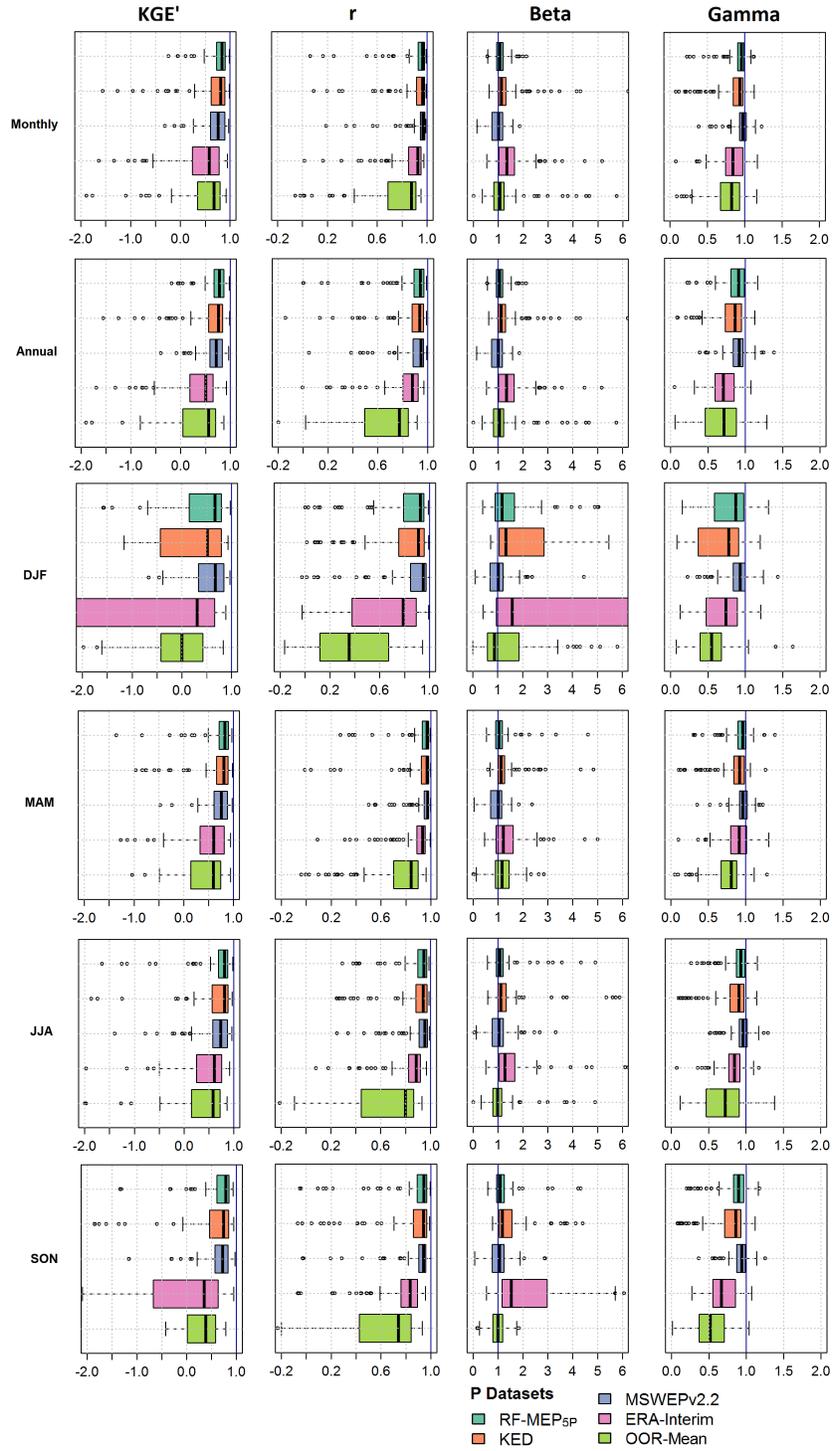


Figure S1: The KGE' (a), r (b), β (c), and γ (d) for RF-MEP_{5P}, Kriging with external drift (KED), MSWEPv2.2, ERA-Interim, and one-outlier-removed (OOR) arithmetic mean products using the ground-based validation dataset.

APPENDIX B.

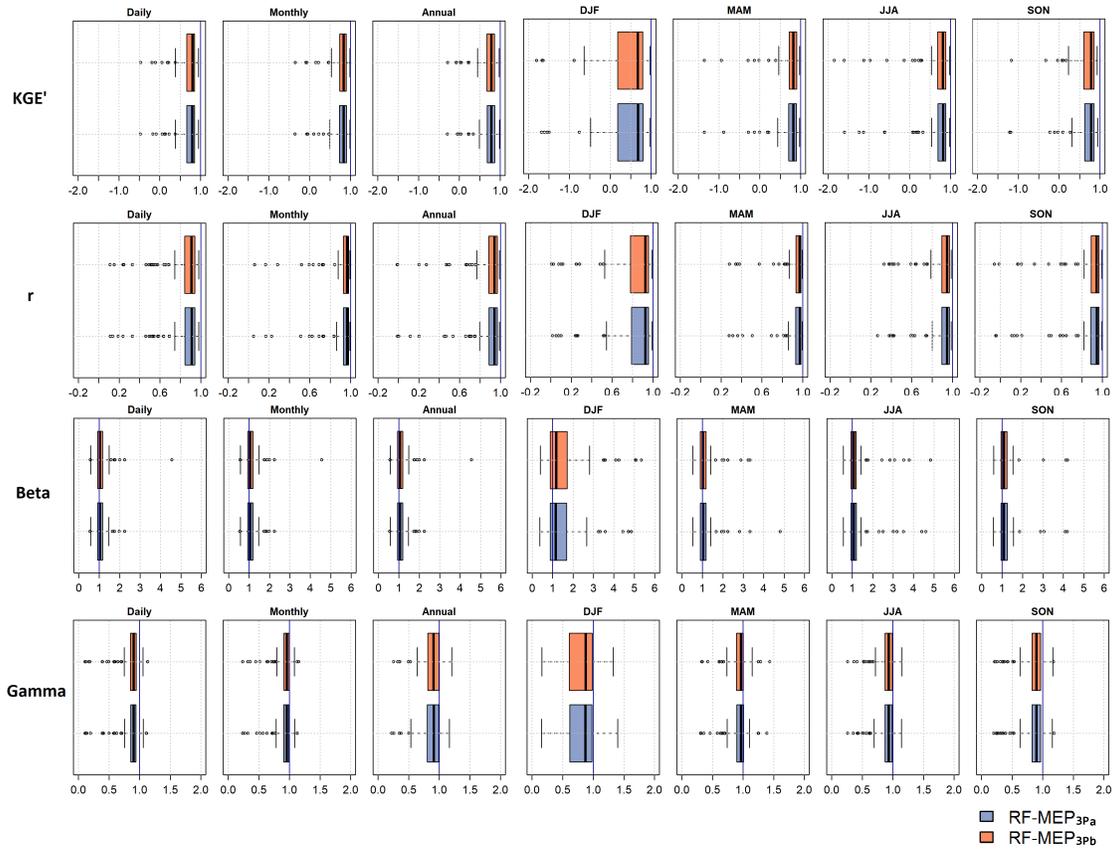


Figure S2: Performance of two RF-MEP_{3P} products according to the KGE' and its components. RF-MEP_{3Pa} was generated using ERA-Interim, PERSIANN-CDR, and CHIRPSv2, while RF-MEP_{3Pb} used ERA-Interim, CMORPHv1, and TRMM 3B42v7.

APPENDIX B.

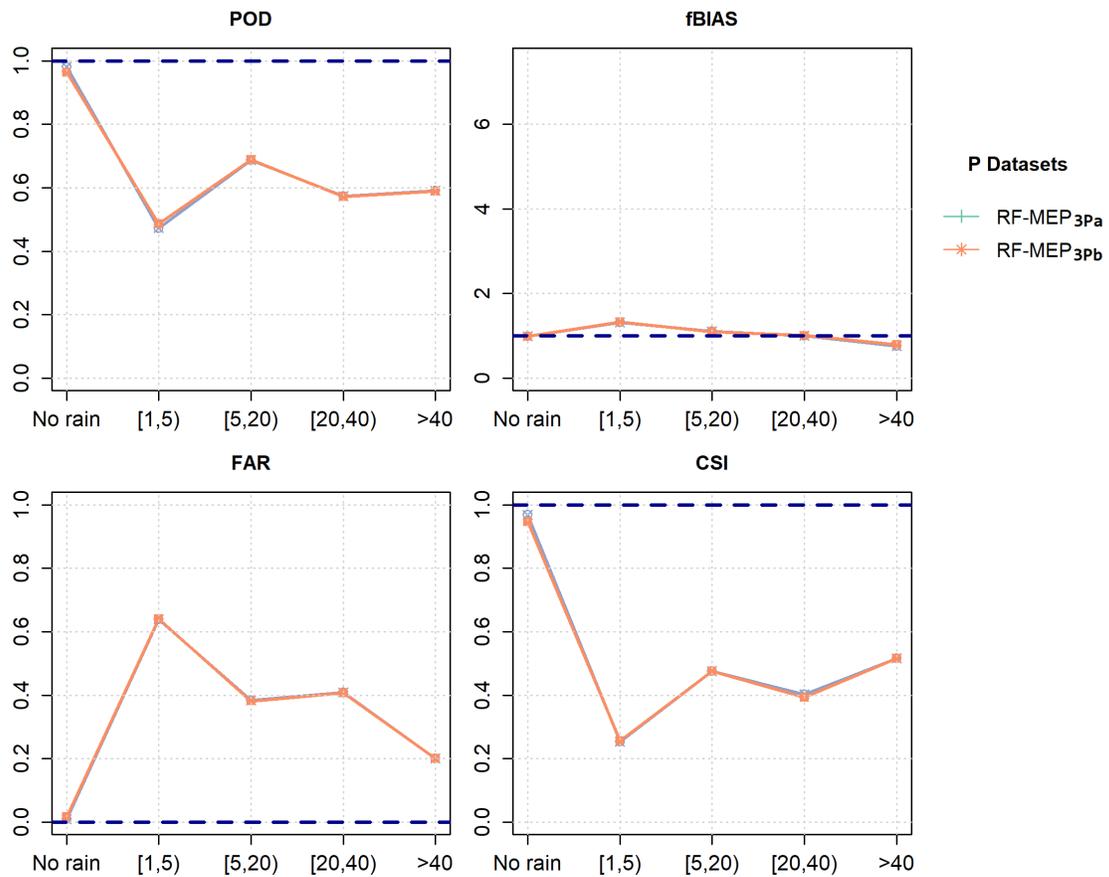


Figure S3: Performance for two RF-MEP_{3P} products according to four categorical indices. RF-MEP_{3Pa} was generated using ERA-Interim, PERSIANN-CDR, and CHIRPSv2, while RF-MEP_{3Pb} used ERA-Interim, CMORPHv1, and TRMM 3B42v7. From top to the bottom and left to right: POD, *f*BIAS, FAR, and CSI.

Appendix C

Regionalisation step

APPENDIX C.

**On the selection of precipitation products for the
regionalisation of hydrological model parameters**

Oscar Manuel Baez-Villanueva; Mauricio Zambrano-Bigiarini;
Lars Ribbe; Alexandra Nauditt; Juan Diego Giraldo-Osorio; Nguyen
Xuan Thinh

Hydrology and Earth System Sciences, Volume 25, 2021,
Pages 5805-5837.

<https://doi.org/10.5194/hess-25-5805-2021>



On the selection of precipitation products for the regionalisation of hydrological model parameters

Oscar M. Baez-Villanueva^{1,2}, Mauricio Zambrano-Bigiarini^{3,4}, Pablo A. Mendoza^{5,6}, Ian McNamara¹, Hylke E. Beck⁷, Joschka Thurner¹, Alexandra Nauditt¹, Lars Ribbe¹, and Nguyen Xuan Thinh²

¹Institute for Technology and Resources Management in the Tropics and Subtropics (ITT), TH Köln, Cologne, Germany

²Faculty of Spatial Planning, TU Dortmund University, Dortmund, Germany

³Department of Civil Engineering, Universidad de la Frontera, Temuco, Chile

⁴Center for Climate and Resilience Research, Universidad de Chile, Santiago, Chile

⁵Department of Civil Engineering, Universidad de Chile, Santiago, Chile

⁶Advanced Mining Technology Center (AMTC), Universidad de Chile, Santiago, Chile

⁷GloH2O, Almere, the Netherlands

Correspondence: Mauricio Zambrano-Bigiarini (mauricio.zambrano@ufrontera.cl)

Received: 16 March 2021 – Discussion started: 23 April 2021

Revised: 11 August 2021 – Accepted: 12 October 2021 – Published: 11 November 2021

Abstract. Over the past decades, novel parameter regionalisation techniques have been developed to predict streamflow in data-scarce regions. In this paper, we examined how the choice of gridded daily precipitation (P) products affects the relative performance of three well-known parameter regionalisation techniques (spatial proximity, feature similarity, and parameter regression) over 100 near-natural catchments with diverse hydrological regimes across Chile. We set up and calibrated a conceptual semi-distributed HBV-like hydrological model (TUWmodel) for each catchment, using four P products (CR2MET, RF-MEP, ERA5, and MSWEPv2.8). We assessed the ability of these regionalisation techniques to transfer the parameters of a rainfall-runoff model, implementing a leave-one-out cross-validation procedure for each P product. Despite differences in the spatio-temporal distribution of P , all products provided good performance during calibration (median Kling–Gupta efficiencies ($KGE's$) > 0.77), two independent verification periods (median $KGE's > 0.70$ and 0.61 , for near-normal and dry conditions, respectively), and regionalisation (median $KGE's$ for the best method ranging from 0.56 to 0.63). We show how model calibration is able to compensate, to some extent, differences between P forcings by adjusting model parameters and thus the water balance components. Overall, feature similarity provided the best results, followed by spatial proximity, while parameter regression resulted in the worst performance, reinforcing the

importance of transferring complete model parameter sets to ungauged catchments. Our results suggest that (i) merging P products and ground-based measurements does not necessarily translate into an improved hydrologic model performance; (ii) the spatial resolution of P products does not substantially affect the regionalisation performance; (iii) a P product that provides the best individual model performance during calibration and verification does not necessarily yield the best performance in terms of parameter regionalisation; and (iv) the model parameters and the performance of regionalisation methods are affected by the hydrological regime, with the best results for spatial proximity and feature similarity obtained for rain-dominated catchments with a minor snowmelt component.

1 Introduction

Daily streamflow (Q) data are crucial for a wide range of scientific and operational water resources applications, such as climate change impact assessment (e.g. Kling et al., 2012; Rojas et al., 2013; Mendoza et al., 2016; Galleguillos et al., 2021), Q and flood forecasting (e.g. Clark and Hay, 2004; Addor et al., 2011; Coughlan de Perez et al., 2016; Sharma et al., 2018), and catchment classification (e.g. Wagener et al., 2007; Sawicz et al., 2011; Kuentz et al., 2017; Jehn

et al., 2020), among others. Q is typically estimated through the implementation of hydrological models, which rely on parameters to represent hypotheses about the dominant processes in a catchment (Beven, 2006). In most cases, these parameters cannot be measured at the scales relevant for model applications (Beven, 1989; Uhlenbrook et al., 1999; Beven, 2000; Wagener et al., 2001) and are therefore estimated through model calibration. To this end, optimisation techniques are used to provide reliable estimates of model parameters, requiring the comparison of observed Q against simulated Q data (Yapo et al., 1998; Vrugt et al., 2003, 2009; Pokhrel et al., 2012; Shafii and Tolson, 2015; Pool et al., 2017). Because the vast majority of streams worldwide remain ungauged (Young, 2006; Beck et al., 2016), the scientific initiative Prediction in Ungauged Basins (PUB; see review by Hrachowitz et al., 2013) has fostered the development of novel regionalisation techniques to predict Q in ungauged basins, a task that is far from complete (Yang et al., 2019; Dallery et al., 2020). The spatial transfer of hydrological model parameters from monitored to ungauged catchments, a process known as regionalisation (Oudin et al., 2008), remains an active research topic (see review by Guo et al., 2021).

In the hydrological modelling literature, there are three main regionalisation approaches (Oudin et al., 2008; Parajka et al., 2013): (i) spatial proximity, (ii) feature similarity, and (iii) parameter regression. Spatial proximity assumes that climatic and physiographic characteristics are relatively homogeneous within a region, and, therefore, neighbouring catchments exhibit similar hydrological behaviour (Vandewiele and Elias, 1995; Oudin et al., 2008). Although this method requires a dense network of gauging stations to perform well, it may lead to inadequate representations of rainfall-runoff behaviour over areas with heterogeneous climate and geomorphological characteristics (Beck et al., 2016). Feature similarity techniques transfer calibrated model parameter sets from donor to ungauged catchments based on geomorphological and climatic similarities (McIntyre et al., 2005; Carrillo et al., 2011; Beck et al., 2016). Finally, parameter regression methods develop statistical relationships between calibrated model parameters and catchment characteristics, which are subsequently used to estimate parameter values for ungauged catchments (Fernandez et al., 2000; Carrillo et al., 2011). Recently, Samaniego et al. (2010) and Beck et al. (2020a) applied multiscale parameter regionalisation techniques that link model parameters to predictors related to geomorphological and climatological characteristics by optimising coefficients in transfer equations, which helps to account for problems related to equifinality. The performances of these three regionalisation techniques vary due to many factors, including the selected sample of catchments, the presence of nested catchments, hydroclimatic conditions, physiographic catchment properties, model configuration (including meteorological forcings, model structure,

and simulation setup), and evaluation criteria (Parajka et al., 2013; Neri et al., 2020; Guo et al., 2021).

Most regionalisation studies have been conducted over regions with a dense network of meteorological stations (see Table 1), including Europe (e.g. McIntyre et al., 2005; Parajka et al., 2005; Oudin et al., 2008; Singh et al., 2012; Zelelew and Alfreksen, 2014; Garambois et al., 2015; Rakovec et al., 2016; Neri et al., 2020), the conterminous United States (Athira et al., 2016; Saadi et al., 2019), India (Swain and Patra, 2017), and China (Bao et al., 2012). However, in developing countries, P has traditionally been estimated through interpolation within sparse rain gauge networks, which is subject to large uncertainties (Hofstra et al., 2010; Woldemeskel et al., 2013; Adhikary et al., 2015; Xavier et al., 2016), hindering an accurate spatio-temporal representation of P patterns. Over the last decades, the emergence of near-global and high-resolution gridded P products has introduced new possibilities for hydrological modelling in data-scarce regions (Maggioni and Massari, 2018; Sun et al., 2018), despite these products still being affected by systematic, random, and detection errors (Ren and Li, 2007; Sevruk et al., 2009; Zambrano-Bigiarini et al., 2017; Baez-Villanueva et al., 2018), which are more pronounced over mountainous regions (Maggioni and Massari, 2018; Beck et al., 2019). Although hydrological model calibration can partly compensate for errors in the representation of P (Elsner et al., 2014; Maggioni and Massari, 2018), this may lead to unrealistic model behaviour (Nikolopoulos et al., 2013; Xue et al., 2013; Ciabatta et al., 2016), thus affecting the quality of parameter regionalisation results.

To date, few regionalisation studies have used gridded P products at the daily temporal scale. Beck et al. (2016) used the Climate Prediction Center unified gauge-based P product (CPC) to provide spatially distributed HBV parameters at the global scale. They selected CPC because it yielded better performance than ERA-Interim during calibration. Rakovec et al. (2016) used the European daily high-resolution gridded dataset (E-OBSv8.0) to force a mesoscale hydrological model over 400 catchments in Europe, providing regionalised model parameters through a multivariate parameter estimation technique. More recently, Beck et al. (2020a) combined MSWEPv2.2 with a novel multiscale parameter regionalisation approach to provide global gridded parameter estimates using daily Q observations from 4229 catchments. Although these studies have successfully used gridded P products for parameter regionalisation, they only selected one product, and thus the effects that the choice of a P dataset can have on regionalisation results remain unknown. This study aims to answer the following questions:

- (i) To what extent does the choice of gridded P forcing used in calibration affect the relative performance of regionalisation techniques?
- (ii) How does this relative performance vary across catchments with different hydrological regimes?

2 Study area and selection of catchments

Our study domain is continental Chile (Fig. 1), which is bounded to the west by the Pacific Ocean, to the north by Peru, and to the east by Bolivia and Argentina. The territory spans 4300 km of latitudinal extension ($17.5\text{--}56.0^\circ\text{S}$) and on average 180 km of longitudinal extension ($76.0\text{--}66.0^\circ\text{W}$), with elevation (Jarvis et al., 2008) ranging from 0 to 6892 m a.s.l. in the Andes Mountains. Figure 1 shows the elevation, land cover (Zhao et al., 2016), Köppen–Geiger climate classification (Beck et al., 2018), and hydrological regimes for the five major macroclimatic zones presented in Zambrano-Bigiarini et al. (2017). A large variety of climates are present across the country, with the macroclimatic zones transitioning from the (hyper)arid and semi-arid climates in the Far North ($17.50\text{--}26.00^\circ\text{S}$) and Near North ($26.00\text{--}32.18^\circ\text{S}$), through temperate climates in Central Chile ($32.18\text{--}36.40^\circ\text{S}$), to more humid and polar climates in the South ($36.40\text{--}43.70^\circ\text{S}$) and Far South ($43.70\text{--}56.00^\circ\text{S}$). P increases with elevation and latitude (in the southern direction), ranging from almost zero in the Atacama Desert to $\sim 6000\text{ mm yr}^{-1}$ in the surroundings of Puerto Cárdenas ($\sim 43.2^\circ\text{S}$). Similar to the P patterns, both the mean annual Q and rainfall-runoff ratio tend to increase from north to south (Alvarez-Garreton et al., 2018; Vásquez et al., 2021).

The El Niño–Southern Oscillation (ENSO) has a large impact on winter P , with negative anomalies during La Niña and positive anomalies during El Niño events (Verbist et al., 2010; Robertson et al., 2014). Although neutral ENSO conditions have prevailed since 2011 (except for a strong El Niño event during 2015), an uninterrupted sequence of dry years with increased temperatures has been observed from 2010–2018, with annual P deficits of about 25%–45% across Chile. This long-term deficit in P volume, also known as the Chilean megadrought (Boisier et al., 2016; Garreaud et al., 2017), has reduced snow cover, river flows, reservoir storage, and groundwater levels across Chile (Garreaud et al., 2017, 2020).

Hydroclimatic indices and characteristics for 516 catchments in continental Chile were acquired from the Catchment Attributes and MEteorology for Large-sample Studies dataset in Chile (CAMELS-CL; Alvarez-Garreton et al., 2018). The dataset includes location, topography, geology, soil types, land cover, hydrological signatures, and human intervention degree, among others. Q data were obtained from the Center for Climate and Resilience Research (CR2; <http://www.cr2.cl/datos-de-caudales/>, last access: October 2020) for 1930–2018 because Q data from CAMELS-CL ended in 2016 at the time of conducting this study. We selected the near-natural catchments from the CAMELS-CL database that fulfilled the following criteria:

1. less than 25 % of missing values in the daily Q time series for 1990–2018 (may be non-consecutive)

2. absence of large dams ($\text{big_dam} = 0$)
3. less than 10 % of Q allocated to consumptive uses ($\text{interv_degree} < 0.1$)
4. not dominated by glaciers ($\text{lc_glacier} < 5\%$)
5. less than 5 % of the area defined as urban ($\text{imp_frac} < 5\%$)
6. absence of substantial irrigation abstractions ($\text{crop_frac} < 20\%$)
7. less than 20 % of the area covered by forest plantations ($\text{fp_frac} < 20\%$)
8. no signs of artificial regulation in the hydrograph (10 excluded in total).

The drainage areas of the selected catchments (100) range from 35 to 11 137 km², with a median value of 645 km². The selected catchments contain 42 nested catchments (i.e. catchments that are contained in a larger catchment). We adjusted the classification of these catchments according to hydrological regime, building on the classifications presented in several national and regional technical reports (e.g. DGA, 1998, 1999, 2004a, b, c, 2006, 2016a, b, 2018), by visually analysing the contribution of solid and liquid P to the mean monthly Q values. These regimes were classified as (i) snow-dominated; (ii) nivo-pluvial, i.e. snow-dominated with a rain component; (iii) pluvio-nival, i.e. rain-dominated with a snow component; and (iv) rain-dominated, as shown in Fig. 1d. Figure A1 shows conceptual hydrographs for each of these regimes and is presented in Appendix A.

3 Methods

3.1 Meteorological forcings

3.1.1 Precipitation products

Four P products were used to investigate how the choice of P forcing affects the performance of regionalisation techniques. The P products are presented in Table 2 and were selected because previous studies have reported good agreement when evaluated against in situ measurements over continental Chile (Zambrano-Bigiarini et al., 2017; Boisier et al., 2018; Baez-Villanueva et al., 2018, 2020).

The Center for Climate and Resilience Research Meteorological dataset version 2.0 (CR2MET; Boisier et al., 2018) provides daily gridded P estimates over continental Chile at a 5 km spatial resolution for 1979–2018. These estimates are produced by combining rain gauge observations with re-analysis data from ERA5, while CR2MET version 1.0 of this product was produced using ERA-Interim data (Boisier et al., 2018). As CR2MET was developed specifically for Chile and uses all the Chilean rain gauges (874 across Chile; see Fig. S1

Table 1. Summary of selected regionalisation studies that used spatial proximity (SP), feature similarity (FS), parameter regression (PR), or multiscale parameter regionalisation (MPR). This study has been added for completeness.

Study	Region	Catchments (donor/evaluation)	Approach	Relevant conclusion
McIntyre et al. (2005)	United Kingdom	127/leave-one-out cross-validation	SP and FS	The transfer of complete model parameter sets increased the performance of regionalisation. The use of the 10 best model parameter sets provided a more robust representation of flood peaks and generated a better ensemble of the overall flow regime, although flow peaks were underestimated. A comparison against the PR approach showed that FS produced better results.
Parajka et al. (2005)	Austria	320/leave-one-out cross-validation	SP, FS, and PR	All methods performed better than the average of the model parameters of all catchments. Two methods performed the best: FS and an SP kriging approach, where the model parameters were regionalised independently based on their spatial correlation. Local regression methods outperformed the global regression method, highlighting the importance of accounting for regional differences during PR.
Oudin et al. (2008)	France	913/leave-one-out cross-validation	SP, FS, and PR	SP performed the best, followed closely by FS. The reduced performance of FS was attributed to the lack of soil-related properties used as inputs. To construct the ensemble output using multiple catchments, averaging the Q time series performed better than averaging the model parameters. They concluded that the dense network of catchments favoured the SP method.
Samaniego et al. (2010)	Germany	1/10 stations within the study area	MPR	The MPR method showed improved results compared to the standard PR when the global parameters were calibrated at a coarser modelling scale and then transferred to a finer scale.
Bao et al. (2012)	China	55/leave-one-out cross-validation	FS and PR	FS outperformed PR over both humid and arid regions. Moving from humid to arid regions, the degree to which the FS approach outperformed PR increased.
Zezelew and Alfredeisen (2014)	Southern Norway	11/Leave-one-out cross-validation	SP and FS	The ensemble of the 10 most similar catchments outperformed the other approaches (the performance increased when two to six catchments were used). They recommended identifying the parameters that influence the model response in order to minimise the model parametric dimensionality.
Garambois et al. (2015)	Southern France	16/leave-one-out cross-validation	SP and FS	FS outperformed SP. They reported only a small decrease in performance from calibration/verification to regionalisation ($\sim 10\%$) when evaluated during flash flood events. Using an ensemble of two to four donor catchments yielded the best regionalisation performance. Using well-modelled catchments does not always produce good performances during regionalisation, and parameter sets from low-performing catchments can produce higher performances when transferred to ungauged settings.
Athira et al. (2016)	Conterminous United States	8/leave-one-out cross-validation	PR	The parameter values using multi-linear regression models were different to those obtained through model calibration, indicating the deficiency of regionalising the parameters directly as a function of catchment attributes. For the one catchment where SP was also tested, PR performed better.
Beck et al. (2016)	Global	674/1113; independent evaluation	FS	The derived global maps of HBV parameter sets conform well with large-scale climate patterns, demonstrating the effect of climate on rainfall-runoff patterns. For 79 % of catchments, the averaging of model outputs (from 10 donor catchments) outperformed the use of spatially uniform parameters. P underestimation appeared to be the dominant cause of low calibration scores, particularly for tropical and arid catchments.
Rakovec et al. (2016)	Europe	36/400, cross-validation	MPR	The model performed well in simulating daily Q over a wide range of physiographic and climatic conditions, with median KGE's greater than 0.55. This performance was reduced in heavily regulated catchments. Further evaluation against complementary datasets showed the best agreement for evaporation, followed by total water storage, and the lowest for soil moisture.

Table 1. Continued.

Study	Region	Catchments (donor/evaluation)	Approach	Relevant conclusion
Swain and Patra (2017)	India	32/leave-one-out cross-validation	SP, FS, and PR	SP (both kriging and inverse distance weighting, IDW) outperformed PR and FS. The methods were evaluated against a global mean approach, which produced worse results than all tested regionalisation methods.
Beck et al. (2020a)	Global	4229/10-fold cross-validation	MPR	They incorporated within-catchment variability in climate and landscape and yielded an improvement in 88 % of the catchments (median KGE' improved from 0.19 to 0.46). They found a weak positive correlation between regionalisation performance and catchment humidity. Considerable improvements were obtained for catchments located both near and far from those used for optimisation. Q simulation performance was best in humid regions and worst in arid regions.
Neri et al. (2020)	Austria	209/leave-one-out cross-validation	SP and FS	Compared to the results of the independent calibration/verification, the regionalisation performance using the TUWmodel deteriorated less than using the GR6J model. With a high density of gauged stations, both the SP and FS performed similarly well, but the results deteriorated with reduced gauge density (especially for SP). Transferring the parameter sets of more than one single catchment improves the regionalisation performance.
This study	Chile	100/leave-one-out cross-validation	SP, FS, and PR	FS was the best-performing method, followed by SP. The use of merged P products does not necessarily translate into an improved hydrological modelling performance. Strong performance of a P product for calibration and validation does not necessarily translate into strong performance for regionalisation. The performance of regionalisation methods depends on the hydrological regime.

Table 2. Gridded P products used in this study.

P product	Period	Spatial and temporal resolution	References
CR2MET	1979–2018	0.05°; daily	Boisier et al. (2018)
RF-MEP	1983–2018	0.05°; daily	Baez-Villanueva et al. (2020)
ERA5	1950–present	~0.28°; hourly	Hersbach et al. (2020)
MSWEPv2.8	1979–present	0.10°; 3-hourly	Beck et al. (2017b, 2019)

in the Supplement), it is considered as the “reference” P product of Chile.

The random forest merging procedure (RF-MEP; Baez-Villanueva et al., 2020) combines gridded P products, ground-based measurements, and other spatial covariates to generate P estimates. We applied this methodology to generate a spatially distributed, daily P product for continental Chile, using daily records from 334 rain gauges (obtained from CR2; <http://www.cr2.cl/datos-de-precipitacion/>, last access: 10 January 2021), gridded P data from the ERA5 reanalysis (Hersbach et al., 2020) aggregated to the Chilean time, and elevation (SRTMv4.1; Jarvis et al., 2008) as covariates. This RF-MEP version 2 product (hereafter, RF-MEP) was generated for 1990–2018 with a spatial resolution of 0.05° using the RFmerge R package (Zambrano-Bigiarini et al., 2020).

ERA5 (Hersbach et al., 2020) is a reanalysis product that provides hourly P estimates (as well as other variables) from 1950 to present at a spatial resolution of around 30 km (~0.28°). There are important improvements in its P estimates compared to its predecessor ERA-Interim, such as improved (i) representation of mixed-phase clouds, (ii) prognostics variables for rain and snow, (iii) parameterisation of microphysics, and (iv) representation of tropical variability (Hersbach et al., 2020). Although ERA5 also assimilates NCEP Stage IV P estimates over the conterminous United States, which combine NEXRAD data with in situ measurements, it does not incorporate information from any ground-based P stations over Chile. Hourly ERA5 estimates were aggregated into daily P values, taking into account the reporting times of the Chilean rain gauges (08:00–07:59 local time, which represents 11:00–10:59 UTC). Although this product has a relatively low spatial resolution com-

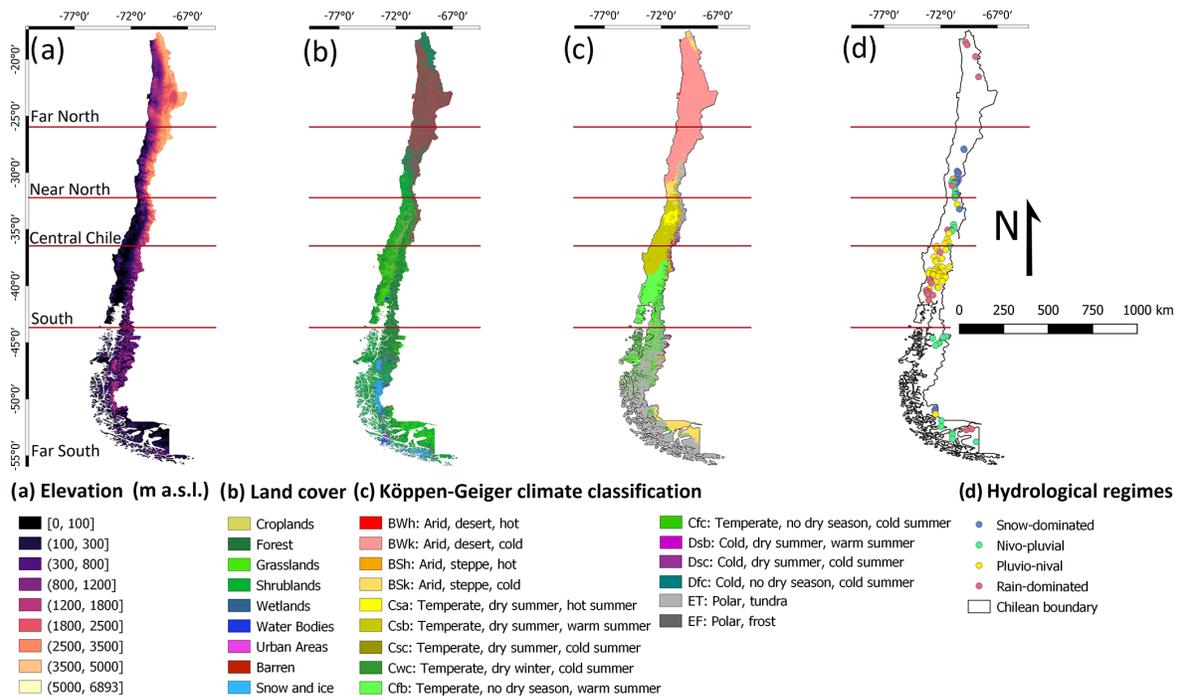


Figure 1. Study area: (a) elevation (SRTMv4.1; Jarvis et al., 2008); (b) land cover classification (Zhao et al., 2016); (c) Köppen–Geiger climate classification (Beck et al., 2018); and (d) hydrological regimes of the selected catchments over the five major macroclimatic zones according to Zambrano-Bigiarini et al. (2017).

pared to the other selected products, we included it because (i) Chile is dominated by large-scale, frontal systems (Zhang and Wang, 2021), and therefore, coarse-resolution products may perform well even over small catchments; (ii) reanalysis products tend to perform well at high latitudes (Beck et al., 2017a); and (iii) we consider that its inclusion represents a realistic situation that may exist in many practical applications (i.e. where a catchment size is small relative to *P* product resolution).

The Multi-Source Weighted-Ensemble Precipitation (MSWEPv2.8; Beck et al., 2017b, 2019) is a 3-hourly *P* product with a spatial resolution of 0.10°, which takes advantage of the complementary strengths of satellite, reanalysis, and ground-based data. MSWEPv2.8 applies daily and monthly corrections to its estimates using data from around 77 000 rain gauge stations globally (628 of these are over Chile; see Fig. S1), accounting for their local reporting times. The 3-hourly MSWEPv2.8 estimates were also aggregated into daily *P* to account for the difference in the reporting times.

Figure 2a shows the spatial distribution of mean annual *P* for all products over 1990–2018, while Fig. 2b shows box plots of the mean monthly *P* averaged over catchments located within each macroclimatic zone. All *P* products show relatively similar patterns of spatial variability across con-

tinental Chile; however, there are substantial differences in their total *P* amounts. In general, *P* increases from the (hyper-arid) Far North to the South and decreases again in the Far South. *P* also increases from the west coast towards the Andes Mountains. ERA5 provides higher *P* amounts over all five macroclimatic zones, while RF-MEP generally yields the lowest annual *P* values. Over the Far North, all products show a marked rainy season during December–March due to summer convective *P*, which differs from the marked seasonality evident over the Near North, Central Chile, and South regions. Over the Far North, ERA5 presents the highest mean annual *P* (157 mm), which is almost twice the amount provided by the second-highest product MSWEPv2.8 (83 mm), followed by CR2MET (63 mm), while RF-MEP has the lowest mean annual *P* (40 mm). Although ERA5 presents the highest mean annual *P* values over the Near North, Central Chile, and South regions (208, 902, and 2172 mm, respectively), when considering only our case study catchments (Fig. 2b), CR2MET has the highest mean monthly values over the Central Chile and South regions during April–June. RF-MEP and MSWEPv2.8 have similar mean annual *P* values over Central Chile (670 mm for both products) and the South (1670 and 1735 mm, respectively) regions, although RF-MEP consistently shows the largest monthly *P* amounts of the two products over the

corresponding catchments. ERA5 provides the highest mean annual P values over the Far South (3018 mm), followed by CR2MET (1888 mm), MSWEPv2.8 (1714 mm), and RF-MEP (815 mm). Finally, each product shows low seasonality over the Far South. Here, ERA5 presents higher monthly P values throughout the year, with the largest difference from the other products between January–March and September–December.

To gain a deeper understanding of the differences between the four P products, we examined the spatial distribution of median annual values of four Climdex indices (Karl et al., 1999) for 1990–2018 (Fig. 3). First, to account for days without rain ($P < 1$ mm), we used the consecutive dry days index (CDD; Fig. 3a), which retrieves the maximum dry spell length. It is evident that CR2MET yields longer dry spells, mainly across the Far North and Near North regions, while ERA5 has shorter dry spells over these regions, especially over the Andes Mountains. CR2MET, RF-MEP, and MSWEPv2.8 have similar spatial patterns over the Central Chile and South regions, while ERA5 has fewer consecutive dry days over the Andes Mountains. Similarly, ERA5 provides shorter dry spells over the Far South, while CR2MET and RF-MEP present similar patterns. These results are consistent with the consecutive wet days index (CWD; Fig. 3b), which assesses the frequency and intermittency of P . ERA5 provides the highest CWD values over the driest regions (Far North and Near North), with medians ranging from 0 to 25 d, followed by MSWEPv2.8 (0 to 15 d). ERA5 also shows higher CWD values over high-elevation areas in Central Chile, while the remaining products show similar spatial patterns to each other. The four products show agreement in the CWD over the South region, with values ranging from 5 to 25 d. Finally, RF-MEP shows the lowest consecutive days with P in the Far South, followed by CR2MET and MSWEPv2.8, while ERA5 shows substantially higher CWD values at latitudes greater than 47° S.

To characterise high P intensities, we used the Rx5day (Fig. 3c) and R95pTOT (Fig. 3d) indices, which represent the maximum P accumulated over 5 consecutive days and the total P above the 95th percentile of the daily P for wet days, respectively. Figure 3c shows that ERA5 and CR2MET generally yield the highest Rx5day values, followed by MSWEPv2.8 and RF-MEP. A similar spatial variability is obtained with R95pTOT (Fig. 3d), indicating that there is a greater contribution of P from extreme events in ERA5 over high-elevation areas. These spatial patterns are replicated to some extent by CR2MET, which provides R95pTOT values up to 1200 mm over the Andes Mountains in Central Chile.

3.1.2 Air temperature and potential evaporation

Maximum and minimum daily air temperature (T) at a spatial resolution of 0.05° were taken from CR2MET. T is estimated using multivariate regression from the Moder-

ate Resolution Imaging Spectroradiometer (MODIS) land surface temperature (LST) and ERA5 estimates as covariates (Alvarez-Garreton et al., 2018; Boisier et al., 2018). The Hargreaves–Samani equation (Hargreaves and Samani, 1985) was used to obtain daily potential evaporation (PE) from CR2MET maximum and minimum daily T at the same spatial resolution (0.05°).

3.2 Hydrological model

The TUWmodel (Viglione and Parajka, 2020) is a conceptual hydrological model that follows the structure of the Hydrologiska Byråns Vattenbalansavdelning (HBV) model (Bergström, 1976; Bergström, 1995; Lindström, 1997). The model simulates the catchment-scale water balance at daily time steps, including processes related to snow accumulation and melting, change of moisture in the soil profile, and surface flow in the drainage network. The TUWmodel was validated over 320 catchments in Austria (Parajka et al., 2007) and has subsequently been used in numerous studies (e.g. Parajka et al., 2016; Zessner et al., 2017; Melsen et al., 2018; Slezciak et al., 2020). We selected a HBV-like conceptual model because it has shown good results in (i) many regionalisation studies (e.g. Parajka et al., 2005; Singh et al., 2012; Beck et al., 2016; Neri et al., 2020) and (ii) catchments with diverse hydroclimatic and geomorphological characteristics (Vetter et al., 2015; Ding et al., 2016; Unduche et al., 2018; Huang et al., 2019).

The TUWmodel requires as inputs daily time series of P , T , and PE . The parameters used by the TUWmodel to represent the hydrological processes are listed in Table 3, including the ranges selected for model calibration, which were adopted from previous studies (Parajka et al., 2007; Ceola et al., 2015) that calibrated the TUWmodel over a large number of mountainous catchments with snow influence. We ran the TUWmodel with a semi-distributed configuration for the period 1990–2018 based on meteorological and Q data availability. For each catchment, the number of equal-area elevation bands (EZ) was defined as $EZ = (H_{\max} - H_{\min})/200$, where H represents elevation. In cases where $EZ > 10$, EZ was set to 10 to reduce the computational demand of the simulations. Furthermore, in catchments with H_{\min} below 900 m a.s.l., the upper bound of the first EZ band was set to 900 m under the assumption that there is no snow influence below this elevation for the particular case of continental Chile. For more details about the TUWmodel implementation in R and the comparison of different HBV-like models, readers are referred to Astagneau et al. (2021) and Jansen et al. (2021), respectively.

3.3 Independent catchment calibration and verification

The simulation period used for this study was 1990–2018. For calibration purposes, we used the first 10 years as a conservative warm-up period to initialise the model stores, as

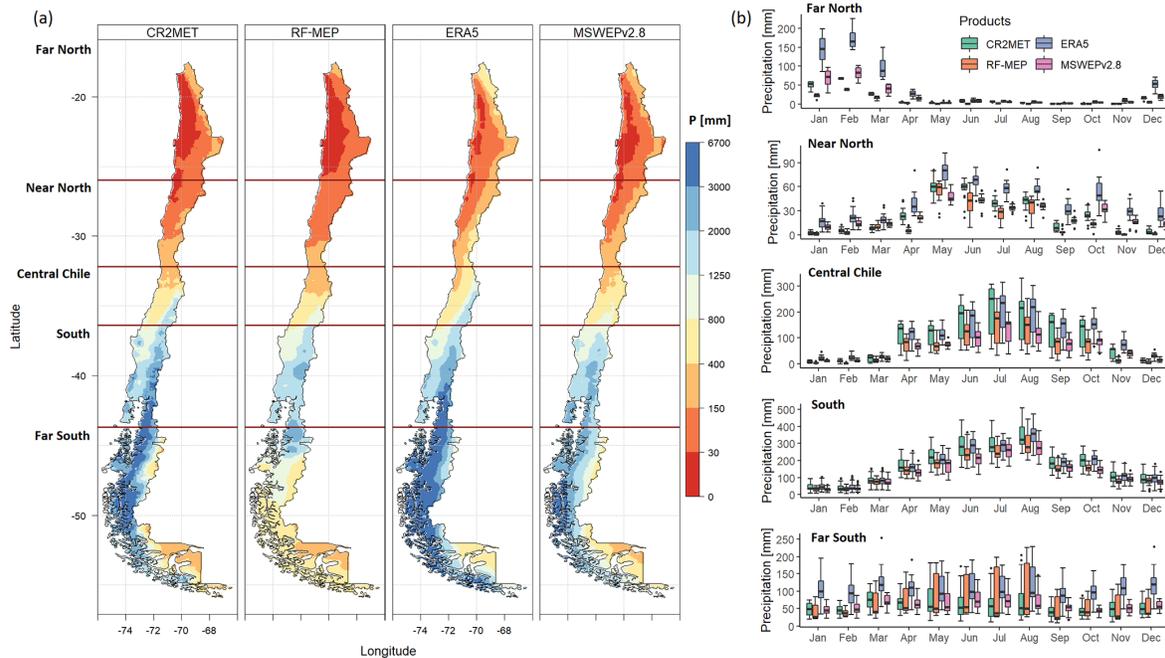


Figure 2. Comparison of *P* products over 1990–2018 (full time period): **(a)** mean annual *P* for each product resampled to a 0.05° spatial resolution using the nearest neighbour method. The dark red horizontal lines represent the limits of each major macroclimatic zone and **(b)** mean monthly *P* averaged over each catchment located within each macroclimatic zone (see Fig. 1d).

Table 3. Summary of the TUWmodel parameters considered for calibration, following the conceptualisation presented in Széles et al. (2020).

No.	Parameter ID	Description	Units	Process	Range
1	SCF	Snow correction factor	–	Snow	0.9–1.5
2	DDF	Degree-day factor	mm °C d ⁻¹	Snow	0.0–5.0
3	Twb	Wet bulb temperature	°C	Snow	–3.0–3.0
4	Tm	Threshold temperature above which melting starts	°C	Snow	–2.0–2.0
5	LPrat	Parameter related to the limit for potential evaporation	–	Evaporation	0.0–1.0
6	FC	Field capacity	mm	Infiltration	0.0–600
7	Beta	Non-linear parameter for runoff production	–	Infiltration	0.0–20
8	cperc	Constant percolation rate	mm d ⁻¹	Infiltration	0.0–8.0
9	k0	Storage coefficient for very fast response	d	Runoff	0.0–2.0
10	k1	Storage coefficient for fast response	d	Runoff	2.0–30
11	k2	Storage coefficient for slow response	d	Runoff	30–250
12	lsuz	Threshold storage state	mm	Runoff	1.0–100
13	bmax	Maximum base at low flows	d	Runoff	0.0–30
14	croute	Free scaling parameter	d ² mm ⁻¹	Runoff	0.0–50

in Beck et al. (2020a). The calibration period (2000–2014) includes near-normal conditions and the beginning of the Chilean megadrought. The first evaluation period (hereafter Verification 1, 1990–1999) represents near-normal/wet hydroclimatic conditions, while the second evaluation period (hereafter Verification 2, 2015–2018) spans the second half of the Chilean megadrought and was used to test the ability of the hydrological simulations to represent dry conditions.

To initialise model stores for the Verification 1 period, we used an 8-year warm-up period due to *P* product availability. We replicated Figs. 2 and 3 for these three periods to analyse the differences between the selected *P* products (see the Supplement, Figs. S2–S7).

We used the modified Kling–Gupta efficiency (KGE', Eq. 1; Kling et al., 2012) to calibrate the TUWmodel, which typically provides better hydrograph simulations than other

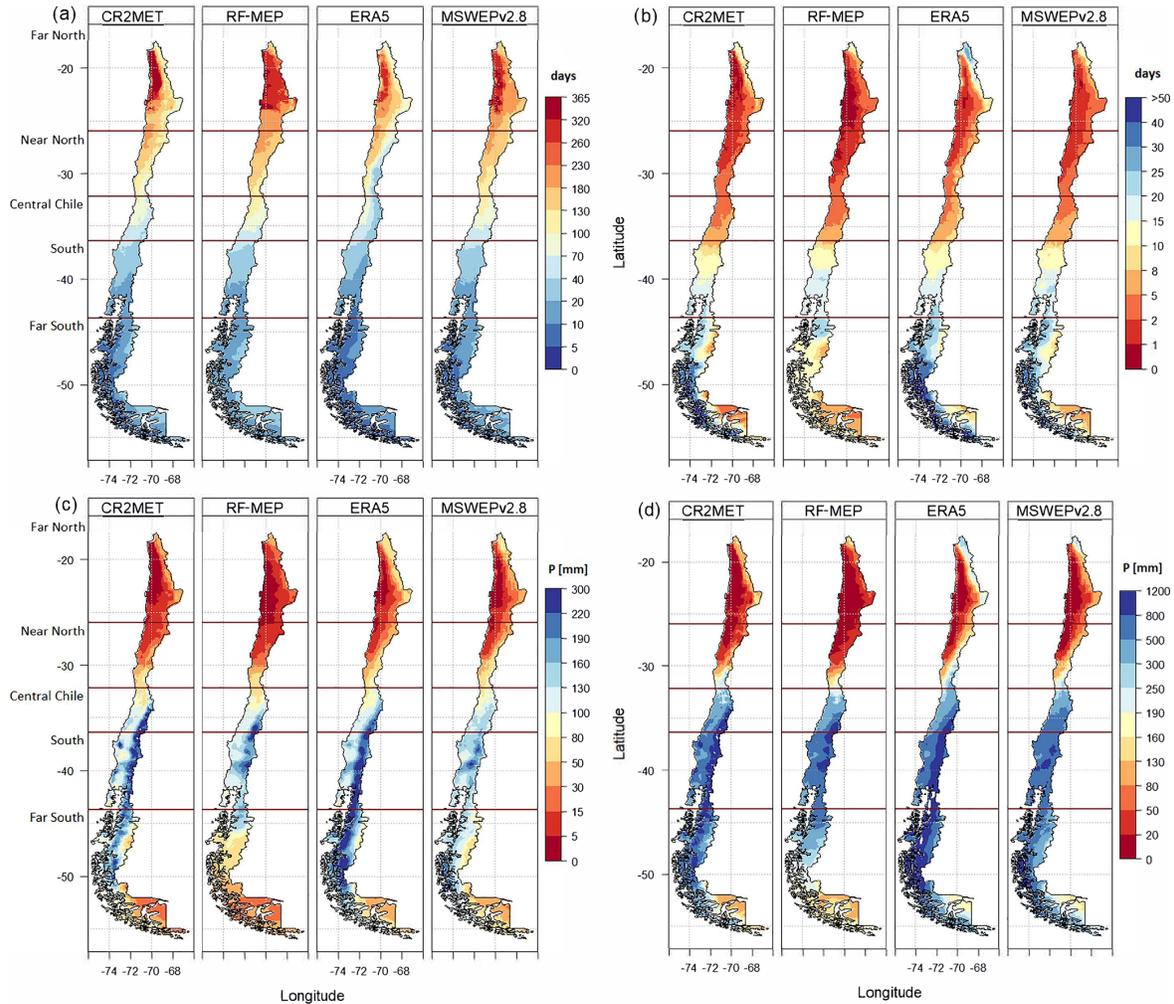


Figure 3. Median annual values of four Climdex indices over 1990–2018 (full period): (a) number of consecutive dry days (CDD), (b) number of consecutive wet days (CWD), (c) maximum *P* over five consecutive days (Rx5day), and (d) annual *P* that is above the 95th percentile of *P* for wet days (R95pTOT). The dark red horizontal lines represent the limits of each macroclimatic zone.

squared-error indices (Gupta et al., 2009; Kling et al., 2012; Mizukami et al., 2019) and has been used in numerous studies (e.g. Garcia et al., 2017; Beck et al., 2019; Baez-Villanueva et al., 2020; Neri et al., 2020; Széles et al., 2020). The KGE' has three components: the Pearson correlation coefficient (r ; Eq. 2), the bias ratio (β ; Eq. 3), and the variability ratio (γ ; Eq. 4). μ is the mean Q , CV is the coefficient of variation, σ represents the standard deviation of Q , and the subscripts “s” and “o” represent simulated and observed Q , respectively. The KGE' and its components have their optimum value at 1, and its optimisation seeks to reproduce the temporal dynamics (measured by r) while preserving the volume and variability of Q , measured by β and γ , respectively

(Kling et al., 2012).

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

$$r = \frac{\sum_{i=1}^n (O_i - \bar{O})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}} \quad (2)$$

$$\beta = \frac{\mu_s}{\mu_o} \quad (3)$$

$$\gamma = \frac{CV_s}{CV_o} = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} \quad (4)$$

To calibrate the model parameters, we used the hydroPSO global optimisation algorithm (Zambrano-Bigiarini and Rojas, 2013), which implements a state-of-the-art version of

the particle swarm optimisation technique (PSO; Eberhart and Kennedy, 1995; Kennedy and Eberhart, 1995). We used the standard PSO 2011 algorithm (Clerc, 2011a, b), defined as *sps02011* in the hydroPSO R package (Zambrano-Bigiarini and Rojas, 2013). We set the number of particles in the swarm ($npart = 80$), the maximum number of iterations ($maxit = 100$), and the relative convergence tolerance ($reltol = 1 \times 10^{-10}$), while the default values were used for all other parameters. Over the last decade, hydroPSO has been successfully used to calibrate numerous hydrological and environmental models (e.g. Brauer et al., 2014; Silal et al., 2015; Bisselink et al., 2016; Kundu et al., 2017; Kearney and Maino, 2018; Abdelaziz et al., 2019; Ollivier et al., 2020; Hann et al., 2021). For more details on the use of the hydroPSO package to calibrate the TUWmodel, readers are referred to Zambrano-Bigiarini and Baez-Villanueva (2020).

3.4 Regionalisation techniques

After obtaining catchment-specific model parameters through independent catchment calibration (Sect. 3.3), we compared three parameter regionalisation techniques: (i) spatial proximity, (ii) feature similarity, and (iii) parameter regression. We assessed performance through a leave-one-out cross-validation exercise, which consists of leaving out each one of the 100 catchments, transferring model parameters, conducting Q simulations, and computing performance evaluation metrics.

3.4.1 Spatial proximity

The spatial proximity method assumes that climatic and physical characteristics are relatively homogeneous over a region (Oudin et al., 2008). We quantified the spatial proximity between the target pseudo-ungauged and the remaining catchments using the Euclidean distance between catchment centroids, computed with geographic coordinates (i.e. latitude and longitude):

$$ED_{ij} = \sqrt{\sum_{k=1}^n (x_{k,i} - x_{k,j})^2}. \quad (5)$$

For each pseudo-ungauged catchment, the donor was chosen according to the minimum Euclidean distance, and the full parameter set obtained during the independent calibration of the donor catchment was transferred to the pseudo-ungauged catchment.

3.4.2 Feature similarity

In the feature similarity method, we transferred the calibrated parameter sets from 10 donor catchments to the pseudo-ungauged catchment based on similarity between climatic and geomorphological features, quantified using the catchment characteristics presented in Table 4. To exclude re-

dundant information, we first performed correlation analyses between catchment descriptors using the Pearson and Spearman rank correlation coefficients (to account for linear and monotonic correlation, respectively) and discarded three descriptors with high correlations (mean elevation, mean annual PE , and the Simple Precipitation Intensity Index (SDII); see Appendix B). Also, we discarded snow cover because it was found to be unreliable, leaving nine catchment features for this method. To assign equal weight to each catchment characteristic, they were normalised into the range $[0, 1]$ using Eq. (6):

$$Z_f = \frac{x_f - x_{\min}}{x_{\max} - x_{\min}}, \quad (6)$$

where x_f is the value of the characteristic for catchment f , while x_{\max} and x_{\min} are the maximum and minimum values of the characteristic x over all catchments. After normalising all catchment characteristics, we calculated the dissimilarity as follows:

$$S_{i,j} = \sum_{m=1}^n |Z_{i,m} - Z_{j,m}|, \quad (7)$$

where $S_{i,j}$ is the dissimilarity index between catchments i and j ; $Z_{i,m}$ and $Z_{j,m}$ are the normalised values of the m catchment characteristic for catchments i and j , respectively; and n is the total number of characteristics.

For each pseudo-ungauged catchment i , the 10 catchments j with the lowest dissimilarity indices ($S_{i,j}$) were selected as donors (Oudin et al., 2008; Zhang and Chiew, 2009; Zhang et al., 2015; Beck et al., 2016). The full parameter sets obtained during the independent calibrations of each donor catchment were used to run TUWmodel in the pseudo-ungauged catchment, thus producing an ensemble of 10 Q simulations, as in previous studies (McIntyre et al., 2005; Zelelew and Alfredsen, 2014; Beck et al., 2016). The 10 Q time series were then averaged to produce a single Q time series.

3.4.3 Parameter regression

The parameter regression technique aims to detect statistical relationships between parameter values and catchment characteristics and uses these relationships to estimate model parameters for ungauged catchments (Parajka et al., 2005; Oudin et al., 2008; Swain and Patra, 2017). To account for non-linear relationships between model parameters and catchment characteristics, we implemented the random forest machine learning algorithm (RF; Breiman, 2001; Prasad et al., 2006; Biau and Scornet, 2016) provided in the RandomForest R package (Liaw and Wiener, 2002). RF uses an ensemble of decision trees between predictand and predictor values (also known as covariates) for regression and supervised classification and has the capability to deal with high-dimensional feature spaces and small sample sizes (Biau and

Table 4. Selected climatic and physiographic characteristics to quantify feature similarity between catchments. All variables related to P were computed using the corresponding P product used as an input to the TUWmodel for 1990–2018.

No.	Variable	Data source	Importance
1	Mean elevation	CAMELS-CL	Composite indicator that influences a range of processes such as long-term P and T and hence soil moisture availability. In some environments, it is also related to aridity and snow processes.
2	Median elevation	SRTMv4.1	Same as mean elevation but provides a more robust representation of elevation over mountainous catchments.
3	Catchment area	CAMELS-CL	Related to the degree of aggregation of catchment processes related to scale effects. Additionally, it is an indicator of total catchment storage capacity.
4	Slope	CAMELS-CL	Related to the response of the catchment, routing, and infiltration processes.
5	Forest cover	CAMELS-CL	Forested catchments are associated with a trade-off between high water consumption rates and enhanced soil.
6	Snow cover	CAMELS-CL	Related to the influence of snow processes within the catchment.
7	Mean annual precipitation	P product	Related to the generation of runoff and P related to orographic gradients (e.g. coastal areas).
8	Mean annual air temperature	CR2MET	Indicator of snow processes in cold environments. It is also related to aridity and consequently to the evaporative demand.
9	Mean annual potential evaporation	Computed from CR2MET	A measure of the atmospheric water demand (especially at the annual temporal scale).
10	Aridity index	CR2MET and P product	Represents the competition between energy and water availability.
11	Daily temperature range	CR2MET	Monthly mean difference between daily maximum and minimum T . Related to variations in the diurnal cycle and evaporative demands.
12	Simple precipitation intensity index	P product	Relation of annual P to the number of wet days ($P > 1$ mm). Serves as a proxy for seasonality and intensity of P events.
13	Maximum consecutive 5 d precipitation	P product	Related to extreme P events.

Scornet, 2016). Previous studies have shown that RF can deal with several covariates as well as non-informative predictors because it does not lead to overfitting or biased estimates (Díaz-Urriarte and Alvarez de Andrés, 2006; Biau and Scornet, 2016; Hengl et al., 2018), which is why it has been used for numerous hydrological applications (Saadi et al., 2019; Baez-Villanueva et al., 2020; Beck et al., 2020b; Zhang et al., 2021). For a more detailed description of RF, we refer the reader to Prasad et al. (2006), Biau and Scornet (2016), and Addor et al. (2018).

For this study, we developed one RF model for each TUW-model parameter, using all 13 independent catchment characteristics listed in Table 4 as covariates. Our experimental setup used an ensemble of 2000 regression trees, a minimum of five terminal nodes for each model, and $p/3$ variables randomly sampled as candidates at each split, where p represents the number of predictors. The trained RF models were then used to predict parameter values in the pseudo-ungauged catchments.

3.5 Influence of nested catchments

To evaluate the influence of nested catchments on the performance of the three regionalisation methods, we repeated the three regionalisation methods for each target catchment,

with catchments considered to be nested (in relation to the pseudo-ungauged catchment) excluded from the set of potential donor catchments. Following Neri et al. (2020), we used a cut-off point of 10 % of drainage area, meaning that only catchments that cover more than 10 % of the area of the parent catchment were considered to be nested.

3.6 Influence of donor catchments for feature similarity

To evaluate the influence of the number of donors used in feature similarity, we repeated the process followed in Sect. 3.4.2 to assess the performance of this regionalisation method when 1, 2, 4, 6, 8, and 10 donor catchments are selected. This analysis evaluates the impact of averaging varying numbers of simulations compared to the results that are based on only the most similar catchment.

We performed all analyses using the R Project of Statistical Computing (R Core Team, 2020). In addition to the R packages described in the methodology, we used the hydroGOF (Zambrano-Bigiarini, 2020a), hydroTSM (Zambrano-Bigiarini, 2020b), lfst (Koffler et al., 2016), raster (Hijmans, 2020), rasterVis (Perpiñán and Hijmans, 2020), rgdal (Bivand et al., 2020), and rgeos (Bivand and Rundel, 2020) packages.

4 Results

4.1 Performance of P products

4.1.1 Calibration and verification

Figure 4 shows the performance of the TUWmodel during calibration (2000–2014) and the two verification periods (1990–1999 and 2015–2018), prior to any regionalisation procedure. CR2MET provided the best performance for all evaluated periods, with median KGE's of 0.84, 0.76, and 0.66, for calibration, Verification 1 (1990–1999, near-normal/wet), and Verification 2 (2015–2018, dry), respectively, followed closely by RF-MEP. Surprisingly, MSWEPv2.8 provided the poorest performance for calibration and Verification 1. For all P products, the lowest performances were obtained during the (dry) Verification 2 period, emphasising the challenges of estimating Q in dry conditions, as discussed by Maggioni et al. (2013) and Beck et al. (2016). Despite the substantial variations between P products (see Sect. 3.1.1), the TUWmodel performed well for all P products in the calibration, Verification 1, and Verification 2 periods, with median KGE' values greater than 0.77, 0.71, and 0.62, respectively. The calibrated model parameters lay well within the selected parameter ranges in the large majority of the cases (see Fig. S8 of the Supplement). In other words, the selected parameter ranges were wide enough so that calibrated parameter values were not concentrated at their lower or upper limits.

Figure 5 shows the performance of the TUWmodel during calibration, Verification 1, and Verification 2 per hydrological regime (see Fig. 1d). The TUWmodel performed better over the pluvio-nival catchments, with median KGE' values above 0.77, 0.76, and 0.69 for calibration, Verification 1, and Verification 2, respectively. During the calibration period, there was no clear second best regime. For instance, the snow-dominated catchments presented slightly higher median KGE' values but a more pronounced dispersion, while the pluvio-nival and rain-dominated catchments presented lower dispersion but reduced median values. The snow-dominated catchments presented a more pronounced decrease from calibration (median KGE' > 0.85) to both verification periods (> 0.55 and 0.23 for Verification 1 and Verification 2, respectively). During both verification periods, the rain-dominated catchments presented the highest dispersion increases in both verification periods compared to calibration.

Over the snow-dominated catchments, ERA5 performed the worst as it presented the highest dispersion and the lowest median KGE' values during Verification 1 (0.55) and Verification 2 (0.25), despite having the highest median KGE' during calibration (0.87). RF-MEP performed the best during Verification 1 (0.68), while MSWEPv2.8 performed the best during the dry Verification 2 period (median KGE' of 0.60). CR2MET performed the best over the nivo-pluvial

catchments, with median KGE' values above 0.64, while RF-MEP performed relatively worse for both verification periods, with median KGE' values above 0.48 and a larger dispersion than the other products, despite having a similar median KGE' (0.62) in Verification 1 to ERA5 and MSWEPv2.8 (0.61, and 0.60, respectively). Over the pluvio-nival catchments, all products showed a relatively good performance, with CR2MET being the best P product in calibration and Verification 1 (median KGE's of 0.87 and 0.84, respectively), while ERA5 performed the best during Verification 2 (median KGE' of 0.78). RF-MEP performed the best over the rain-dominated catchments in calibration and Verification 1, with median KGE' values of 0.84 and 0.77, respectively, while ERA5 performed the worst (median KGE' values of 0.69 and 0.70). Finally, CR2MET performed the best in Verification 2 (median KGE' of 0.72), followed by MSWEPv2.8 (median KGE' of 0.69).

4.1.2 Performance during regionalisation

Figure 6 summarises the leave-one-out cross-validation results obtained from the application of three regionalisation methods, for each P product. The results are displayed for the calibration (2000–2014; panel a), Verification 1 (1990–1999; panel b), and Verification 2 (2015–2018; panel c) periods. Overall, the median performance of all P products was the best for feature similarity, with median KGE' values between 0.44–0.62 for all periods, followed by spatial proximity (0.39–0.55) and parameter regression (−0.12–0.51). In addition to exhibiting a considerably lower overall performance, parameter regression returned a larger spread in KGE's for all periods.

The overall performances obtained for feature similarity and spatial proximity are relatively close for different P products over each period (Fig. 6). For feature similarity, all P products generate acceptable KGE' results (median KGE' > 0.54) during the calibration and Verification 1 periods, while the median KGE' values during the dry Verification 2 period lowered to a median KGE' of > 0.44. The best model performance for feature similarity was obtained by CR2MET, with median KGE' values of 0.62 for calibration and Verification 1 and 0.53 for Verification 2, followed closely by RF-MEP for calibration (0.59), ERA5 for Verification 1 (0.59), and MSWEPv2.8 for Verification 2 (0.52). In the case of spatial proximity, MSWEPv2.8 yielded the best performance in the calibration period (0.55), followed closely by RF-MEP (0.56 but with a higher dispersion), and CR2MET (0.53). For Verification 1, RF-MEP provided the best performance (0.54), while MSWEPv2.8 produced the best results over Verification 2 (0.48). For spatial proximity, ERA5 performed the worst over the three evaluated periods. Finally, parameter regression yielded the lowest results, with CR2MET and ERA5 showing the highest median KGE' values (> 0.42 for calibration and Verification 1 and > 0.22 for Verification 2).

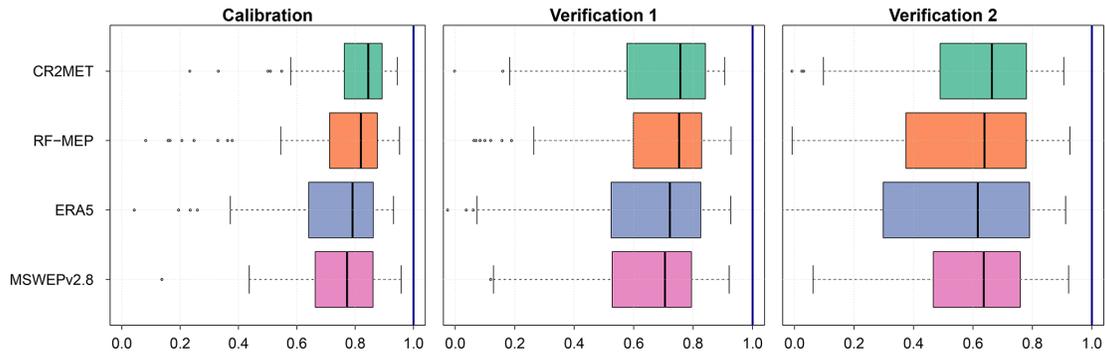


Figure 4. Performance of the TUWmodel during the calibration (2000–2014), Verification 1 (1990–1999), and Verification 2 (2015–2018), prior to any regionalisation, using the modified Kling–Gupta efficiency (KGE'). The solid line represents the median value, the edges of the boxes represent the first and third quartiles, and the whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box. The blue line indicates the optimal value for the KGE' .

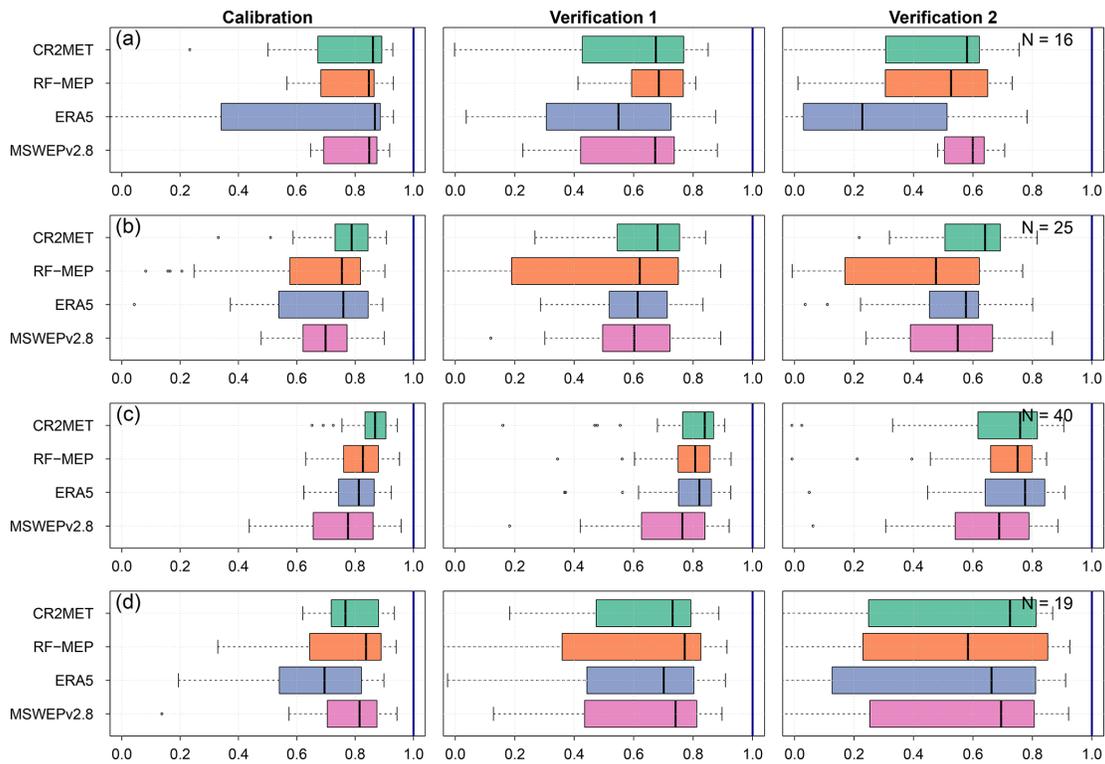


Figure 5. Performance of TUWmodel during calibration (2000–2014), Verification 1 (1990–1999), and Verification 2 (2015–2018), prior to any regionalisation, over catchments with different hydrological regimes: (a) snow-dominated, (b) nivo-pluvial, (c) pluvio-nival, and (d) rain-dominated.

For each regionalisation technique, Fig. 7 summarises the spatial distribution of the performance of each *P* product for the calibration, Verification 1, and Verification 2 periods. The spatial patterns obtained for all regionalisation methods were similar, independent of the *P* product or the eval-

uated period, except for parameter regression, which yielded poor results over high-elevation catchments and under dry conditions (Verification 2). These results indicate that spatial proximity and feature similarity present very similar spatial

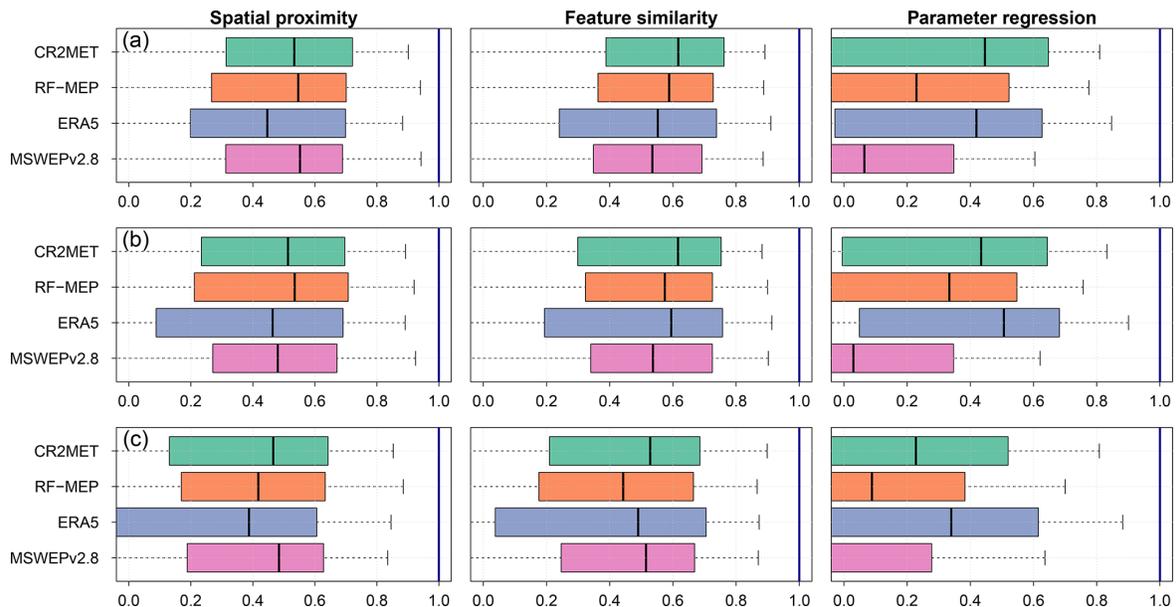


Figure 6. Leave-one-out cross-validation results for the three regionalisation methods applied with different P products during the (a) calibration (2000–2014), (b) Verification 1 (1990–1999), and (c) Verification 2 (2015–2018) periods.

performance patterns, with feature similarity yielding higher KGE' values over the three evaluated periods.

All P products performed better in the Central Chile and South regions than in the Far North, Near North, and Far South regions. The low performance of regionalisation in the arid north is very likely due to the convective nature of storms occurring in the highlands of the Chilean Altiplano (elevations above 4000 m a.s.l.) and the low density of Q stations over this area. Despite this general low performance, RF-MEP was the best-performing P product over the Far North region for both spatial proximity (median KGE' of 0.28) and feature similarity (median KGE' of 0.46) in the calibration period, suggesting that merging P products and ground-based observations helps to improve, to some extent, the performance of hydrological modelling across arid regions. Conversely, all products outperformed RF-MEP over the Far South. Figure 7 also highlights that spatial proximity provides the best performance over the Far South, with median KGE' values higher than 0.46, 0.27, 0.30, and 0.35 for CR2MET, RF-MEP, ERA5, and MSWEPv2.8, respectively. The systematic lower performance of feature similarity compared to spatial proximity over the Far South (except for the case of ERA5) could be attributed to (i) the lack of catchment characteristics that represent the hydrological behaviour of this complex area dominated by polar and temperate climates and (ii) the low number of potential donor catchments (11 for latitudes $> 49^\circ$ S), combined with their varied hydrological regimes. For the most southern catchments, the highest P intensities occur during March–May, while the lowest

P occurs between June–August, which differs from catchments throughout the rest of the country (Alvarez-Garreton et al., 2018, their Fig. 9). This may affect the hydrological simulations when model parameters from catchments located $< 49^\circ$ S are transferred to these far southern catchments.

4.2 Evaluation of regionalisation techniques

4.2.1 Overall performance

For each P product, Fig. 8 compares the performances of the three regionalisation techniques with those obtained in the independent calibration and verification periods. The independent calibration of each catchment represents the highest model performance that can be obtained for a specific combination of hydrological model, objective function, and catchment (i.e. an absolute benchmark), whereas the two verification periods were used to evaluate the performance of the regionalisation techniques over independent time periods (i.e. as verification benchmarks). There are marked differences in performance according to the P product used to force the TUWmodel, regardless of the regionalisation method and the evaluated period. For example, ERA5 has more dispersion in the KGE' values compared to other products for the cases of feature similarity and spatial proximity, while for parameter regression, it tends to perform the best. For all P products and evaluation periods, feature similarity performed the best, followed by spatial proximity and parameter regression, which is consistent with results from multiple studies (e.g. Parajka et al., 2005; Oudin et al., 2008; Bao

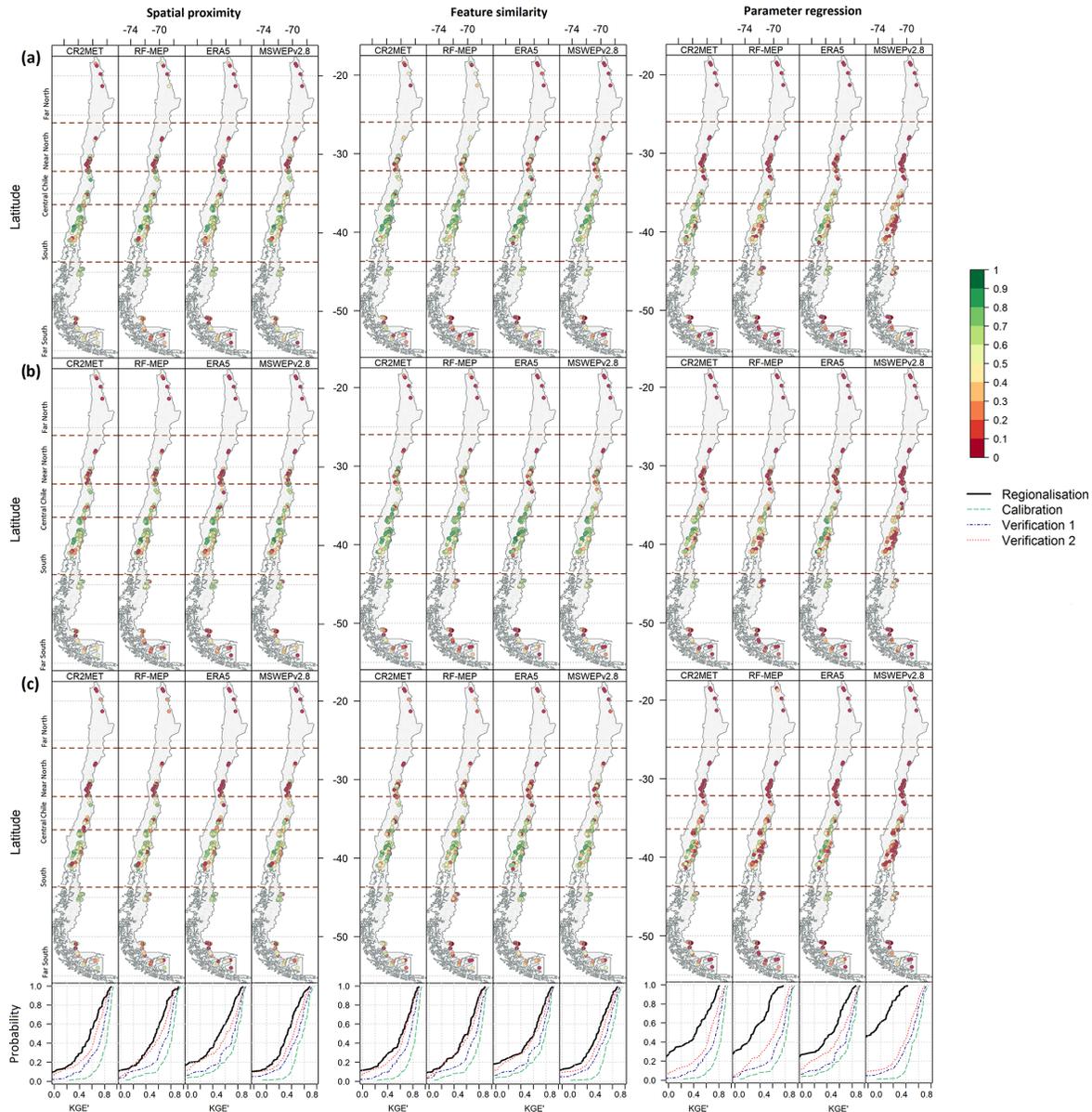


Figure 7. Spatial performance of the leave-one-out cross-validation results for the three regionalisation methods according to *P* products used to force TUWmodel. Results are presented for the (a) calibration (2000–2014), (b) Verification 1 (1990–1999), and (c) Verification 2 (2015–2018) periods. The panels beneath the map plots refer to the ECDFs of the corresponding regionalisation technique for the entire period of analysis (1990–2018) and *P* products (black) against the performances during the independent calibration (green), Verification 1 (blue), and Verification 2 (red) periods.

et al., 2012; Garambois et al., 2015; Neri et al., 2020). Parameter regression had both the lowest median KGE's as well as the largest spread. Comparing the two verification periods, results obtained during the (near-normal/wet) Verification 1 period were close to those obtained during calibration, while those obtained during the (dry) Verification 2 period were

substantially lower, especially for spatial proximity and parameter regression.

These results are in agreement with the lower panels located below each map in Fig. 7, which show the empirical cumulative distribution functions (ECDFs) of the performance of each regionalisation technique during the complete

period of analysis (1990–2018). These ECDFs compare the relative performance of each regionalisation method against those obtained from the independent calibration and verification of each catchment (used as benchmarks). As expected, all regionalisation methods presented a lower performance than the independent calibration and verification, with this reduction more pronounced for parameter regression.

4.2.2 Impact of hydrological regimes

Figure 9 shows the performance of the regionalisation techniques according to hydrological regime for all P products during the calibration period (and Figs. S9 and S10 of the Supplement show the same for the two verification periods). Feature similarity provided the best median performance for all hydrological regimes and P products except for snow-dominated catchments, where spatial proximity performed the best for MSWEPv2.8 for calibration and Verification 2. These results demonstrate that there was no single P product that outperformed the others for all regionalisation techniques and hydrological regimes. In other words, the best-performing P product depends on the hydrological regime and chosen regionalisation method for our case study. For feature similarity in snow-dominated catchments, RF-MEP performed the best for calibration and Verification 1, while CR2MET performed the best during Verification 2. For nivo-pluvial catchments, CR2MET provided the best performance during calibration and Verification 1, while MSWEPv2.8 performed the best during Verification 2. CR2MET and ERA5 performed the best in pluvio-nival catchments for the case of feature similarity, while all products performed similarly for spatial proximity. Finally, ERA5 performed the best for feature similarity in all periods across the rain-dominated catchments.

4.3 Impact of nested catchments

We evaluated the influence of the nested catchments on the regionalisation results. Figure 10 shows the performance of the three regionalisation methods for the subset of 56 nested catchments that share a common area with at least one other catchment (i.e. the 42 nested catchments as well as all corresponding parent catchments). Here, we compare the regionalisation performance using all potential donors (dark colours) with the performance when excluding nested catchments as potential donors (light colours). The order of performance of the regionalisation methods and P products did not vary when the nested catchments were excluded, as feature similarity and CR2MET remained the best-performing method and product, respectively. As expected, the regionalisation technique with the largest reduction in performance when excluding nested catchments was spatial proximity, followed closely by feature similarity. All P products showed a slight performance reduction and increased dispersion for spatial proximity, except for MSWEPv2.8, which

showed a slight increase in the KGE' median value. Feature similarity showed a slight reduction in performance when the nested catchments were excluded; however, the median values remained almost the same. The change in performance of parameter regression was negligible after the exclusion of nested catchments because, in the particular case of Chile, excluding only a few catchments had a negligible effect on the non-linear relationships between model parameters and the selected climatic and physiographic characteristics (see Table 4).

4.4 Impact of the number of donors in feature similarity

Figure 11 shows the performance of feature similarity during the calibration and both verification periods when varying the number of donors used to transfer model parameters to ungauged catchments (see Sect. 3.6). In general, the highest median performance is obtained when using four or more donor catchments. However, the application of a t test demonstrated that the improvement in the KGE' values obtained when increasing to more than one donor was not statistically significant. The results show that the performance varies according to the P product and selected period of analysis. For the calibration period, feature similarity produced similar median values to those obtained with spatial proximity when one donor was used, while the performance improved as more donors were included. For both verification periods, feature similarity (median KGE' values from 0.44 to 0.64) outperformed spatial proximity (median KGE' values ranging from 0.39 to 0.54). For all three periods, feature similarity provided better performance considering the distribution of the KGE' values.

5 Discussion

5.1 Performance of P products

During the independent catchment calibration (2000–2014) and two verification periods (1990–1999 and 2015–2018), good performances were obtained with all P products (see Fig. 4). When decomposing the results of the KGE' objective function into its three components (see Appendix C), r exhibited the lowest performance, while β and γ values were generally closer to their optimal values, particularly for calibration and Verification 1. The results obtained with ERA5, which is a reanalysis product, were as good or even better than those obtained with the gauge-corrected products CR2MET, RF-MEP, and MSWEPv2.8 (e.g. see results for the pluvio-nival catchments in Fig. 5). This is in agreement with Tarek et al. (2020), who concluded that ERA5 should be considered a high-potential dataset for hydrological modelling in data-scarce regions. The good performance of ERA5 suggests that, for the particular case of Chile, merging P products with ground-based measurements does not

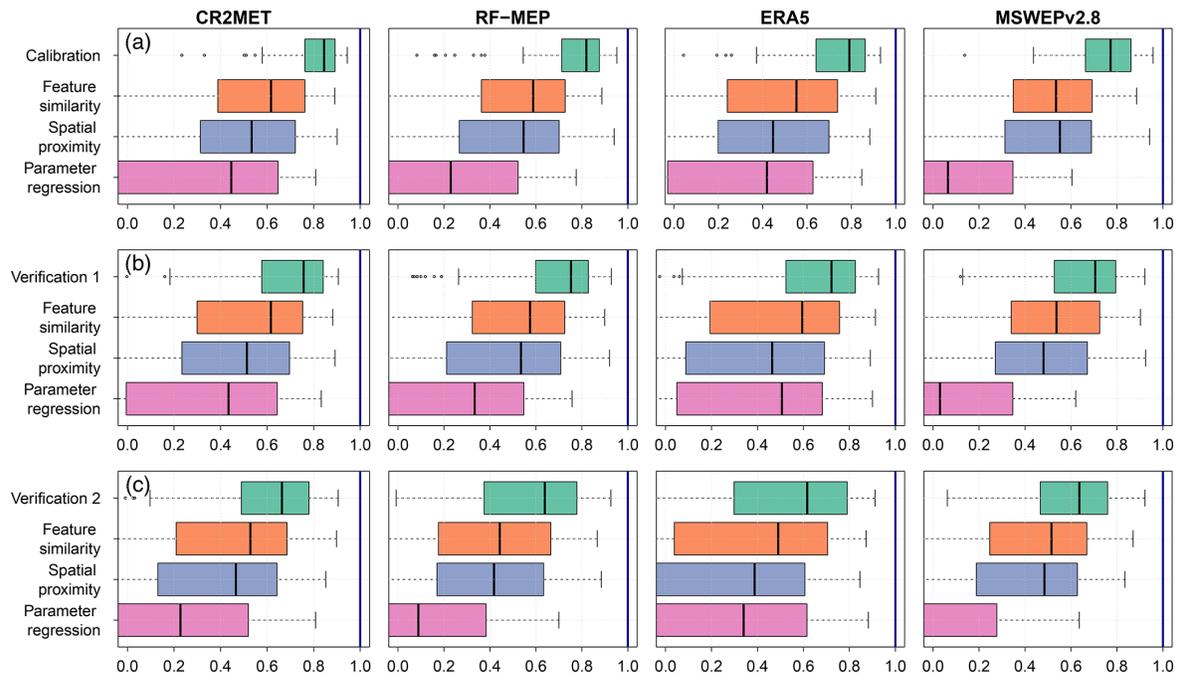


Figure 8. Performance of the regionalisation methods during the (a) calibration (2000–2014), (b) Verification 1 (1990–1999), and (c) Verification 2 (2015–2018) periods.

necessarily translate into improved hydrological model performance, which may be attributed to (i) the lack of P rain gauges in the Andes Mountains; (ii) the ability of the rainfall-runoff model to compensate for the P forcing (visible in the performances of the β and γ components, Appendix C), and (iii) the fact that P products still have errors in the detection of P events that could impact the representation of the modelled Q dynamics (as suggested by the relative lower performance of the r component of the KGE').

Furthermore, the similar performances obtained with uncorrected (ERA5) and gauge-corrected (CR2MET, RF-MEP, and MSWEPv2.8) P products, both in wet and dry periods, highlight that there was no single P dataset outperforming the others in all periods. These results demonstrate that the calibration of hydrological model parameters smooths out, to some extent, the spatio-temporal differences between P products (see Figs. 2, 3, 6 and 9), which is in agreement with previous studies that have demonstrated that model calibration with each P product improves the performance of Q simulations (e.g. Artan et al., 2007; Stisen and Sandholt, 2010; Bitew et al., 2012; Thiemig et al., 2013). The decomposition of the KGE' into its components also demonstrated the ability of the TUWmodel to compensate for the total volume of P , as the β component was close to the optimum value, particularly for calibration and Verification 1 (see Appendix C), which can be attributed to the improved detection of P events of the merged products (regarding RF-MEP,

see Baez-Villanueva et al., 2020). This can also be observed for MSWEPv2.8, as it produced the best performance over snow-dominated catchments under dry conditions (Verification 2).

Regarding the suitability of P products for parameter regionalisation, RF-MEP provided slightly better results in the Far North for the calibration period using both spatial proximity and feature similarity, suggesting that P products that are merged with ground-based information over arid climates can improve regionalisation performance. The lower performance obtained in regionalisation with ERA5 in the Far North compared to the other P products (median values < 0.18 for feature similarity in all periods) can be attributed to its high P values, which are likely due to the lack of ground-based P stations over Chile in the development of the product. The incorporation of ground-based stations has the potential to (i) compensate for overestimations caused by the evaporation of hydrometeors before they reach the ground (Maggioni and Massari, 2018) and (ii) improve event-based detection skills (Baez-Villanueva et al., 2020; Zhang et al., 2021). The latter is evident in CR2MET and MSWEPv2.8, which are both based on ERA5 but included several rain gauges in the Far North and have a higher performance than ERA5 (see Figs. 2, 3, and S1).

Despite the low performance of all P products in the Far North and Near North (median KGE' values < 0.58 ; see Fig. 7), the TUWmodel appears to be flexible enough to com-

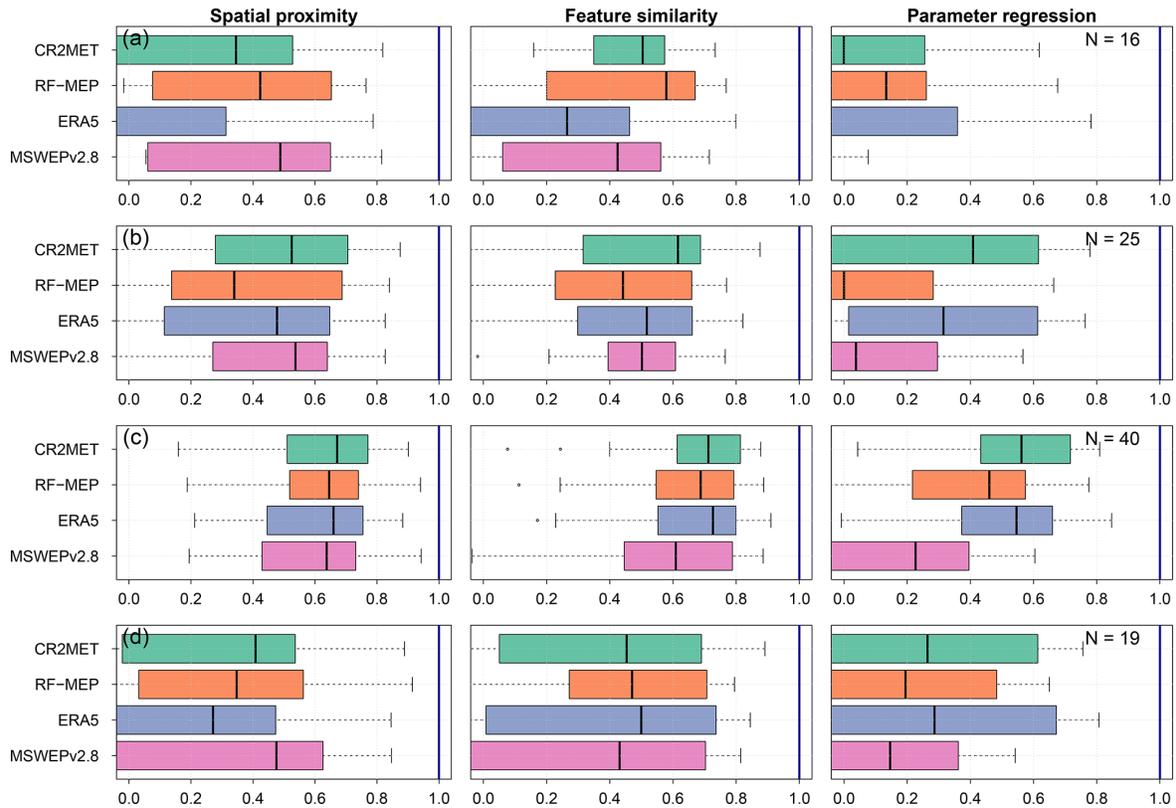


Figure 9. Performance of regionalisation methods for calibration (2000–2014) according to the hydrological regime: (a) snow-dominated, (b) nivo-pluvial, (c) pluvio-nival, and (d) rain-dominated. N denotes the number of catchments per hydrological regime.

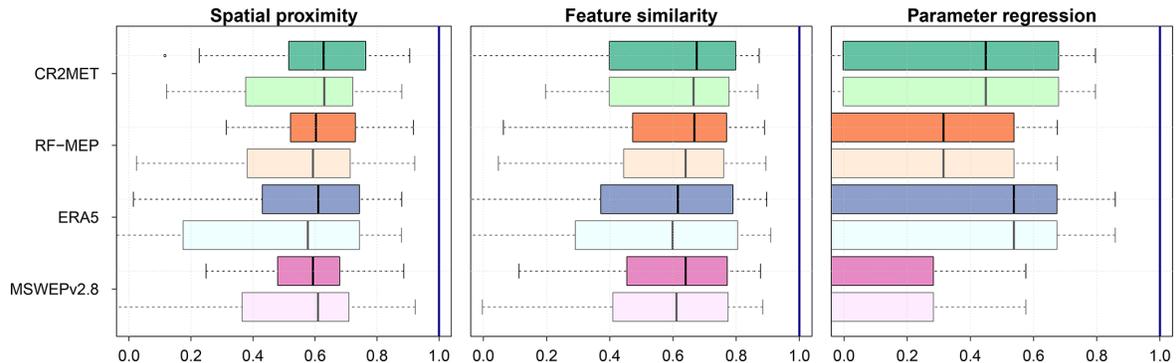


Figure 10. Comparison of regionalisation performance using all catchments as potential donors (dark colours) against the performance when nested catchments are excluded as potential donors (light colours).

pensate, to some extent, for differences between P products. A similar conclusion was obtained by Elsner et al. (2014), who examined differences between four meteorological forcing datasets and their implications in hydrological model calibration in the western United States using the variable infiltration capacity model (VIC; Liang et al., 1994). Our results

are also in agreement with Bisselink et al. (2016), who concluded that parameter sets obtained during calibration partially compensated for the bias of seven P products used to force the fully distributed LISFLOOD model in four catchments in southern Africa.

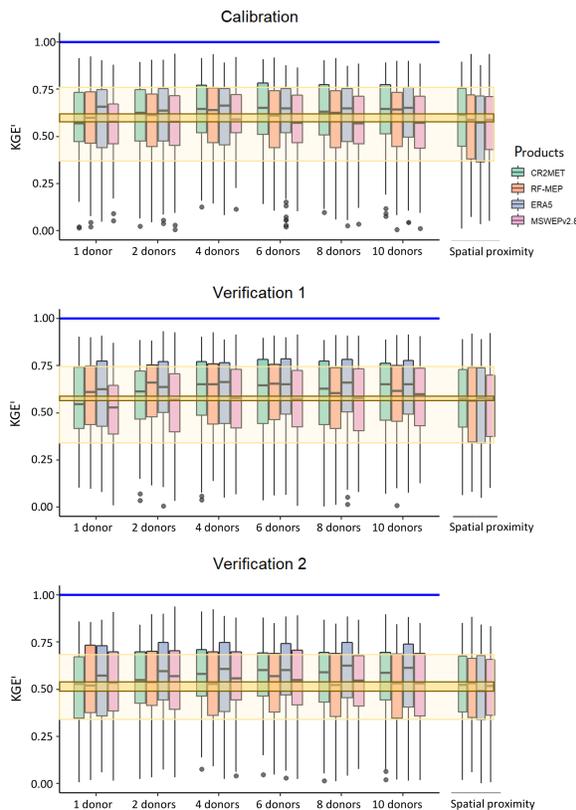


Figure 11. Influence of the number of donors used for feature similarity for calibration (2000–2014), Verification 1 (1990–1999), and Verification 2 (2015–2018). The results from spatial proximity are included on the right of each panel for comparison purposes. The dark yellow box denotes the upper and lower bounds of the median performance (of the four P products) obtained with spatial proximity, the lighter yellow box represents the upper and lower bounds of the interquartile range for spatial proximity, and the blue lines represent the optimum KGE' value.

An unexpected result from this study is that the spatial resolution of the P products did not play a major role in model performance during calibration, verification and regionalisation; although CR2MET and RF-MEP have a higher spatial resolution (0.05° ; $\sim 25 \text{ km}^2$) than MSWEPv2.8 ($\sim 0.10^\circ$; $\sim 100 \text{ km}^2$) and ERA5 ($\sim 0.28^\circ$; $\sim 625 \text{ km}^2$), all four products performed well during the independent calibration of the hydrological model and the two verification periods. The performance of ERA5 over the 25 smallest catchments during regionalisation (area $< 353.1 \text{ km}^2$) was similar to that obtained with products with a higher spatial resolution (Fig. S11 of the Supplement). This can be attributed to the fact that Chile is dominated by large-scale frontal systems (Zhang and Wang, 2021); and therefore, coarse-resolution products may perform well over small catchments. Our results also align with the findings of Maggioni et al. (2013),

who concluded that the loss of spatial information associated with coarser resolution (e.g. ERA5) can be compensated for through model calibration.

5.2 How does the calibration of the TUWmodel compensate for differences in P ?

The calibration of TUWmodel was able to compensate, to some extent, for differences in annual and intra-annual P amounts, intermittency, and extremes (see Figs. 2 and 3) among the four products. Using the example of the nivo-pluvial catchments, Fig. 12 illustrates how TUWmodel parameters compensate for differences between the P forcings used in calibration, while Fig. 13 shows the corresponding variations in the mean monthly water balance components. Similar figures for snow-dominated, pluvio-nival, and rain-dominated catchments can be found in the Supplement (Figs. S12–S17).

In general, the calibrated parameters behave as expected for each hydrological regime. A notable exception is ERA5, which shows low values for the snow correction factor (SCF) in nivo-pluvial and snow-dominated catchments (Figs. 12 and S12). These catchments are primarily located in the arid Near North region (see Fig. 2 and Figure S15), where the estimated winter P is substantially lower for CR2MET, RF-MEP, and MSWEPv2.8, and a high SCF corrects this apparent underestimation. The lower P amounts presented in these products may reflect the incorporation of information from rain gauges located in drier, low-lying areas to correct their P estimates (see Fig. S1).

ERA5 presented relatively low SCF values over nivo-pluvial catchments compared to the other P products (Fig. 13), which is expected because it exhibits the highest P values. Conversely, because RF-MEP has the lowest mean monthly P over the nivo-pluvial catchments, the model adjusts the evaporation, snow water equivalent, and soil moisture components (Fig. 13), thus increasing the simulated Q (to match the observed Q). Substantial differences were obtained for LPrat and field capacity (FC), which directly affect evaporation and soil moisture. For example, over the nivo-pluvial catchments, the LPrat and FC values for RF-MEP are similar to those of ERA5, despite RF-MEP having substantially lower P amounts, which in turn is reflected in the reduced soil moisture and evaporation amounts. The differences between LPrat and FC according to P product are even more pronounced for snow-dominated catchments (Fig. S12).

Finally, higher values of the nonlinear parameter for runoff production Beta reduce the amount of water that leaves the catchment as runoff (Széles et al., 2020, their Eq. 7). For all hydrological regimes except pluvio-nival, the median Beta parameter is substantially higher for ERA5 than for the other P products. The larger Beta values obtained with ERA5 are expected to attenuate the runoff generation from extreme P events (see Fig. 3c and d). Interestingly, the Beta

parameter is zero in some pluvio-nival catchments, which means that all liquid P and snowmelt was used to generate runoff (Fig. S16). This behaviour was more pronounced with RF-MEP and MSWEPv2.8, which exhibited the lowest P amounts and longer dry spells (Fig. 3a) over these catchments. In general, the storage components obtained from each P product (computed as the sum of the two deepest reservoirs of the model; see Széles et al., 2020, their Fig. 3) are similar for all four P products.

5.3 Evaluation of regionalisation techniques

The compensation due to the flexibility of the TUWmodel observed during the independent calibration and verification (see Sect. 5.2) also influences the regionalisation performance. Feature similarity provided the best performance when the TUWmodel was forced with all P products (Fig. 8), while spatial proximity provided similar performance to feature similarity over the Central Chile and South regions, where there is a high density of Q stations. These results are in agreement with Parajka et al. (2005), Oudin et al. (2008), and Neri et al. (2020), who demonstrated that spatial proximity performs well over densely gauged regions.

The inclusion of donor catchments with low model performance introduces a diversity that has the potential to benefit Q prediction in ungauged catchments, as discussed by Oudin et al. (2008). We decided to incorporate these catchments in the regionalisation process because of the diversity of climates and physiographic characteristics across continental Chile (see Fig. 1), with the potential downside that this may lead to errors in the transferred model parameters. Additionally, the similarity between the performance of spatial proximity and feature similarity can be partially attributed to the fact that six of the nine selected catchment characteristics are directly or indirectly related to climate, which in Chile is highly related to the geographical locations of the catchments. Parameter regression was the regionalisation method that provided the worst results (Figs. 6 and 8); however, Fig. 7 shows that this method generated good results over low-elevated areas of the Central Chile and South regions, where there are many potential donor catchments located nearby.

The compensation for P differences obtained through model calibration also affected the relative performance of regionalisation techniques, producing unrealistic parameter sets in some donor catchments. In particular, such compensation may have impacted the spatial transferability of model parameters with the parameter regression method. The main reason for this is that, unlike techniques that transfer the entire parameter sets, the regression process denatures the already uncertain model parameters by applying independent regression procedures using climate and physiographic characteristics (Arsenault and Brissette, 2014). This challenge can be overcome by simultaneously optimising both the model parameters and the regression equations (e.g.

Samaniego et al., 2010; Rakovec et al., 2016; Beck et al., 2020a), but such an exercise is outside of the scope of this study.

For both spatial proximity and feature similarity, the best and worst results were obtained for pluvio-nival catchments and rain-dominated catchments, respectively. Figure 9 shows the performances of the three regionalisation techniques according to hydrological regimes (see Fig. 1d) for the calibration period. Comparing Figs. 5 and 9, it is evident that the snow-dominated catchments performed substantially worse than in the independent performance during the same period (Fig. 5). On the other hand, the pluvio-nival catchments performed systematically better in the independent calibration and verification as well as in regionalisation. This could be attributed to (i) the ability of the model to reproduce Q in this regime and (ii) the increased likelihood of transferring model parameters from a catchment with the same hydrological regime, as they are grouped closed together and form 40 % of the total number of catchments.

5.4 Impact of nested catchments

Nested catchments play an important role in the performance of regionalisation methods as they are more likely to have a strong climatological and physiological similarity to each other. As observed in Fig. 10, the regionalisation method that was most impacted by the exclusion of nested catchments was spatial proximity, followed by feature similarity. These results are in agreement with previous studies, where the exclusion of nested catchments reduced the performance of regionalisation techniques (Merz and Blöschl, 2004; Oudin et al., 2008; Neri et al., 2020). Feature similarity only presented a slight decrease when the nested catchments were neglected, which can be attributed to the low degree of nestedness (i.e. the number of catchments that are nested in a larger one). As expected, the exclusion of nested catchments had a negligible effect on parameter regression, as the removal of relatively few catchments had a negligible impact on the non-linear relationships between the climatic and physiographic characteristics and the model parameters that were determined using all potential donor catchments. The reduction of regionalisation performance when the nested catchments were removed was lower than the reduction reported in a case study over Austria (Neri et al., 2020, their Figure 9a), which could be attributed to (i) the degree of nestedness, as the unique geography of Chile limits, to some extent, the number of nested catchments within any larger catchment (only 10 of the 100 selected catchments contained more than three nested catchments); and (ii) the percentage of catchments that are nested (42 % in this study, compared to 65 % in the Austrian case study).

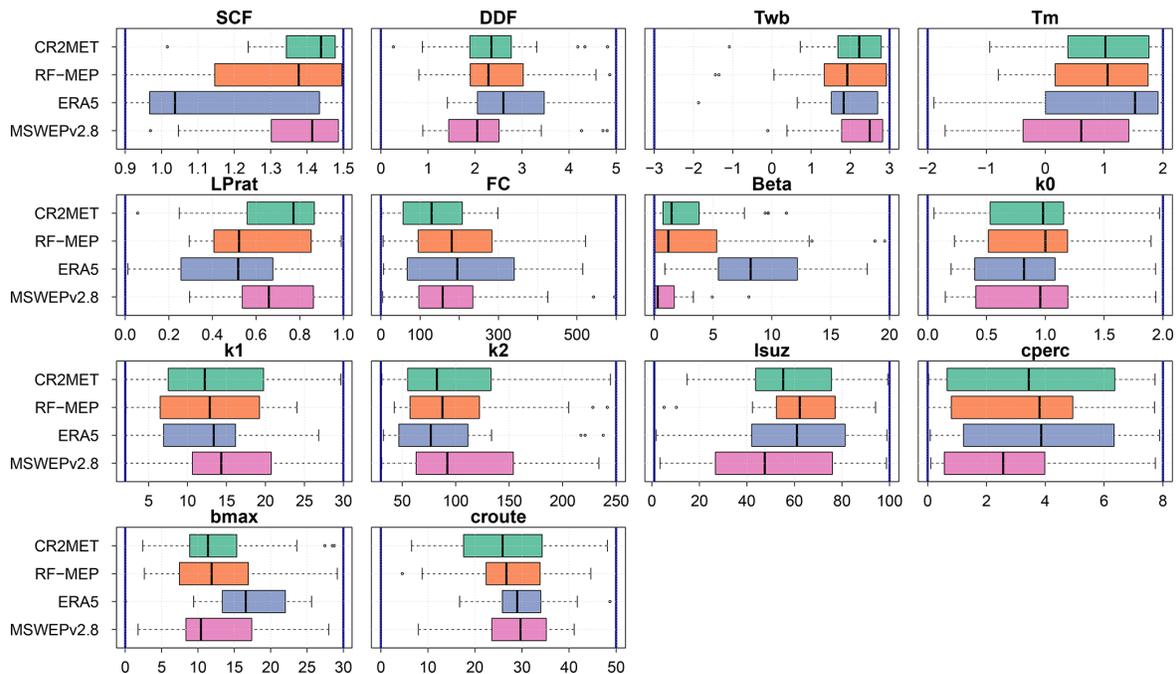


Figure 12. Model parameters obtained through calibration in nivo-pluvial catchments. The vertical blue lines indicate the upper and lower limits of the parameter ranges.

5.5 Impact of number of donor catchments

Increasing the number of donor catchments in feature similarity improved the regionalisation performance. This is in agreement with several studies that have demonstrated that using an ensemble of multiple donor catchments improves regionalisation results (McIntyre et al., 2005; Zelelew and Alfredsen, 2014; Garambois et al., 2015; Beck et al., 2016; Neri et al., 2020). Figure 11 shows that there is a slight increase in performance when four donors or more are used, independent of the P product and evaluated period. These results are similar to those of Neri et al. (2020), who determined that three donors were optimal for the TUWmodel over Austrian catchments. Feature similarity still outperformed spatial proximity when only one catchment was used to transfer the model parameters to the ungauged catchments, which is in agreement with multiple studies that have shown the ability of this method to produce good regionalisation results (Parajka et al., 2005; Oudin et al., 2008; Bao et al., 2012; Garambois et al., 2015; Neri et al., 2020).

6 Conclusion

Accurate streamflow predictions in ungauged catchments are critical for water resources management, and their generation is challenged by uncertainties arising from P products. In this paper, we assessed the relative performance of

three common regionalisation techniques (spatial proximity, feature similarity, and parameter regression) over 100 near-natural catchments located in the topographically and climatologically diverse Chilean territory. Four P products (CR2MET, RF-MEP, ERA5, and MSWEPv2.8) were used to force the semi-distributed TUWmodel at the daily timescale, using the KGE' as the calibration objective function and metric to assess (i) the impact of selecting different P forcings on the relative performance of regionalisation techniques and (ii) possible connections between regionalisation performance and hydrological regimes. Our key findings are as follows:

1. For the selected P products, the one that provided the best (worst) performance during independent calibration and verification did not necessarily yield the best (worst) results during regionalisation.
2. The P products corrected with daily ground-based measurements did not necessarily yield the best hydrological model performance. However, we expect that P products with lower performances than the ones used in this study might benefit from such a correction.
3. The spatial resolution of the P products did not noticeably affect model performance during the calibration and verification periods.

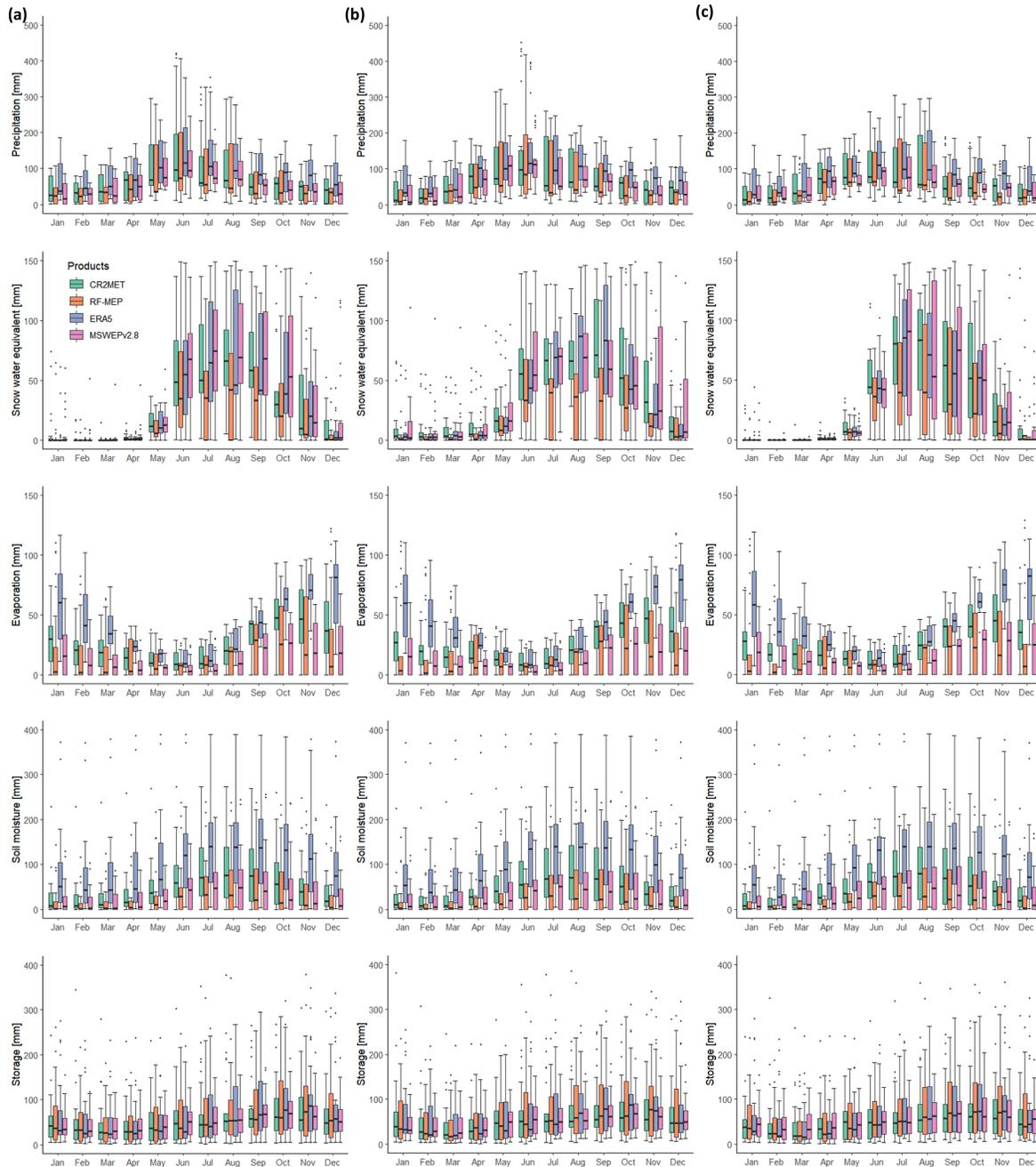


Figure 13. Mean monthly water balance components over nivo-pluvial catchments, obtained by forcing the TUW model with different *P* products for the (a) calibration (2000–2014), (b) Verification 1 (1990–1999), and (c) Verification 2 (2015–2018) periods. Mean monthly *P* was added for comparison purposes.

4. The TUWmodel was able to compensate, to some extent, the differences between *P* products through model calibration by adjusting the model parameters and, therefore, adjusting the water balance components (e.g. snow water equivalent, evaporation, and soil moisture).
5. Feature similarity was the best-performing regionalisation technique, regardless of the choice of gridded *P* product or hydrological regime.
6. Spatial proximity was the second best-performing regionalisation method because, in our study area, spatial proximity is a good proxy for climatic similarity for most neighbouring catchments.
7. Parameter regression provided the worst regionalisation performance, reinforcing the importance of transferring complete parameter sets to ungauged catchments.
8. The performance of regionalisation techniques can depend on the hydrological regime. We obtained the best results in pluvio-nival catchments with spatial proximity and feature similarity, while the same techniques provided the worst performance in rain-dominated catchments.
9. The exclusion of (relatively few) nested catchments had a minimal impact on the non-linear relationships between the climatic and physiographic characteristics (i.e. predictors) and model parameters (i.e. predictands), having a negligible effect on parameter regression results.
10. The performance of feature similarity increased when four or more catchments were used as donors; however, the differences in performance were not statistically significant when compared to the results of using only one donor.

modelling platforms (e.g. Clark et al., 2008; Knoben et al., 2019); and (v) the sensitivity of regionalisation results with respect to modified climate scenarios.

The results presented here are valid only for near-natural catchments across continental Chile. Nevertheless, they provide guidance for ongoing and future studies involving the application of gridded *P* products for regionalising hydrological model parameters in ungauged basins. The feature similarity procedure described here could be used to refine the parameter regionalisation approach adopted for national-scale hydrological characterisations in Chile (e.g. Bambach et al., 2018; Lagos et al., 2019). Additionally, further analyses could address (i) the effects that objective functions may have on the simulation of streamflow-derived hydrological signatures (e.g. Pool et al., 2017); (ii) other states and fluxes derived from remote sensing data (e.g. Dembélé et al., 2020); (iii) the influence of parameter equifinality (mainly for parameter regression), which can be accounted for by simultaneously optimising the model parameters and the regression equations, as described in Beck et al. (2020a); (iv) the use of additional model structures, implemented through flexible

Appendix A: Conceptual figure of hydrological regimes

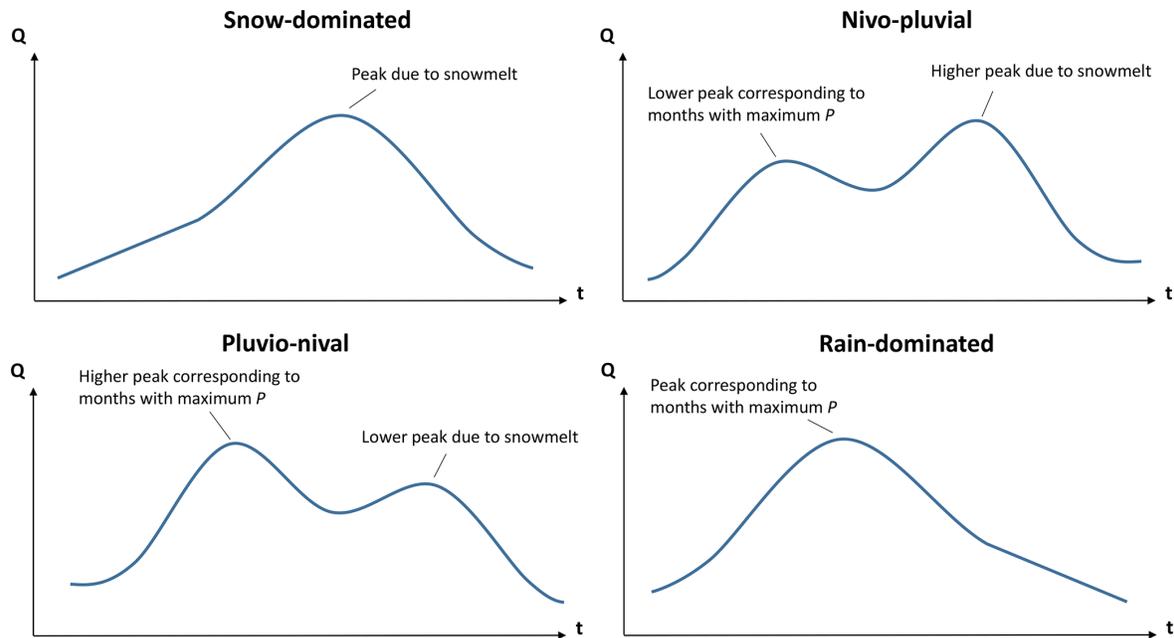


Figure A1. Conceptual illustration of the hydrological regimes used to classify the 100 near-natural catchments used in this study.

Appendix B: Selection of catchment characteristics for feature similarity

To avoid including redundant information when quantifying catchment similarity, we examined the correlations between the catchment characteristics described in Table 4. Figure B1 shows correlation matrices between catchment characteristics using the Pearson correlation (a) and the Spearman rank (b) correlation coefficients. We only present correlations obtained with CR2MET, since very similar results were obtained with the remaining P products. Because the mean and median elevation are highly correlated (values of 1.0 and 0.99 for the Pearson and Spearman correlation coefficients, respectively), we decided to keep the median elevation under the assumption that it is more representative of topographic conditions, given the pronounced elevation gradients in continental Chile. Similarly, mean annual PE was excluded because of its high correlation with mean annual T (0.87 and 0.86 for the Pearson and Spearman correlation coefficients, respectively), notwithstanding that T was used to calculate PE . SDII was also excluded due to its high correlation to the Rx5day (0.97 for both coefficients). Finally, we excluded the snow cover from CAMELS-CL, as we found it to be unreliable over the snow-dominated catchments selected in our analysis.

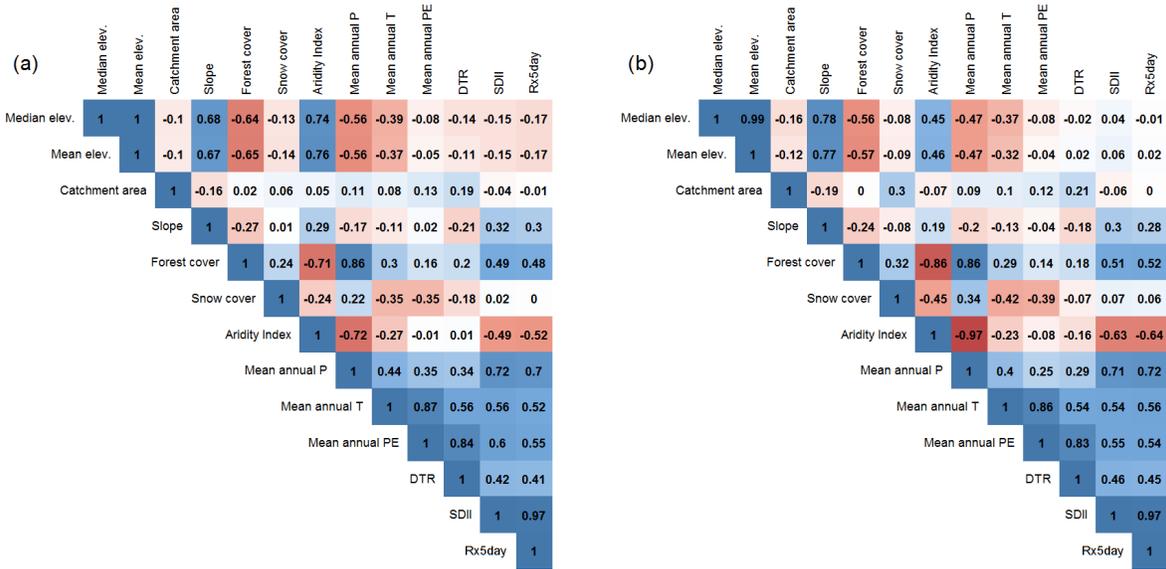


Figure B1. Correlation matrices of the catchment characteristics described in Table 4 using CR2MET as the *P* product for (a) the Pearson correlation, to evaluate linear correlation, and (b) the Spearman correlation, to evaluate the monotonic correlation.

Appendix C: Performance of the components of the KGE'

Table C1. Quantiles 0.25 and 0.75 of the correlation coefficient (*r*) of the KGE' over the selected catchments.

Pearson correlation (<i>r</i>)	CR2MET	RF-MEP	ERA5	MSWEPv2.8
Calibration (cal.)	0.78–0.90	0.77–0.88	0.71–0.86	0.77–0.88
Verification 1 (Ver. 1)	0.74–0.88	0.72–0.87	0.67–0.87	0.69–0.86
Verification 2 (Ver. 2)	0.68–0.86	0.59–0.85	0.59–0.86	0.67–0.85
Spatial proximity (cal.)	0.70–0.87	0.68–0.84	0.57–0.82	0.66–0.84
Spatial proximity (Ver. 1)	0.66–0.86	0.63–0.84	0.61–0.84	0.62–0.84
Spatial proximity (Ver. 2)	0.61–0.83	0.51–0.82	0.56–0.83	0.59–0.82
Feature similarity (cal.)	0.74–0.89	0.71–0.88	0.69–0.85	0.72–0.88
Feature similarity (Ver. 1)	0.69–0.88	0.70–0.88	0.67–0.88	0.69–0.86
Feature similarity (Ver. 2)	0.64–0.87	0.59–0.85	0.64–0.87	0.65–0.84
Parameter regression (cal.)	0.54–0.80	0.54–0.69	0.60–0.82	0.42–0.63
Parameter regression (Ver. 1)	0.58–0.80	0.50–0.68	0.64–0.86	0.43–0.62
Parameter regression (Ver. 2)	0.50–0.79	0.43–0.65	0.59–0.84	0.37–0.57

Table C2. Quantiles 0.25 and 0.75 of the bias ratio (β) of the KGE' over the selected catchments.

Bias ratio (β)	CR2MET	RF-MEP	ERA5	MSWEPv2.8
Calibration (cal.)	0.95–0.99	0.93–1.01	0.97–1.02	0.90–1.02
Verification 1 (Ver. 1)	0.89–1.03	0.84–1.02	0.90–1.12	0.77–1.04
Verification 2 (Ver. 2)	0.96–1.19	0.86–1.11	1.00–1.25	0.74–1.06
Spatial proximity (cal.)	0.73–1.09	0.70–1.15	0.74–1.22	0.70–1.13
Spatial proximity (Ver. 1)	0.72–1.12	0.70–1.12	0.72–1.22	0.69–1.08
Spatial proximity (Ver. 2)	0.73–1.30	0.73–1.23	0.77–1.46	0.68–1.14
Feature similarity (cal.)	0.81–1.19	0.78–1.29	0.81–1.35	0.68–1.3
Feature similarity (Ver. 1)	0.80–1.17	0.74–1.24	0.80–1.36	0.69–1.29
Feature similarity (Ver. 2)	0.86–1.40	0.77–1.40	0.86–1.57	0.69–1.27
Parameter regression (cal.)	0.99–2.04	0.89–1.72	0.76–1.78	0.82–3.07
Parameter regression (Ver. 1)	0.99–1.73	0.87–1.65	0.76–1.62	0.83–2.64
Parameter regression (Ver. 2)	1.10–2.05	0.90–1.83	0.88–1.94	0.83–2.54

Table C3. Quantiles 0.25 and 0.75 of the variability ratio (γ) of the KGE' over the selected catchments.

Variability ratio (γ)	CR2MET	RF-MEP	ERA5	MSWEPv2.8
Calibration (cal.)	0.97–1.00	0.95–1.00	0.95–1.01	0.96–1.01
Verification 1 (Ver. 1)	0.93–1.07	0.92–1.06	0.93–1.07	0.93–1.11
Verification 2 (Ver. 2)	0.92–1.13	0.91–1.17	0.91–1.12	0.79–1.05
Spatial proximity (cal.)	0.84–1.20	0.84–1.23	0.88–1.24	0.88–1.22
Spatial proximity (Ver. 1)	0.89–1.24	0.84–1.30	0.85–1.32	0.86–1.27
Spatial proximity (Ver. 2)	0.88–1.34	0.85–1.37	0.85–1.38	0.75–1.19
Feature similarity (cal.)	0.74–1.06	0.75–1.06	0.75–1.10	0.78–1.07
Feature similarity (Ver. 1)	0.79–1.04	0.76–1.06	0.77–1.07	0.81–1.03
Feature similarity (Ver. 2)	0.79–1.13	0.75–1.12	0.79–1.15	0.66–0.97
Parameter regression (cal.)	0.80–1.18	1.02–1.50	0.84–1.23	1.26–1.89
Parameter regression (Ver. 1)	0.82–1.20	1.02–1.35	0.87–1.25	1.27–1.69
Parameter regression (Ver. 2)	0.86–1.38	1.15–1.83	0.86–1.46	1.22–1.82

Code availability. The codes used in the development of all analyses will be made available upon request.

Data availability. The datasets used in this study are open-access and can be retrieved from their respective websites. Please see Sect. 3.3.1 and 3.1.2 and Table 4.

Supplement. The supplement related to this article is available online at: <https://doi.org/10.5194/hess-25-5805-2021-supplement>.

Author contributions. OMBV led the investigation, conducted the analysis, and wrote the original draft; OMBV, MZB, and PAM conceived and developed the methodology of the manuscript, and supervised the project; HEB and IM provided methodological feedback; JT supported in the development of the algorithms; MZB, PAM, HEB, IM, JT, AN, LR, and NXT reviewed and edited the manuscript; and LR acquired the funding.

Competing interests. The authors declare that they have no conflict of interest.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. We particularly thank Jim Freer, Juraj Parajka, Elena Toth, and Koray Kamil Yilmaz, whose constructive comments helped to improve the quality of the final paper. We would also like to thank the HESS editorial team for their support; the Centers for Natural Resources and Development (CNRD) PhD programme for their financial support for the main author; the CAMELS-CL dataset (<http://camels.cr2.cl/>, last access: 17 July 2021); Camila Álvarez-Garretón for providing an initial dataset of catchment that could be considered as being undisturbed for our analysis; Juan Pablo Boisier for providing the rain gauges used for CR2METv2; and Rodrigo Marinao Rivas for his support in the classification of the catchments into hydrological regimes. The authors are also grateful to the active R community for unselfish and prompt support, in particular to Robert J. Hijmans, Alberto Viglione, and Juraj Parajka for developing and maintaining the `raster` and `TUWmodel` R packages, respectively.

Financial support. This research has been supported by the Conicyt-Fondecyt (grant nos. 11150861 and 11200142) and the CONICYT/PIA Project (grant no. AFB180004).

Review statement. This paper was edited by Jim Freer and reviewed by Elena Toth, Juraj Parajka, and Koray Kamil Yilmaz.

References

- Abdelaziz, R., Merkel, B. J., Zambrano-Bigiarini, M., and Nair, S.: Particle swarm optimization for the estimation of surface complexation constants with the geochemical model PHREEQC-3.1.2, *Geosci. Model Dev.*, 12, 167–177, <https://doi.org/10.5194/gmd-12-167-2019>, 2019.
- Addor, N., Jaun, S., Fundel, F., and Zappa, M.: An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): skill, case studies and scenarios, *Hydrol. Earth Syst. Sci.*, 15, 2327–2347, <https://doi.org/10.5194/hess-15-2327-2011>, 2011.
- Addor, N., Nearing, G., Prieto, C., Newman, A., Le Vine, N., and Clark, M. P.: A ranking of hydrological signatures based on their predictability in space, *Water Resour. Res.*, 54, 8792–8812, 2018.
- Adhikary, S. K., Yilmaz, A. G., and Muttill, N.: Optimal design of rain gauge network in the Middle Yarra River catchment, Australia, *Hydrol. Process.*, 29, 2582–2599, 2015.
- Alvarez-Garretón, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Puelma, C., Cortes, G., Garreaud, R., McPhee, J., and Ayala, A.: The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies – Chile dataset, *Hydrol. Earth Syst. Sci.*, 22, 5817–5846, <https://doi.org/10.5194/hess-22-5817-2018>, 2018.
- Arsenault, R. and Brissette, F. P.: Continuous streamflow prediction in ungauged basins: The effects of equifinality and parameter set selection on uncertainty in regionalization approaches, *Water Resour. Res.*, 50, 6135–6153, 2014.
- Artan, G., Gadain, H., Smith, J. L., Asante, K., Bandaragoda, C. J., and Verdin, J. P.: Adequacy of satellite derived rainfall data for stream flow modeling, *Nat. Hazards*, 43, 167, <https://doi.org/10.1007/s11069-007-9121-6>, 2007.
- Astagneau, P. C., Thirel, G., Delaigüe, O., Guillaume, J. H. A., Parajka, J., Brauer, C. C., Viglione, A., Buytaert, W., and Beven, K. J.: Technical note: Hydrology modelling R packages – a unified analysis of models and practicalities from a user perspective, *Hydrol. Earth Syst. Sci.*, 25, 3937–3973, <https://doi.org/10.5194/hess-25-3937-2021>, 2021.
- Athira, P., Sudheer, K., Cibir, R., and Chaubey, I.: Predictions in ungauged basins: an approach for regionalization of hydrological models considering the probability distribution of model parameters, *Stoch. Env. Res. Risk A.*, 30, 1131–1149, 2016.
- Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Ribbe, L., Nauditt, A., Giraldo-Osorio, J. D., and Thinh, N. X.: Temporal and spatial evaluation of satellite rainfall estimates over different regions in Latin-America, *Atmos. Res.*, 213, 34–50, <https://doi.org/10.1016/j.atmosres.2018.05.011>, 2018.
- Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Beck, H. E., McNamara, I., Ribbe, L., Nauditt, A., Birkel, C., Verbist, K., Giraldo-Osorio, J. D., and Thinh, N. X.: RF-MEP: A novel Random Forest method for merging gridded precipitation products and ground-based measurements, *Remote Sens. Environ.*, 239, 111606, <https://doi.org/10.1016/j.rse.2019.111606>, 2020.
- Bambach, N., Bustos, E., Meza, F., Morales, D., Suarez, F., and Na, V.: Aplicación de La Metodología de Actualización del Balance Hídrico Nacional en las Cuencas de la Macrozona Norte y Centro, Dirección General de Aguas (DGA), Santiago, 2018.
- Bao, Z., Zhang, J., Liu, J., Fu, G., Wang, G., He, R., Yan, X., Jin, J., and Liu, H.: Comparison of regionalization approaches based on regression and similarity for predictions in ungauged catchments

- under multiple hydro-climatic conditions, *J. Hydrol.*, 466, 37–46, 2012.
- Beck, H. E., van Dijk, A. I., De Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., and Bruijnzeel, L. A.: Global-scale regionalization of hydrologic model parameters, *Water Resour. Res.*, 52, 3599–3622, 2016.
- Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., and de Roo, A.: MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data, *Hydrol. Earth Syst. Sci.*, 21, 589–615, <https://doi.org/10.5194/hess-21-589-2017>, 2017a.
- Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J., and Wood, E. F.: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling, *Hydrol. Earth Syst. Sci.*, 21, 6201–6217, <https://doi.org/10.5194/hess-21-6201-2017>, 2017b.
- Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F.: Present and future Köppen-Geiger climate classification maps at 1-km resolution, *Sci. Data*, 5, 180214, <https://doi.org/10.1038/sdata.2018.214>, 2018.
- Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., Van Dijk, A. I., McVicar, T. R., and Adler, R. F.: MSWEP V2 global 3-hourly 0.1 precipitation: methodology and quantitative assessment, *B. Am. Meteorol. Soc.*, 100, 473–500, 2019.
- Beck, H. E., Pan, M., Lin, P., Seibert, J., van Dijk, A. I., and Wood, E. F.: Global fully distributed parameter regionalization based on observed streamflow from 4,229 headwater catchments, *J. Geophys. Res.-Atmos.*, 125, e2019JD031485, <https://doi.org/10.1029/2019JD031485>, 2020a.
- Beck, H. E., Wood, E. F., McVicar, T. R., Zambrano-Bigiarini, M., Alvarez-Garretón, C., Baez-Villanueva, O. M., Sheffield, J., and Karger, D. N.: Bias correction of global high-resolution precipitation climatologies using streamflow observations from 9372 catchments, *J. Climate*, 33, 1299–1315, 2020b.
- Bergström, S.: Development and application of a conceptual runoff model for Scandinavian catchments, Sveriges Meteorologiska Och Hydrologiska Institut, Norrköping, 1976.
- Bergström, S.: The HBV model, Computer models of watershed hydrology, Water Resources Publications, Highlands Ranch, CO, 1995.
- Beven, K. J.: Changing ideas in hydrology - The case of physically-based models, *J. Hydrol.*, 105, 157–172, [https://doi.org/10.1016/0022-1694\(89\)90101-7](https://doi.org/10.1016/0022-1694(89)90101-7), 1989.
- Beven, K. J.: Uniqueness of place and process representations in hydrological modelling, *Hydrol. Earth Syst. Sci.*, 4, 203–213, <https://doi.org/10.5194/hess-4-203-2000>, 2000.
- Beven, K. J.: A manifesto for the equifinality thesis, *J. Hydrol.*, 320, 18–36, <https://doi.org/10.1016/j.jhydrol.2005.07.007>, 2006.
- Biau, G. and Scornet, E.: A random forest guided tour, *TEST*, 25, 197, <https://doi.org/10.1007/s11749-016-0481-7>, 2016.
- Bisselink, B., Zambrano-Bigiarini, M., Burek, P., and de Roo, A.: Assessing the role of uncertain precipitation estimates on the robustness of hydrological model parameters under highly variable climate conditions, *J. Hydrol.-Regional Studies*, 8, 112–129, <https://doi.org/10.1016/j.ejrh.2016.09.003>, 2016.
- Bitew, M. M., Gebremichael, M., Ghebremichael, L. T., and Bayissa, Y. A.: Evaluation of High-Resolution Satellite Rainfall Products through Streamflow Simulation in a Hydrological Modeling of a Small Mountainous Watershed in Ethiopia, *J. Hydrometeorol.*, 13, 338–350, <https://doi.org/10.1175/2011JHM1292.1>, 2012.
- Bivand, R. and Rundel, C.: rgeos: Interface to Geometry Engine – Open Source (“GEOS”), available at: <https://CRAN.R-project.org/package=rgeos> (last access: 10 December 2020), r package version 0.5-3, 2020.
- Bivand, R., Keitt, T., and Rowlingson, B.: rgdal: Bindings for the “Geospatial” Data Abstraction Library, available at: <https://CRAN.R-project.org/package=rgdal> (last access: 10 December 2020), r package version 1.5-12, 2020.
- Boisier, J. P., Rondanelli, R., Garreaud, R. D., and Muñoz, F.: Anthropogenic and natural contributions to the South-east Pacific precipitation decline and recent megadrought in central Chile, *Geophys. Res. Lett.*, 43, 413–421, <https://doi.org/10.1002/2015GL067265>, 2016.
- Boisier, J. P., Alvarez-Garretón, C., Cepeda, J., Osses, A., Vásquez, N., and Rondanelli, R.: CR2MET: A high-resolution precipitation and temperature dataset for hydroclimatic research in Chile, EGUGA, p. 19739, 2018.
- Brauer, C. C., Torfs, P. J. J. F., Teuling, A. J., and Uijlenhoet, R.: The Wageningen Lowland Runoff Simulator (WALRUS): application to the Hupsel Brook catchment and the Cabauw polder, *Hydrol. Earth Syst. Sci.*, 18, 4007–4028, <https://doi.org/10.5194/hess-18-4007-2014>, 2014.
- Breiman, L.: Random Forests, *Mach. Learn.*, 45, 5–32, <https://doi.org/10.1023/A:1010933404324>, 2001.
- Carrillo, G., Troch, P. A., Sivapalan, M., Wagener, T., Harman, C., and Sawicz, K.: Catchment classification: hydrological analysis of catchment behavior through process-based modeling along a climate gradient, *Hydrol. Earth Syst. Sci.*, 15, 3411–3430, <https://doi.org/10.5194/hess-15-3411-2011>, 2011.
- Ceola, S., Arheimer, B., Baratti, E., Blöschl, G., Capell, R., Castellarin, A., Freer, J., Han, D., Hrachowitz, M., Hundscha, Y., Hutton, C., Lindström, G., Montanari, A., Nijzink, R., Parajka, J., Toth, E., Viglione, A., and Wagener, T.: Virtual laboratories: new opportunities for collaborative water science, *Hydrol. Earth Syst. Sci.*, 19, 2101–2117, <https://doi.org/10.5194/hess-19-2101-2015>, 2015.
- Ciabatta, L., Brocca, L., Massari, C., Moramarco, T., Gabellani, S., Puca, S., and Wagner, W.: Rainfall-runoff modelling by using SM2RAIN-derived and state-of-the-art satellite rainfall products over Italy, *Int. J. Appl. Earth Obs.*, 48, 163–173, 2016.
- Clark, M. P. and Hay, L. E.: Use of medium-range numerical weather prediction model output to produce forecasts of streamflow, *J. Hydrometeorol.*, 5, 15–32, 2004.
- Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, *Water Resour. Res.*, 44, W00B02, <https://doi.org/10.1029/2007WR006735>, 2008.
- Clerc, M.: From theory to practice in particle swarm optimization, in: *Handbook of Swarm Intelligence*, 3–36, Springer, Berlin, Germany, 2011a.
- Clerc, M.: Standard particle swarm optimisation from 2006 to 2011, *Particle Swarm Central*, France, 253, 1–15, 2011b.
- Coughlan de Perez, E., van den Hurk, B., van Aalst, M. K., Amuron, I., Bamanya, D., Hauser, T., Jongma, B., Lopez, A., Mason, S.,

- Mendler de Suarez, J., Pappenberger, F., Rueth, A., Stephens, E., Suarez, P., Wagemaker, J., and Zsoter, E.: Action-based flood forecasting for triggering humanitarian action, *Hydrol. Earth Syst. Sci.*, 20, 3549–3560, <https://doi.org/10.5194/hess-20-3549-2016>, 2016.
- Dallery, D., Squidant, H., De Lavenne, A., Launay, J., and Cudennec, C.: An end-user-friendly hydrological Web Service for hydrograph prediction in ungauged basins, *Hydrol. Sci. J.*, 0, 1–9, <https://doi.org/10.1080/02626667.2020.1797045>, 2020.
- Dembélé, M., Hrachowitz, M., Savenije, H. H., Mariéthoz, G., and Schaefli, B.: Improving the predictive skill of a distributed hydrological model by calibration on spatial patterns with multiple satellite data sets, *Water Resour. Res.*, 56, e2019WR026085, <https://doi.org/10.1029/2019WR026085>, 2020.
- DGA: Plan director para la gestión de los recursos hídricos en la cuenca del río San José, Tech. rep., Dirección General de Aguas, Santiago, available at: <https://snia.mop.gob.cl/sad/ADM600v1.pdf> (last access: 9 June 2021), 1998.
- DGA: Recursos hídricos compartidos con la República Argentina : ficha temática de la cuenca del río Grande de Tierra del Fuego, Tech. rep., Dirección General de Aguas, Santiago, available at: <https://snia.mop.gob.cl/sad/CUH2087.pdf> (last access: 9 June 2021), 1999.
- DGA: Cuenca Quebrada de Tarapacá, Tech. rep., Dirección General de Aguas, Santiago, available at: <https://mma.gob.cl/wp-content/uploads/2017/12/Tarapaca.pdf> (last access: 9 June 2021), 2004a.
- DGA: Cuenca Río Loa, Tech. rep., Dirección General de Aguas, Santiago, available at: <https://mma.gob.cl/wp-content/uploads/2017/12/Loa.pdf> (last access: 9 June 2021), 2004b.
- DGA: Cuenca del Río Elqui, Tech. rep., Dirección General de Aguas, Santiago, available at: <https://mma.gob.cl/wp-content/uploads/2017/12/Elqui.pdf> (last access: 9 June 2021), 2004c.
- DGA: Evaluación de los recursos hídricos superficiales de las cuencas de los ríos Petorca y La Ligua V Región, Tech. rep., Dirección General de Aguas, Santiago, available at: <https://snia.mop.gob.cl/sad/SUP4496.pdf> (last access: 9 June 2021), 2006.
- DGA: Análisis integral de soluciones a la escasez hídrica, región de Arica y Parinacota: informe final, Tech. rep., Dirección General de Aguas, Santiago, available at: <https://snia.mop.gob.cl/sad/REH5720.pdf> (last access: 9 June 2021), 2016a.
- DGA: Actualización de Información y Modelación Hidrológica Acuíferos de la XII Región, de Magallanes y la Antártica Chilena : Informe definitivo etapa II, Tech. rep., Dirección General de Aguas, Santiago, available at: <https://snia.mop.gob.cl/sad/SUB5698.pdf> (last access: 9 June 2021), 2016b.
- DGA: Herramientas de gestión y actualización de los modelos numéricos del acuífero de Copiapó: informe final, Tech. rep., Dirección General de Aguas, Santiago, available at: <https://snia.mop.gob.cl/sad/SUB5851v1.pdf> (last access: 9 June 2021), 2018.
- Díaz-Uriarte, R. and Alvarez de Andrés, S.: Gene selection and classification of microarray data using random forest., *BMC Bioinformatics*, 7, 3, <https://doi.org/10.1186/1471-2105-7-3>, 2006.
- Ding, J., Wallner, M., Müller, H., and Haberlandt, U.: Estimation of instantaneous peak flows from maximum mean daily flows using the HBV hydrological model, *Hydrol. Process.*, 30, 1431–1448, 2016.
- Eberhart, R. and Kennedy, J.: A new optimizer using particle swarm theory, in: *Micro Machine and Human Science*, 1995. MHS '95, Proceedings of the Sixth International Symposium on, 39–43, <https://doi.org/10.1109/MHS.1995.494215>, 1995.
- Elsner, M. M., Gangopadhyay, S., Pruitt, T., Brekke, L. D., Mizukami, N., and Clark, M. P.: How does the choice of distributed meteorological data affect hydrologic model calibration and streamflow simulations?, *J. Hydrometeorol.*, 15, 1384–1403, 2014.
- Fernandez, W., Vogel, R., and Sankarasubramanian, A.: Regional calibration of a watershed model, *Hydrol. Sci. J.*, 45, 689–707, 2000.
- Galleguillos, M., Gimeno, F., Puelma, C., Zambrano-Bigiarini, M., Lara, A., and Rojas, M.: Disentangling the effect of future land use strategies and climate change on streamflow in a Mediterranean catchment dominated by tree plantations, *J. Hydrol.*, 595, 126047, <https://doi.org/10.1016/j.jhydrol.2021.126047>, 2021.
- Garambois, P.-A., Roux, H., Larnier, K., Labat, D., and Dartus, D.: Parameter regionalization for a process-oriented distributed model dedicated to flash floods, *J. Hydrol.*, 525, 383–399, 2015.
- García, F., Folton, N., and Oudin, L.: Which objective function to calibrate rainfall–runoff models for low-flow index simulations?, *Hydrol. Sci. J.*, 62, 1149–1166, 2017.
- Garreaud, R. D., Alvarez-Garreton, C., Barichivich, J., Boisier, J. P., Christie, D., Galleguillos, M., LeQuesne, C., McPhee, J., and Zambrano-Bigiarini, M.: The 2010–2015 megadrought in central Chile: impacts on regional hydroclimate and vegetation, *Hydrol. Earth Syst. Sci.*, 21, 6307–6327, <https://doi.org/10.5194/hess-21-6307-2017>, 2017.
- Garreaud, R. D., Boisier, J. P., Rondanelli, R., Montecinos, A., Sepúlveda, H. H., and Veloso-Aguila, D.: The Central Chile Mega Drought (2010–2018): A climate dynamics perspective, *International J. Climatol.*, 40, 421–439, 2020.
- Guo, Y., Zhang, Y., Zhang, L., and Wang, Z.: Regionalization of hydrological modeling for predicting streamflow in ungauged catchments: A comprehensive review, *Wiley Interdisciplinary Reviews: Water*, e1487, <https://doi.org/10.1002/wat2.1487>, 2021.
- Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, 377, 80–91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>, 2009.
- Hann, H., Nauditt, A., Zambrano-Bigiarini, M., Thurner, J., McNamara, I., and Ribbe, L.: Combining satellite-based rainfall data with rainfall-runoff modelling to simulate low flows in a Southern Andean catchment, *J. Nat. Res. Dev.*, 11, 1–19, <https://doi.org/10.18716/ojs/jnrd/2021.11.02>, 2021.
- Hargreaves, G. H. and Samani, Z. A.: Reference crop evapotranspiration from ambient air temperature, American Society of Agricultural Engineers, (fiche no. 85-2517), (Microfiche collection), USA, 1985.
- Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B., and Gräler, B.: Random Forest as a generic framework for predictive modeling of spatial and spatio-temporal variables, *PeerJ*, 6, e5518, <https://doi.org/10.7717/peerj.5518>, 2018.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Fleming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S.,

- Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, *Q. J. Roy. Meteor. Soc.*, 146, 1999–2049, 2020.
- Hijmans, R. J.: raster: Geographic Data Analysis and Modeling, available at: <https://CRAN.R-project.org/package=raster> (last access: 12 December 2021), r package version 3.3-13, 2020.
- Hofstra, N., New, M., and McSweeney, C.: The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data, *Clim. Dynam.*, 35, 841–858, 2010.
- Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivalapan, M., Pomeroy, J., Arheimer, B., Blume, T., Clark, M., Ehret, U., et al.: A decade of Predictions in Ungauged Basins (PUB) – a review, *Hydrol. Sci. J.*, 58, 1198–1255, 2013.
- Huang, S., Eisner, S., Magnusson, J. O., Lussana, C., Yang, X., and Beldring, S.: Improvements of the spatially distributed hydrological modelling using the HBV model at 1 km resolution for Norway, *J. Hydrol.*, 577, 123585, <https://doi.org/10.1016/j.jhydrol.2019.03.051>, 2019.
- Jansen, K. F., Teuling, A. J., Craig, J. R., Dal Molin, M., Knoben, W. J., Parajka, J., Vis, M., and Melsen, L. A.: Mimicry of a Conceptual Hydrological Model (HBV): What's in a Name?, *Water Resour. Res.*, 57, e2020WR029143, <https://doi.org/10.1029/2020WR029143>, 2021.
- Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, E.: Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90 m Database, available at: <http://srtm.csi.cgiar.org> (last access: 23 November 2020), 15, 25–54, 2008.
- Jehn, F. U., Bestian, K., Breuer, L., Kraft, P., and Houska, T.: Using hydrological and climatic catchment clusters to explore drivers of catchment behavior, *Hydrol. Earth Syst. Sci.*, 24, 1081–1100, <https://doi.org/10.5194/hess-24-1081-2020>, 2020.
- Karl, T. R., Nicholls, N., and Ghazi, A.: Clivar/GCOS/WMO workshop on indices and indicators for climate extremes workshop summary, in: *Weather and climate extremes*, 3–7, Springer, Dordrecht, the Netherlands, 1999.
- Kearney, M. R. and Maino, J. L.: Can next-generation soil data products improve soil moisture modelling at the continental scale? An assessment using a new microclimate package for the R programming environment, *J. Hydrol.*, 561, 662–673, <https://doi.org/10.1016/j.jhydrol.2018.04.040>, 2018.
- Kennedy, J. and Eberhart, R.: Particle swarm optimization, in: *Neural Networks, 1995. Proceedings., IEEE International Conference on*, 4, 1942–1948, <https://doi.org/10.1109/ICNN.1995.488968>, 1995.
- Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, *J. Hydrol.*, 424, 264–277, <https://doi.org/10.1016/j.jhydrol.2012.01.011>, 2012.
- Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, *Hydrol. Earth Syst. Sci.*, 23, 4323–4331, <https://doi.org/10.5194/hess-23-4323-2019>, 2019.
- Koffler, D., Gauster, T., and Laaha, G.: Ifstat: Calculation of Low Flow Statistics for Daily Stream Flow Data, available at: <https://CRAN.R-project.org/package=lfstat> (last access: 12 December 2020), r package version 0.9.4, 2016.
- Kuentz, A., Arheimer, B., Hundecha, Y., and Wagener, T.: Understanding hydrologic variability across Europe through catchment classification, *Hydrol. Earth Syst. Sci.*, 21, 2863–2879, <https://doi.org/10.5194/hess-21-2863-2017>, 2017.
- Kundu, D., Vervoort, R. W., and van Ogtrop, F. F.: The value of remotely sensed surface soil moisture for model calibration using SWAT, *Hydrol. Process.*, 31, 2764–2780, <https://doi.org/10.1002/hyp.11219>, 2017.
- Lagos, M., Mendoza, P., Rondanelli, R., Daniele, D., and Tomaás, G.: Aplicación de La metodología de actualización del balance hídrico nacional en las cuencas de la Macrozona Sur y parte de la Macrozona Austral, Dirección General de Aguas (DGA), Santiago, Chile, 2019.
- Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *J. Geophys. Res.-Atmos.*, 99, 14415–14428, 1994.
- Liaw, A. and Wiener, M.: Classification and Regression by random Forest, *R News*, 2, 18–22, 2002.
- Lindström, G.: A simple automatic calibration routine for the HBV model, *Hydrol. Res.*, 28, 153–168, 1997.
- Maggioni, V. and Massari, C.: On the performance of satellite precipitation products in riverine flood modeling: A review, *J. Hydrol.*, 558, 214–224, 2018.
- Maggioni, V., Vergara, H. J., Anagnostou, E. N., Gourley, J. J., Hong, Y., and Stampoulis, D.: Investigating the applicability of error correction ensembles of satellite rainfall products in river flow simulations, *J. Hydrometeorol.*, 14, 1194–1211, 2013.
- McIntyre, N., Lee, H., Wheeler, H., Young, A., and Wagener, T.: Ensemble predictions of runoff in ungauged catchments, *Water Resour. Res.*, 41, W12434, <https://doi.org/10.1029/2005WR004289>, 2005.
- Melsen, L. A., Addor, N., Mizukami, N., Newman, A. J., Torfs, P. J. J. F., Clark, M. P., Uijlenhoet, R., and Teuling, A. J.: Mapping (dis)agreement in hydrologic projections, *Hydrol. Earth Syst. Sci.*, 22, 1775–1791, <https://doi.org/10.5194/hess-22-1775-2018>, 2018.
- Mendoza, P. A., Clark, M. P., Mizukami, N., Gutmann, E. D., Arnold, J. R., Brekke, L. D., and Rajagopalan, B.: How do hydrologic modeling decisions affect the portrayal of climate change impacts?, *Hydrol. Process.*, 30, 1071–1095, 2016.
- Merz, R. and Blöschl, G.: Regionalisation of catchment model parameters, *J. Hydrol.*, 287, 95–123, 2004.
- Mizukami, N., Rakovec, O., Newman, A. J., Clark, M. P., Wood, A. W., Gupta, H. V., and Kumar, R.: On the choice of calibration metrics for “high-flow” estimation using hydrologic models, *Hydrol. Earth Syst. Sci.*, 23, 2601–2614, <https://doi.org/10.5194/hess-23-2601-2019>, 2019.
- Neri, M., Parajka, J., and Toth, E.: Importance of the informative content in the study area when regionalising rainfall-runoff model parameters: the role of nested catchments and gauging station density, *Hydrol. Earth Syst. Sci.*, 24, 5149–5171, <https://doi.org/10.5194/hess-24-5149-2020>, 2020.
- Nikolopoulos, E. I., Anagnostou, E. N., and Borga, M.: Using high-resolution satellite rainfall products to simulate a major flash flood event in northern Italy, *J. Hydrometeorol.*, 14, 171–185, 2013.
- Ollivier, C., Mazzilli, N., Oliosio, A., Chalikakis, K., Carrière, S. D., Danquigny, C., and Emblanch, C.: Karst

- recharge-discharge semi distributed model to assess spatial variability of flows, *Sci. Total Environ.*, 703, 134368, <https://doi.org/10.1016/j.scitotenv.2019.134368>, 2020.
- Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Le Moine, N.: Spatial proximity, physical similarity, regression and ungauged catchments: A comparison of regionalization approaches based on 913 French catchments, *Water Resour. Res.*, 44, W03413, <https://doi.org/10.1029/2007WR006240>, 2008.
- Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalisation methods for catchment model parameters, *Hydrol. Earth Syst. Sci.*, 9, 157–171, <https://doi.org/10.5194/hess-9-157-2005>, 2005.
- Parajka, J., Merz, R., and Blöschl, G.: Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments, *Hydrol. Process.*, 21, 435–446, 2007.
- Parajka, J., Viglione, A., Rogger, M., Salinas, J. L., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 1: Runoff-hydrograph studies, *Hydrol. Earth Syst. Sci.*, 17, 1783–1795, <https://doi.org/10.5194/hess-17-1783-2013>, 2013.
- Parajka, J., Blaschke, A. P., Blöschl, G., Haslinger, K., Hepp, G., Laaha, G., Schöner, W., Trautvetter, H., Viglione, A., and Zessner, M.: Uncertainty contributions to low-flow projections in Austria, *Hydrol. Earth Syst. Sci.*, 20, 2085–2101, <https://doi.org/10.5194/hess-20-2085-2016>, 2016.
- Perpiñán, O. and Hijmans, R.: rasterVis, available at: <https://oscarperpinan.github.io/rastervis/> (last access: 12 December 2020), r package version 0.49, 2020.
- Pokhrel, P., Yilmaz, K. K., and Gupta, H. V.: Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures, *J. Hydrol.*, 418, 49–60, 2012.
- Pool, S., Vis, M. J. P., Knight, R. R., and Seibert, J.: Streamflow characteristics from modeled runoff time series – importance of calibration criteria selection, *Hydrol. Earth Syst. Sci.*, 21, 5443–5457, <https://doi.org/10.5194/hess-21-5443-2017>, 2017.
- Prasad, A. M., Iverson, L. R., and Liaw, A.: Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction, *Ecosystems*, 9, 181, <https://doi.org/10.1007/s10021-005-0054-1>, 2006.
- R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, available at: <https://www.R-project.org/> (last access: 12 December 2020), 2020.
- Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., Attinger, S., Schäfer, D., Schrön, M., and Samaniego, L.: Multi-scale and multivariate evaluation of water fluxes and states over European river basins, *J. Hydrometeorol.*, 17, 287–307, 2016.
- Ren, Z. and Li, M.: Errors and correction of precipitation measurements in China, *Adv. Atmos. Sci.*, 24, 449–458, <https://doi.org/10.1007/s00376-007-0449-3>, 2007.
- Robertson, A. W., Baethgen, W., Block, P., Lall, U., Sankarasubramanian, A., de Souza Filho, F. d. A., and Verbist, K. M.: Climate risk management for water in semi-arid regions, *Earth Perspectives*, 1, 12, <https://doi.org/10.1186/2194-6434-1-12>, 2014.
- Rojas, R., Feyen, L., and Watkiss, P.: Climate change and river floods in the European Union: Socio-economic consequences and the costs and benefits of adaptation, *Global Environ. Change*, 23, 1737–1751, <https://doi.org/10.1016/j.gloenvcha.2013.08.006>, 2013.
- Saadi, M., Oudin, L., and Ribstein, P.: Random forest ability in regionalizing hourly hydrological model parameters, *Water*, 11, 1540, <https://doi.org/10.3390/w11081540>, 2019.
- Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, *Water Resour. Res.*, 46, W05523, <https://doi.org/10.1029/2008WR007327>, 2010.
- Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G.: Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, *Hydrol. Earth Syst. Sci.*, 15, 2895–2911, <https://doi.org/10.5194/hess-15-2895-2011>, 2011.
- Sevruk, B., Ondrás, M., and Chvíla, B.: The WMO precipitation measurement intercomparisons, *Atmos. Res.*, 92, 376–380, <https://doi.org/10.1016/j.atmosres.2009.01.016>, 2009.
- Shafii, M. and Tolson, B. A.: Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives, *Water Resour. Res.*, 51, 3796–3814, 2015.
- Sharma, S., Siddique, R., Reed, S., Ahnert, P., Mendoza, P., and Mejia, A.: Relative effects of statistical preprocessing and postprocessing on a regional hydrological ensemble prediction system, *Hydrol. Earth Syst. Sci.*, 22, 1831–1849, <https://doi.org/10.5194/hess-22-1831-2018>, 2018.
- Silal, S. P., Little, F., Barnes, K. I., and White, L. J.: Predicting the impact of border control on malaria transmission: a simulated focal screen and treat campaign, *Malar J.*, 14, 268, <https://doi.org/10.1186/s12936-015-0776-2>, 2015.
- Singh, S. K., Bárdossy, A., Götzinger, J., and Sudheer, K.: Effect of spatial resolution on regionalization of hydrological model parameters, *Hydrol. Process.*, 26, 3499–3509, 2012.
- Sleziak, P., Szolgay, J., Hlavčová, K., Danko, M., and Parajka, J.: The effect of the snow weighting on the temporal stability of hydrologic model efficiency and parameters, *J. Hydrol.*, 583, 124639, <https://doi.org/10.1016/j.jhydrol.2020.124639>, 2020.
- Stisen, S. and Sandholt, I.: Evaluation of remote-sensing-based rainfall products through predictive capability in hydrological runoff modelling, *Hydrol. Process.*, 24, 879–891, <https://doi.org/10.1002/hyp.7529>, 2010.
- Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., and Hsu, K.-L.: A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons, *Rev. Geophys.*, 56, 79–107, <https://doi.org/10.1002/2017RG000574>, 2018.
- Swain, J. B. and Patra, K. C.: Streamflow estimation in ungauged catchments using regional flow duration curve: comparative study, *J. Hydrol. Eng.*, 22, 04017010, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001509](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001509), 2017.
- Széles, B., Parajka, J., Hogan, P., Silasari, R., Pavlin, L., Strauss, P., and Blöschl, G.: The added value of different data types for calibrating and testing a hydrologic model in a small catchment, *Water Resour. Res.*, 56, e2019WR026153, <https://doi.org/10.1029/2019WR026153>, 2020.
- Tarek, M., Brissette, F. P., and Arsenaault, R.: Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America, *Hydrol. Earth Syst. Sci.*, 24, 2527–2544, <https://doi.org/10.5194/hess-24-2527-2020>, 2020.
- Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., and De Roo, A.: Hydrological evaluation of satellite-based rainfall estimates over

- the Volta and Baro-Akobo Basin, *J. Hydrol.*, 499, 324–338, <https://doi.org/10.1016/j.jhydrol.2013.07.012>, 2013.
- Uhlenbrook, S., Seibert, J., Leibundgut, C., and Rohde, A.: Prediction uncertainty of conceptual rainfall-runoff models caused by problems to identify model parameters and structure, *Hydrol. Sci. J.*, 44, 779–797, 1999.
- Unduche, F., Tolossa, H., Senbeta, D., and Zhu, E.: Evaluation of four hydrological models for operational flood forecasting in a Canadian Prairie watershed, *Hydrol. Sci. J.*, 63, 1133–1149, 2018.
- Vandewiele, G. and Elias, A.: Monthly water balance of ungauged catchments obtained by geographical regionalization, *J. Hydrol.*, 170, 277–291, 1995.
- Vásquez, N., Cepeda, J., Gómez, T., Mendoza, P. A., Lagos, M., Boisier, J. P., Álvarez-Garretón, C., and Vargas, X.: Catchment-Scale Natural Water Balance in Chile, in: *Water Resources of Chile*, 189–208, Springer, Switzerland, 2021.
- Verbist, K., Robertson, A. W., Cornelis, W. M., and Gabriels, D.: Seasonal predictability of daily rainfall characteristics in central northern Chile for dry-land management, *J. Appl. Meteorol. Climatol.*, 49, 1938–1955, 2010.
- Vetter, T., Huang, S., Aich, V., Yang, T., Wang, X., Krysanova, V., and Hattermann, F.: Multi-model climate impact assessment and intercomparison for three large-scale river basins on three continents, *Earth Syst. Dynam.*, 6, 17–43, <https://doi.org/10.5194/esd-6-17-2015>, 2015.
- Viglione, A. and Parajka, J.: TUVmodel: Lumped/Semi-Distributed Hydrological Model for Education Purposes, available at: <https://CRAN.R-project.org/package=TUVmodel> (last access: 12 December 2020), r package version 1.1-1, 2020.
- Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S.: A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrological model parameters, *Water Resour. Res.*, 39, 1201, <https://doi.org/10.1029/2002WR001642>, 2003.
- Vrugt, J. A., ter Braak, C. J. F., Gupta, H. V., and Robinson, B. A.: Response to comment by Keith Beven on “Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?”, *Stoch. Env. Res. Risk A.*, 23, 1061–1062, <https://doi.org/10.1007/s00477-008-0284-9>, 2009.
- Wagener, T., Boyle, D. P., Lees, M. J., Wheatler, H. S., Gupta, H. V., and Sorooshian, S.: A framework for development and application of hydrological models, *Hydrol. Earth Syst. Sci.*, 5, 13–26, <https://doi.org/10.5194/hess-5-13-2001>, 2001.
- Wagener, T., Sivapalan, M., Troch, P., and Woods, R.: Catchment Classification and Hydrologic Similarity, *Geography Compass*, 1, 901–931, <https://doi.org/10.1111/j.1749-8198.2007.00039.x>, 2007.
- Woldemeskel, F. M., Sivakumar, B., and Sharma, A.: Merging gauge and satellite rainfall with specification of associated uncertainty across Australia, *J. Hydrol.*, 499, 167–176, <https://doi.org/10.1016/j.jhydrol.2013.06.039>, 2013.
- Xavier, A. C., King, C. W., and Scanlon, B. R.: Daily gridded meteorological variables in Brazil (1980–2013), *Int. J. Climatol.*, 36, 2644–2659, 2016.
- Xue, X., Hong, Y., Limaye, A. S., Gourley, J. J., Huffman, G. J., Khan, S. I., Dorji, C., and Chen, S.: Statistical and hydrological evaluation of TRMM-based Multi-satellite Precipitation Analysis over the Wangchu Basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins?, *J. Hydrol.*, 499, 91–99, 2013.
- Yang, Y., Pan, M., Beck, H. E., Fisher, C. K., Beighley, R. E., Kao, S.-C., Hong, Y., and Wood, E. F.: In quest of calibration density and consistency in hydrologic modeling: distributed parameter calibration against streamflow characteristics, *Water Resour. Res.*, 55, 7784–7803, 2019.
- Yapo, P. O., Gupta, H. V., and Sorooshian, S.: Multi-objective global optimization for hydrologic models, *J. Hydrol.*, 204, 83–97, [https://doi.org/10.1016/S0022-1694\(97\)00107-8](https://doi.org/10.1016/S0022-1694(97)00107-8), 1998.
- Young, A. R.: Stream flow simulation within UK ungauged catchments using a daily rainfall-runoff model, *J. Hydrol.*, 320, 155–172, 2006.
- Zambrano-Bigiarini, M.: hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series, r package version 0.4-0, <https://doi.org/10.5281/zenodo.839854>, 2020a.
- Zambrano-Bigiarini, M.: hydroTSM: Time Series Management, Analysis and Interpolation for Hydrological Modelling, r package version 0.6-0, <https://doi.org/10.5281/zenodo.839864>, 2020b.
- Zambrano-Bigiarini, M. and Baez-Villanueva, O.: Tutorial for using hydroPSO to calibrate TUVmodel, <https://doi.org/10.5281/zenodo.3772176>, 2020.
- Zambrano-Bigiarini, M. and Rojas, R.: A model-independent Particle Swarm Optimisation software for model calibration, *Environ. Model. Softw.*, 43, 5–25, <https://doi.org/10.1016/j.envsoft.2013.01.004>, 2013.
- Zambrano-Bigiarini, M., Nauditt, A., Birkel, C., Verbist, K., and Ribbe, L.: Temporal and spatial evaluation of satellite-based rainfall estimates across the complex topographical and climatic gradients of Chile, *Hydrol. Earth Syst. Sci.*, 21, 1295–1320, <https://doi.org/10.5194/hess-21-1295-2017>, 2017.
- Zambrano-Bigiarini, M., Baez-Villanueva, O. M., and Giraldo-Osorio, J.: RFmerge: Merging of Satellite Datasets with Ground Observations using Random Forests, r package version 0.1-10, <https://doi.org/10.5281/zenodo.3581515>, 2020.
- Zeilew, M. B. and Alfredeisen, K.: Transferability of hydrological model parameter spaces in the estimation of runoff in ungauged catchments, *Hydrol. Sci. J.*, 59, 1470–1490, 2014.
- Zessner, M., Schönhart, M., Parajka, J., Trautvetter, H., Mitter, H., Kirchner, M., Hepp, G., Blaschke, A. P., Strenn, B., and Schmid, E.: A novel integrated modelling framework to assess the impacts of climate and socio-economic drivers on land use and water quality, *Sci. Total Environ.*, 579, 1137–1151, 2017.
- Zhang, L., Li, X., Zheng, D., Zhang, K., Ma, Q., Zhao, Y., and Ge, Y.: Merging multiple satellite-based precipitation products and gauge observations using a novel double machine learning approach, *J. Hydrol.*, 594, 125969, <https://doi.org/10.1016/j.jhydrol.2021.125969>, 2021.
- Zhang, Y. and Chiew, F. H.: Relative merits of different methods for runoff predictions in ungauged catchments, *Water Resour. Res.*, 45, W07412, <https://doi.org/10.1029/2008WR007504>, 2009.
- Zhang, Y. and Wang, K.: Global precipitation system size, *Environ. Res. Lett.*, 16, 054005, <https://doi.org/10.1088/1748-9326/abf394>, 2021.
- Zhang, Y., Vaze, J., Chiew, F. H., and Li, M.: Comparing flow duration curve and rainfall-runoff modelling for predicting daily runoff in ungauged catchments, *J. Hydrol.*, 525, 72–86, 2015.

Zhao, Y., Feng, D., Yu, L., Wang, X., Chen, Y., Bai, Y., Hernández, H. J., Galleguillos, M., Estades, C., Biging, G. S., Radke, J. D., and Gong, P.: Detailed dynamic land cover mapping of Chile: Accuracy improvement by integrating multi-temporal data, *Remote Sens. Environ.*, 183, 170–185, <https://doi.org/10.1016/j.rse.2016.05.016>, 2016.

C.1 Supplement material

APPENDIX C.

Supplement of Hydrol. Earth Syst. Sci., 25, 5805–5837, 2021
<https://doi.org/10.5194/hess-25-5805-2021-supplement>
© Author(s) 2021. CC BY 4.0 License.



Hydrology and
Earth System
Sciences  Open Access

Supplement of

On the selection of precipitation products for the regionalisation of hydrological model parameters

Oscar M. Baez-Villanueva et al.

Correspondence to: Mauricio Zambrano-Bigiarini (mauricio.zambrano@ufrontera.cl)

The copyright of individual parts of the supplement might differ from the article licence.

APPENDIX C.

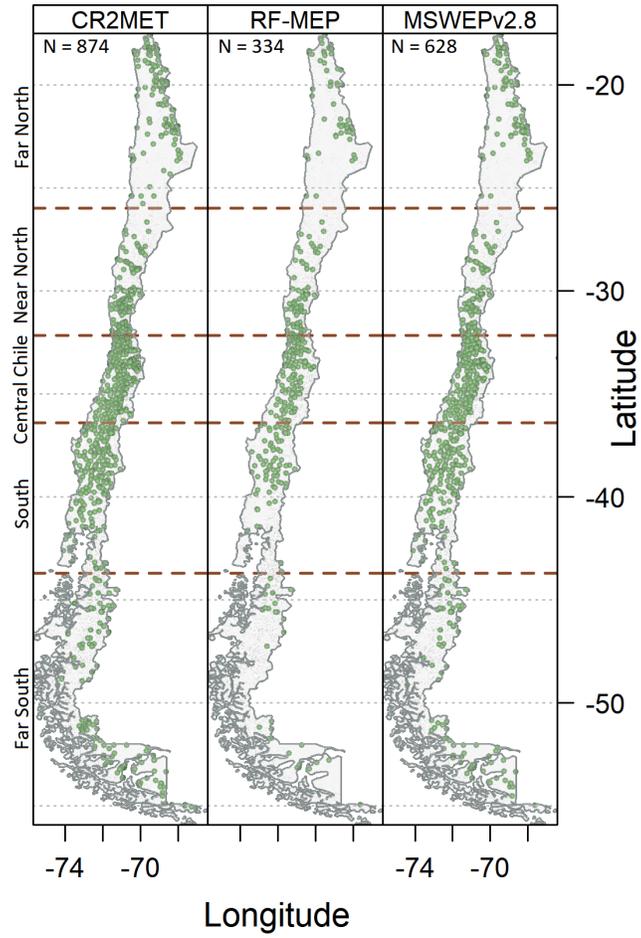


Figure S1: Rain gauges that each merged product used to construct their P estimates over Chile.

APPENDIX C.

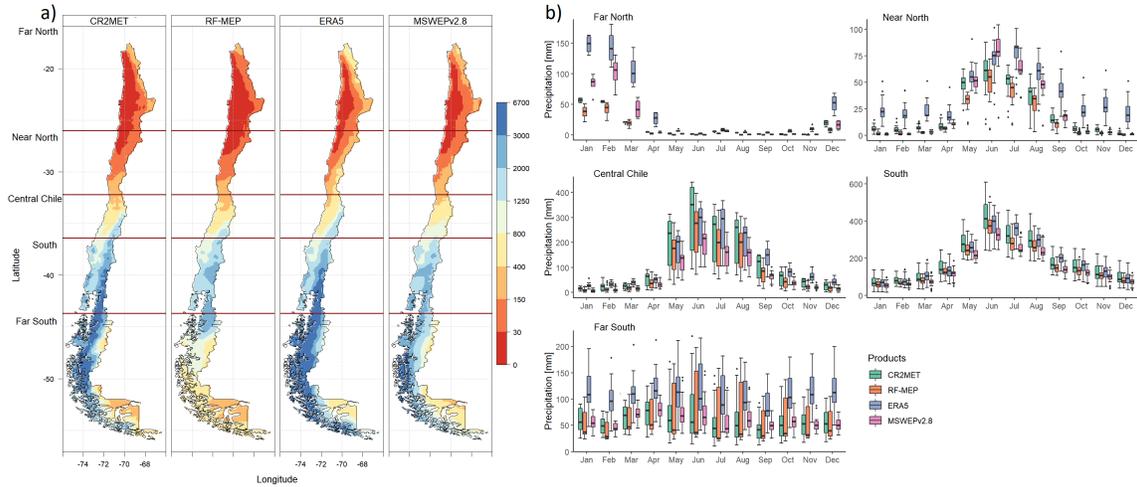


Figure S2: Comparison of P products over 2000–2014 (near-normal): *a*) mean annual P for each product resampled to a 0.05° spatial resolution using the nearest neighbour method. The dark red horizontal lines represent the limits of each major macroclimatic zone; and *b*) mean monthly P averaged over each catchment located within each macroclimatic zone (see Figure 1).

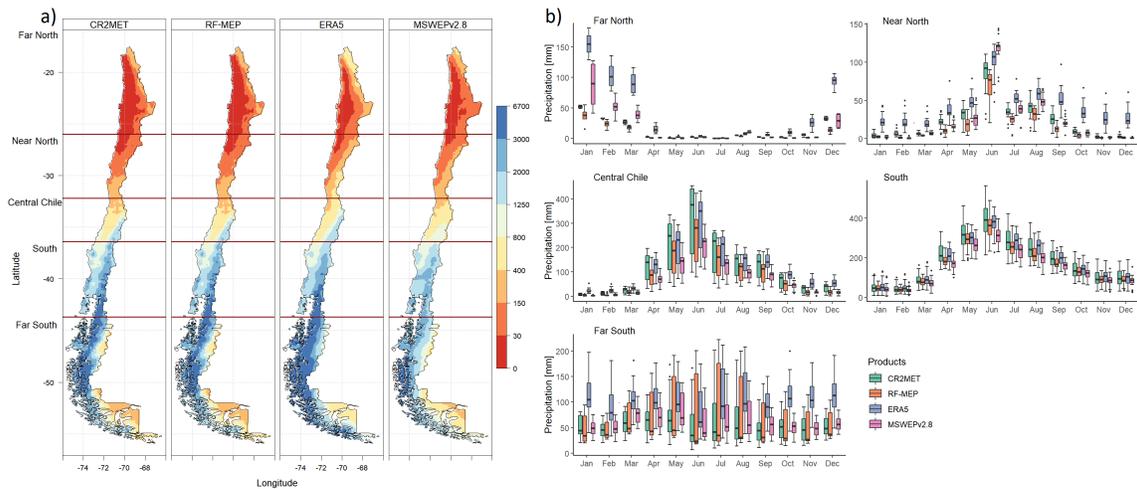


Figure S3: Comparison of P products over 1990–1999 (near-normal): *a*) mean annual P for each product resampled to a 0.05° spatial resolution using the nearest neighbour method. The dark red horizontal lines represent the limits of each major macroclimatic zone; and *b*) mean monthly P averaged over each catchment located within each macroclimatic zone (see Figure 1).

APPENDIX C.

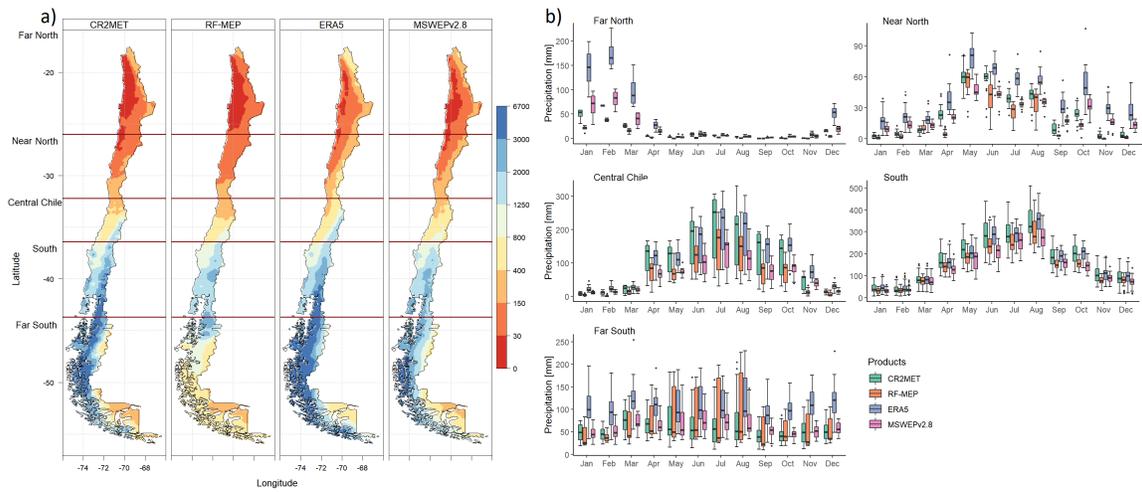


Figure S4: Comparison of P products over 2015–2018 (dry): *a*) mean annual P for each product resampled to a 0.05° spatial resolution using the nearest neighbour method. The dark red horizontal lines represent the limits of each major macroclimatic zone; and *b*) mean monthly P averaged over each catchment located within each macroclimatic zone (see Figure 1).

APPENDIX C.

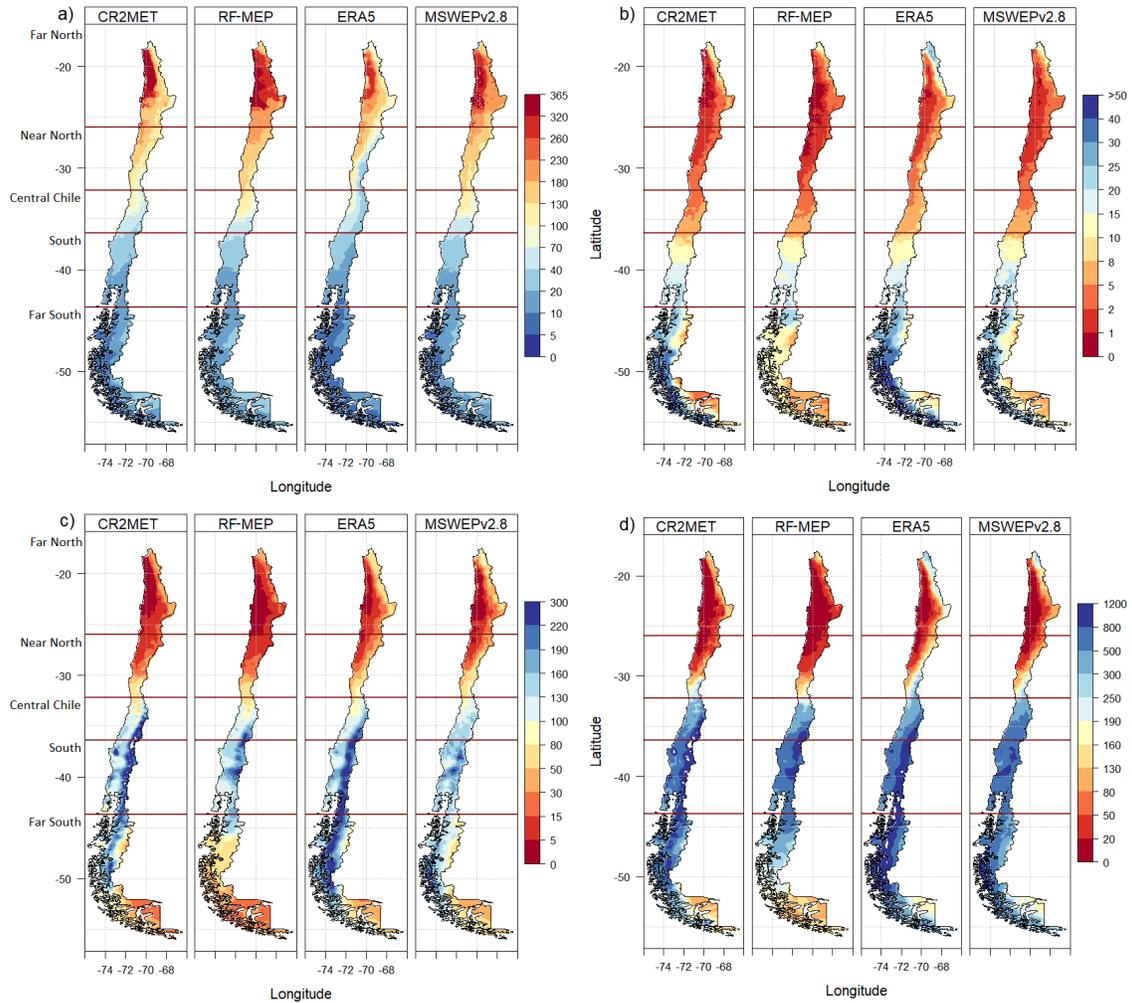


Figure S5: Median annual values of four Climdex indices over 2000–2014 (near-normal): *a*) number of consecutive dry days (CDD); *b*) number of consecutive wet days (CWD); *c*) maximum P over five consecutive days (RX5day); and *d*) annual P that is above the 95th percentile of P accumulated for events that are above the 95th percentile of the daily P for wet days (R95pTOT). The dark red horizontal lines represent the limits of each macroclimatic zone.

APPENDIX C.

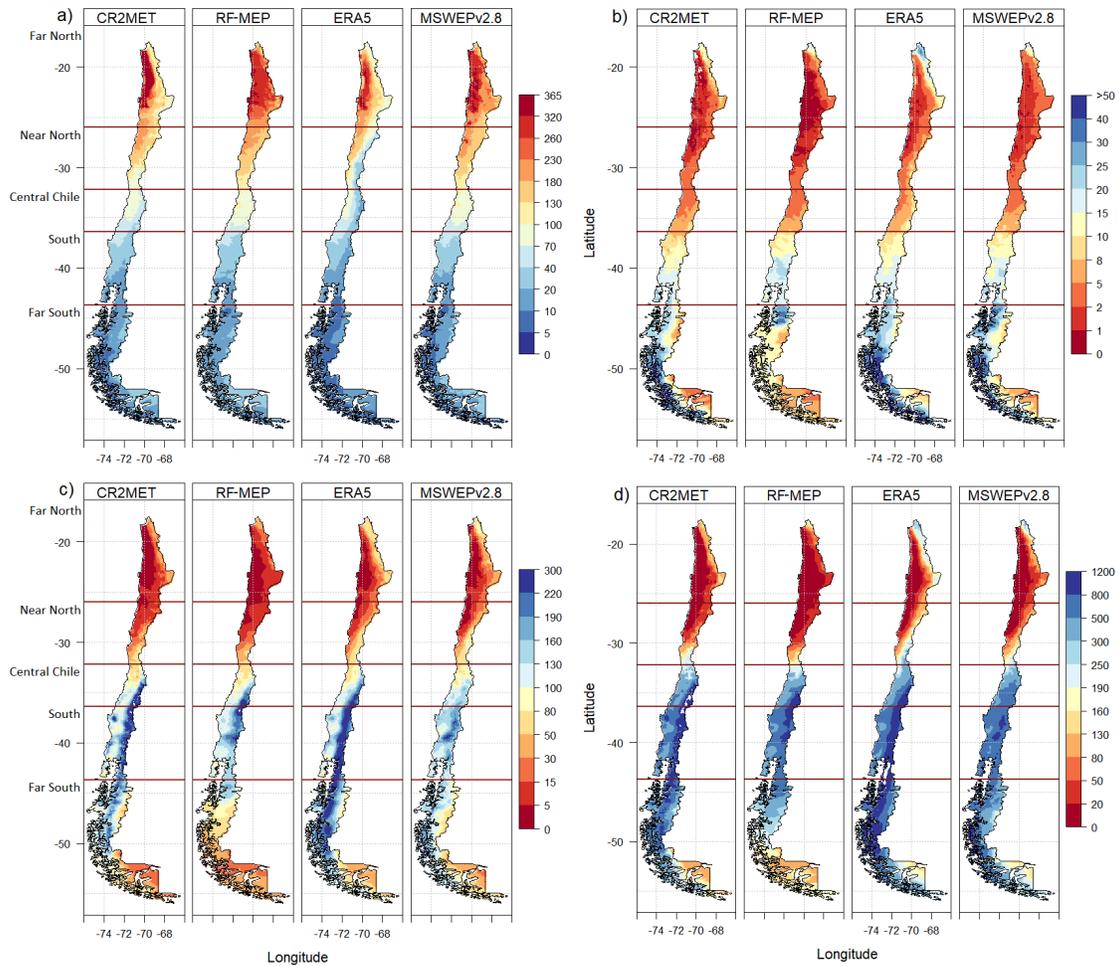


Figure S6: Median annual values of four Climdex indices over 1990–1999 (near-normal): *a*) number of consecutive dry days (CDD); *b*) number of consecutive wet days (CWD); *c*) maximum P over five consecutive days (RX5day); and *d*) annual P that is above the 95th percentile of P accumulated for events that are above the 95th percentile of the daily P for wet days (R95pTOT). The dark red horizontal lines represent the limits of each macroclimatic zone.

APPENDIX C.

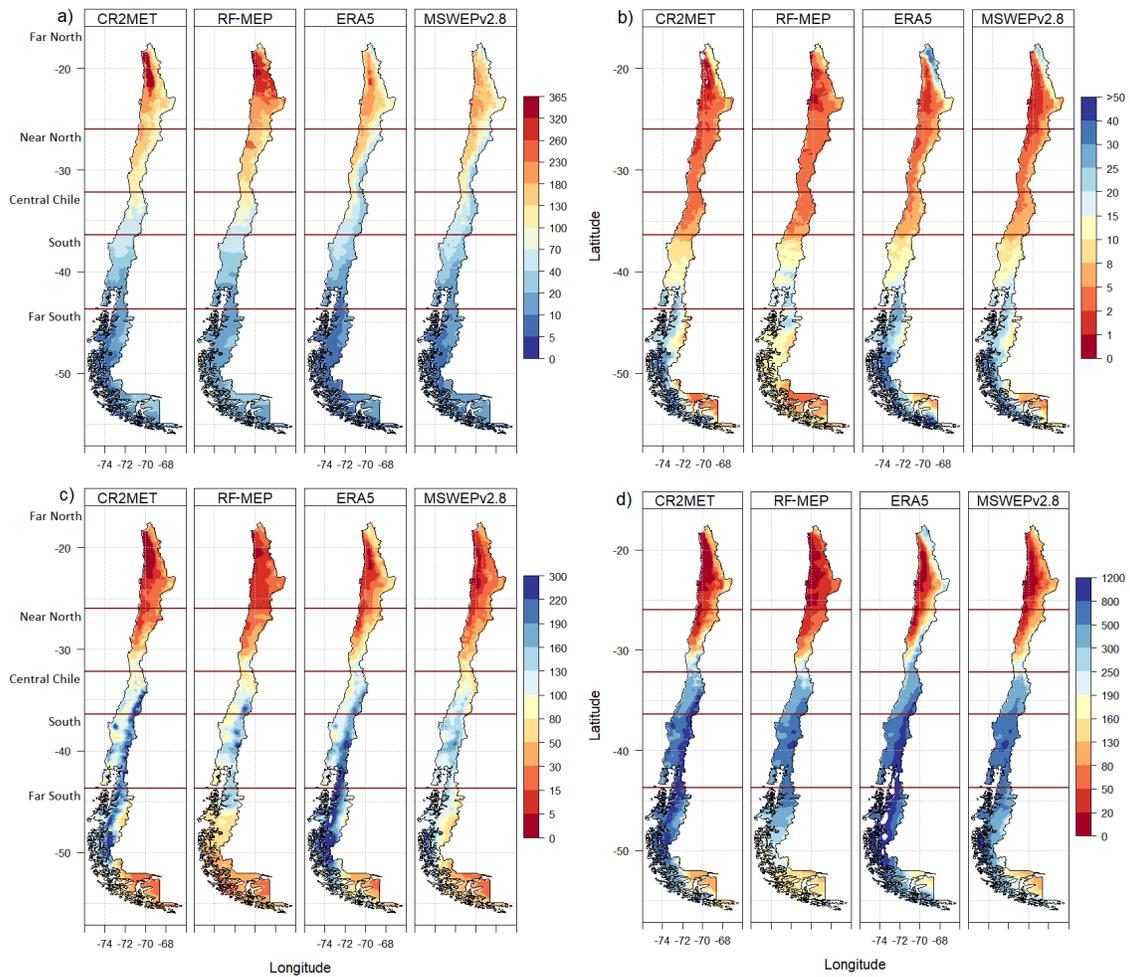


Figure S7: Median annual values of four Climdex indices over 2015–2018 (dry): *a*) number of consecutive dry days (CDD); *b*) number of consecutive wet days (CWD); *c*) maximum P over five consecutive days (RX5day); and *d*) annual P that is above the 95th percentile of P accumulated for events that are above the 95th percentile of the daily P for wet days (R95pTOT). The dark red horizontal lines represent the limits of each macroclimatic zone.

APPENDIX C.

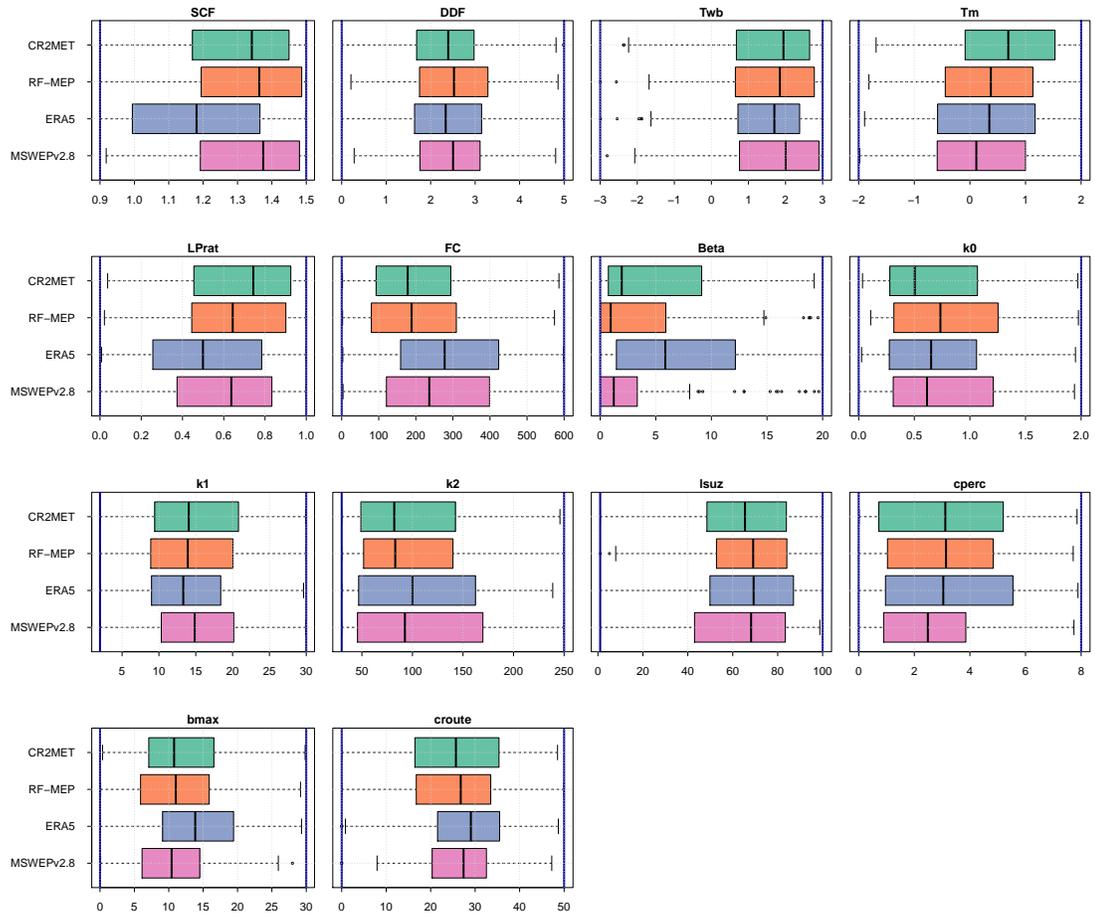


Figure S8: Model parameters obtained through calibration of the 100 selected catchments. The vertical blue lines indicate the upper and lower limits of the parameter ranges.

APPENDIX C.

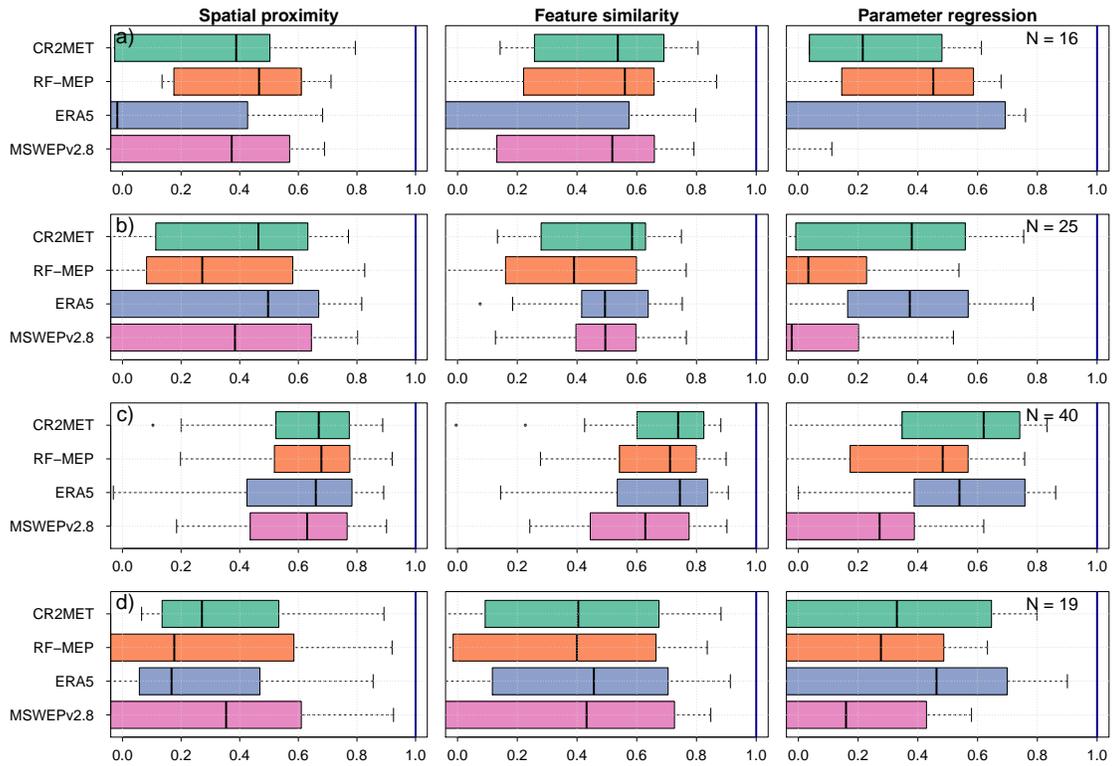


Figure S9: Performance of regionalisation methods for Verification 1 (1990–1999) according to the hydrological regime: a) snow-dominated; b) nivo-pluvial; c) pluvio-nival; and d) rain-dominated. N denotes the number of catchments per hydrological regime.

APPENDIX C.

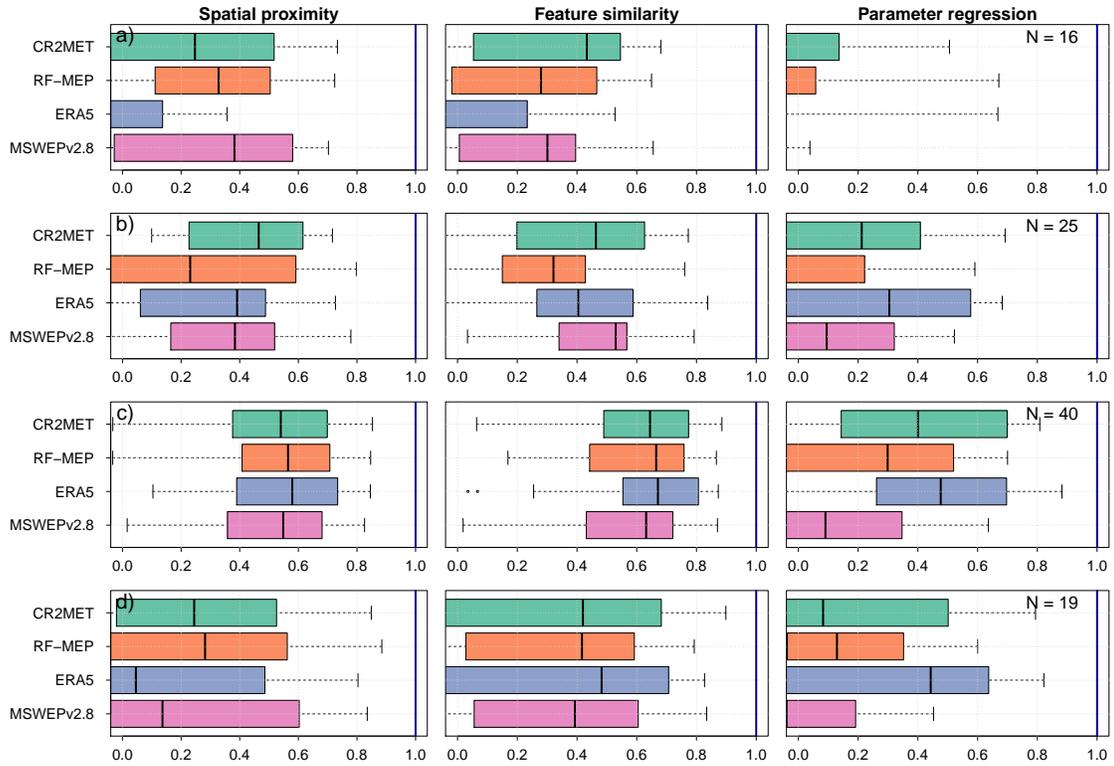


Figure S10: Performance of regionalisation methods for Verification 2 (2015–2018) according to the hydrological regime: *a*) snow-dominated; *b*) nivo-pluvial; *c*) pluvio-nival; and *d*) rain-dominated. *N* denotes the number of catchments per hydrological regime.

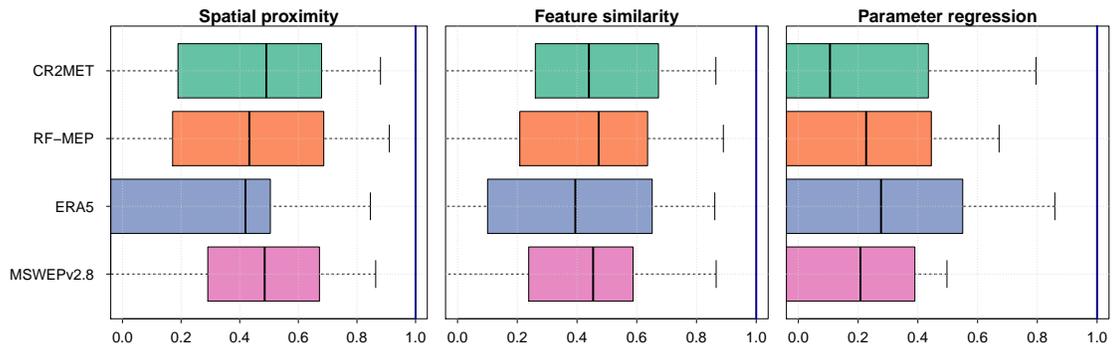


Figure S11: Regionalisation performance of the *P* products over the 25 smallest catchments (area < 353.1 km²).

APPENDIX C.

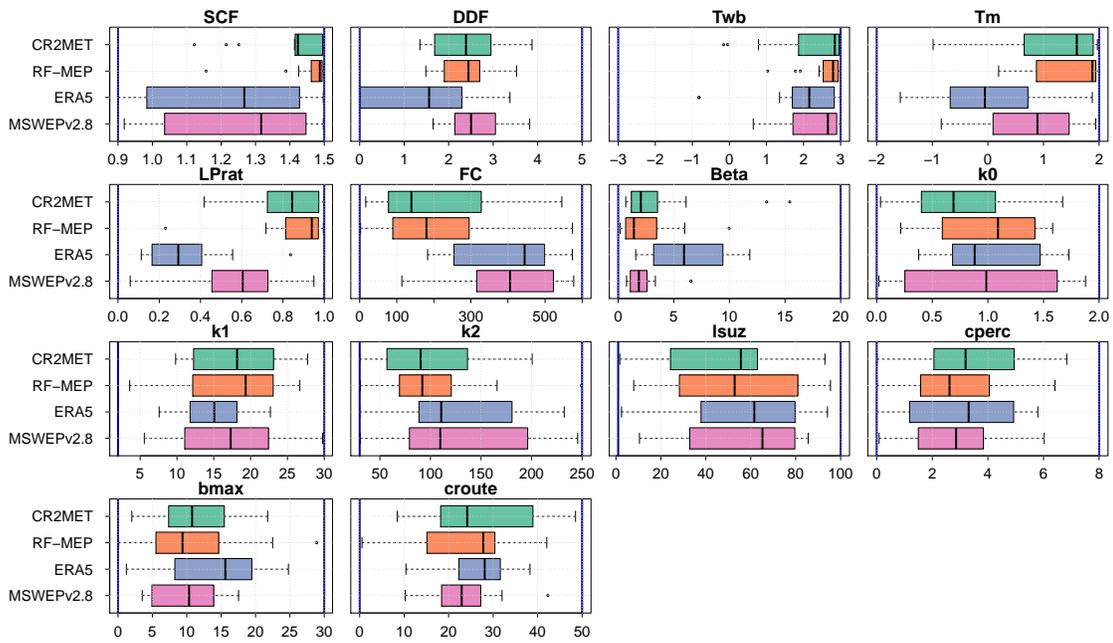


Figure S12: Model parameters obtained through calibration in snow-dominated catchments. The vertical blue lines indicate the upper and lower limits of the parameter ranges.

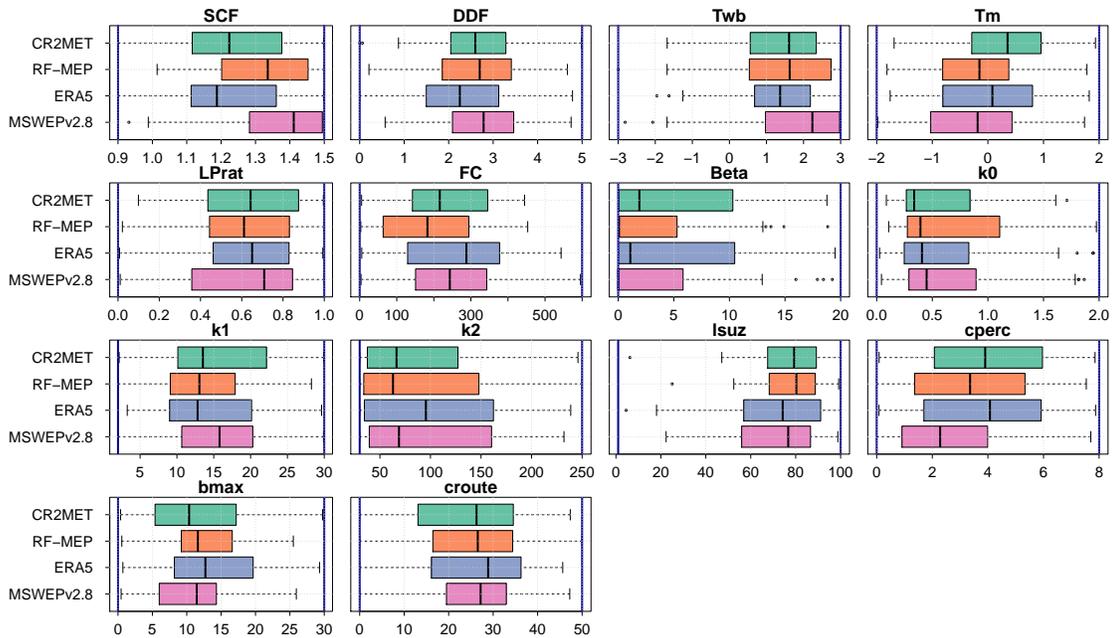


Figure S13: Model parameters obtained through calibration in pluvio-nival catchments. The vertical blue lines indicate the upper and lower limits of the parameter ranges.

APPENDIX C.

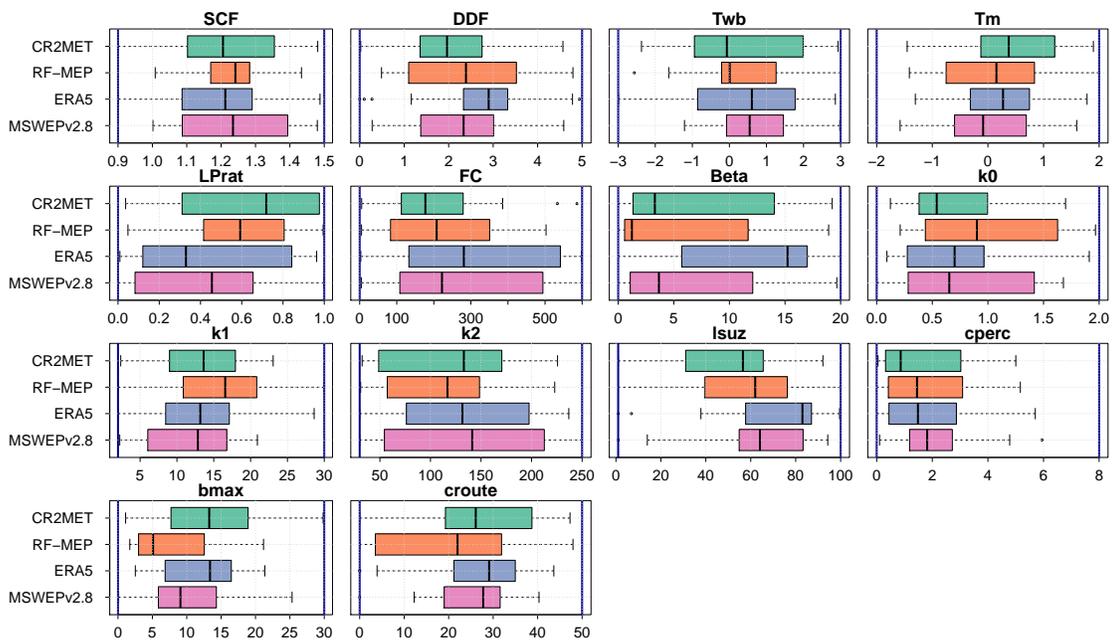


Figure S14: Model parameters obtained through calibration in rain-dominated catchments. The vertical blue lines indicate the upper and lower limits of the parameter ranges.

APPENDIX C.

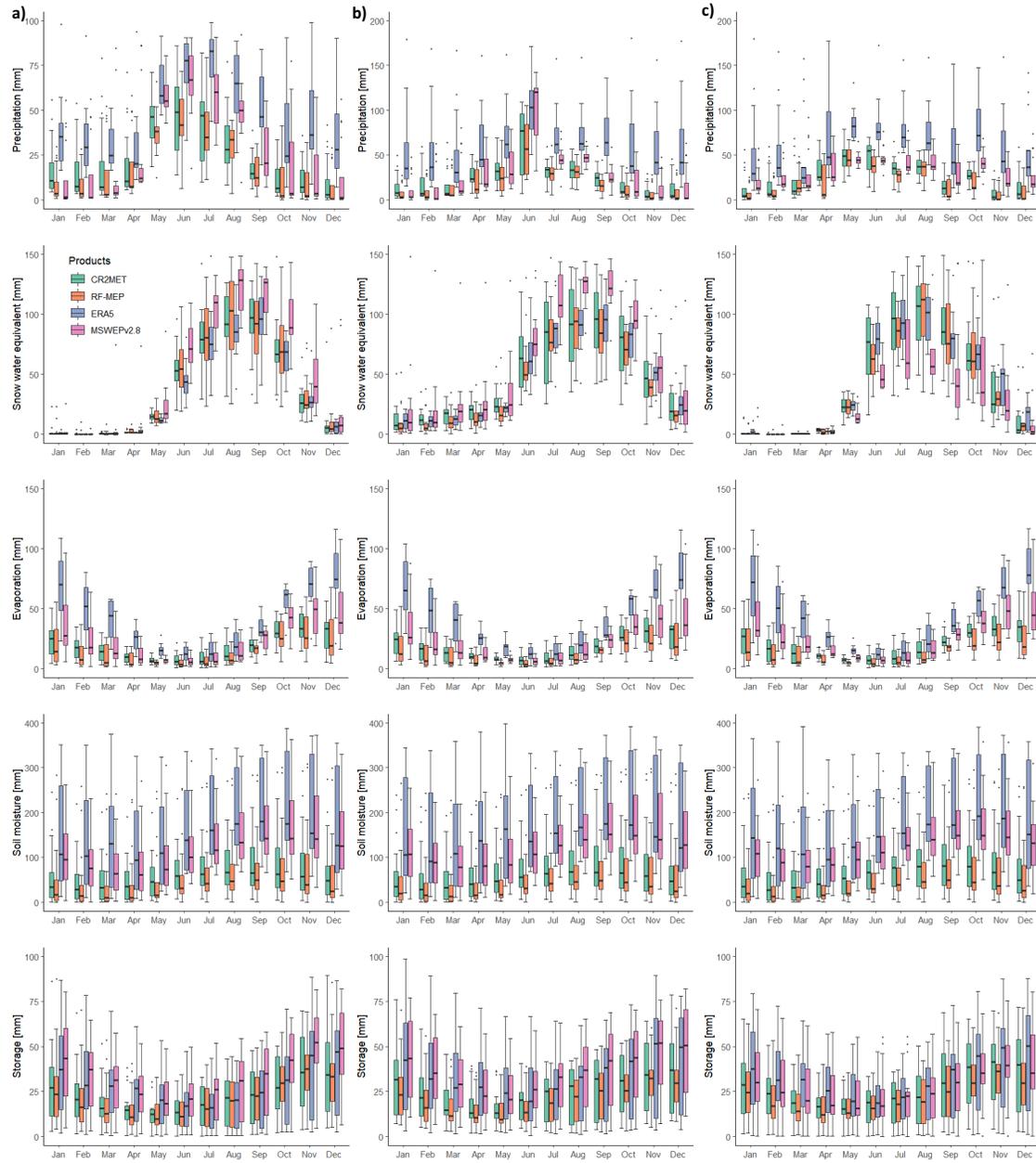


Figure S15: Mean monthly water balance components over snow-dominated catchments, obtained by forcing the TUW model with different P products for the: *a*) calibration (2000–2014); *b*) Verification 1 (1990–1999); and *c*) Verification 2 (2015–2018) periods. The mean monthly P was added for comparison purposes.

APPENDIX C.

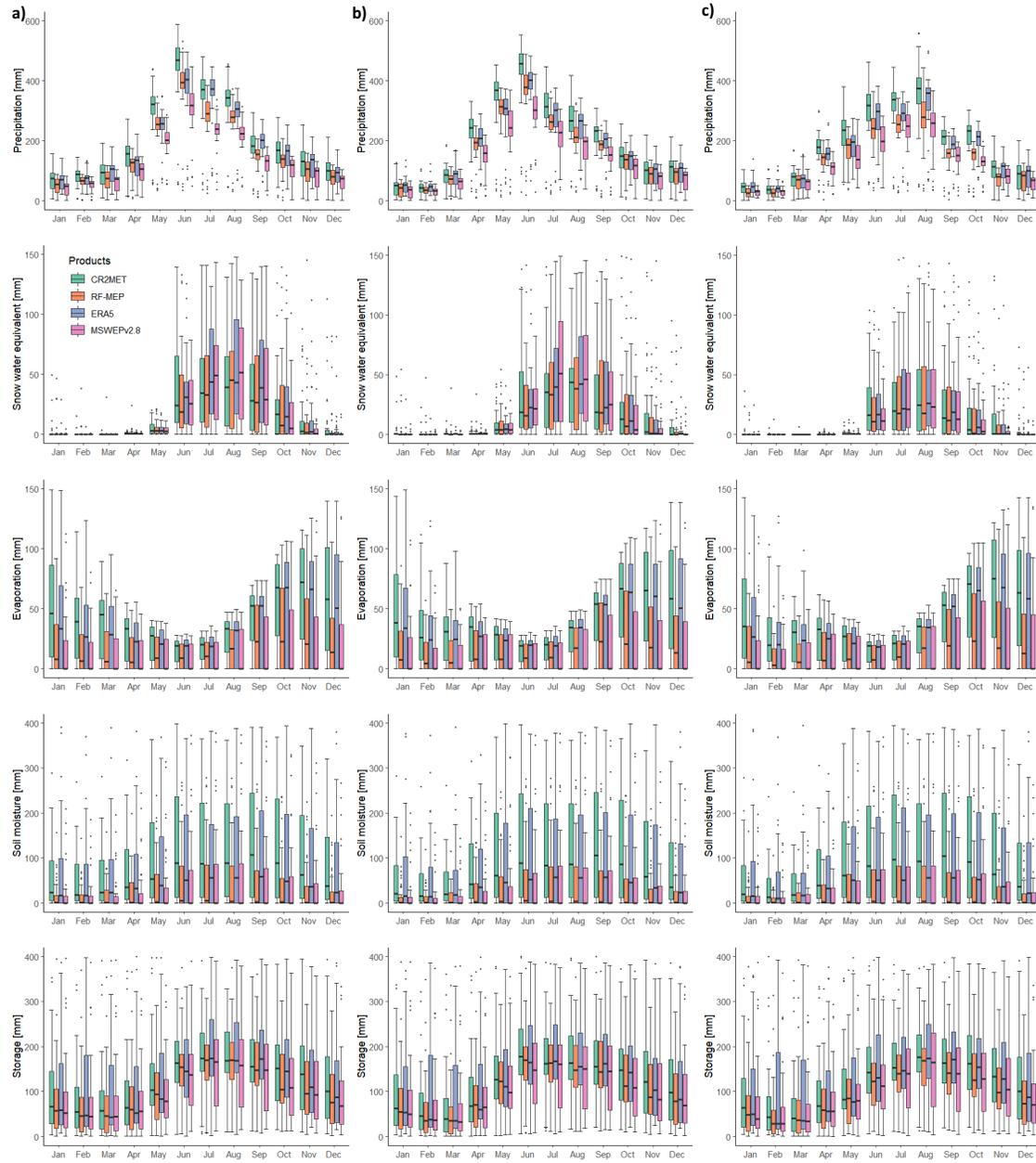


Figure S16: Mean monthly water balance components over pluvio-nival catchments, obtained by forcing the TUW model with different P products for the: *a*) calibration (2000–2014); *b*) Verification 1 (1990–1999); and *c*) Verification 2 (2015–2018) periods. The mean monthly P was added for comparison purposes.

APPENDIX C.

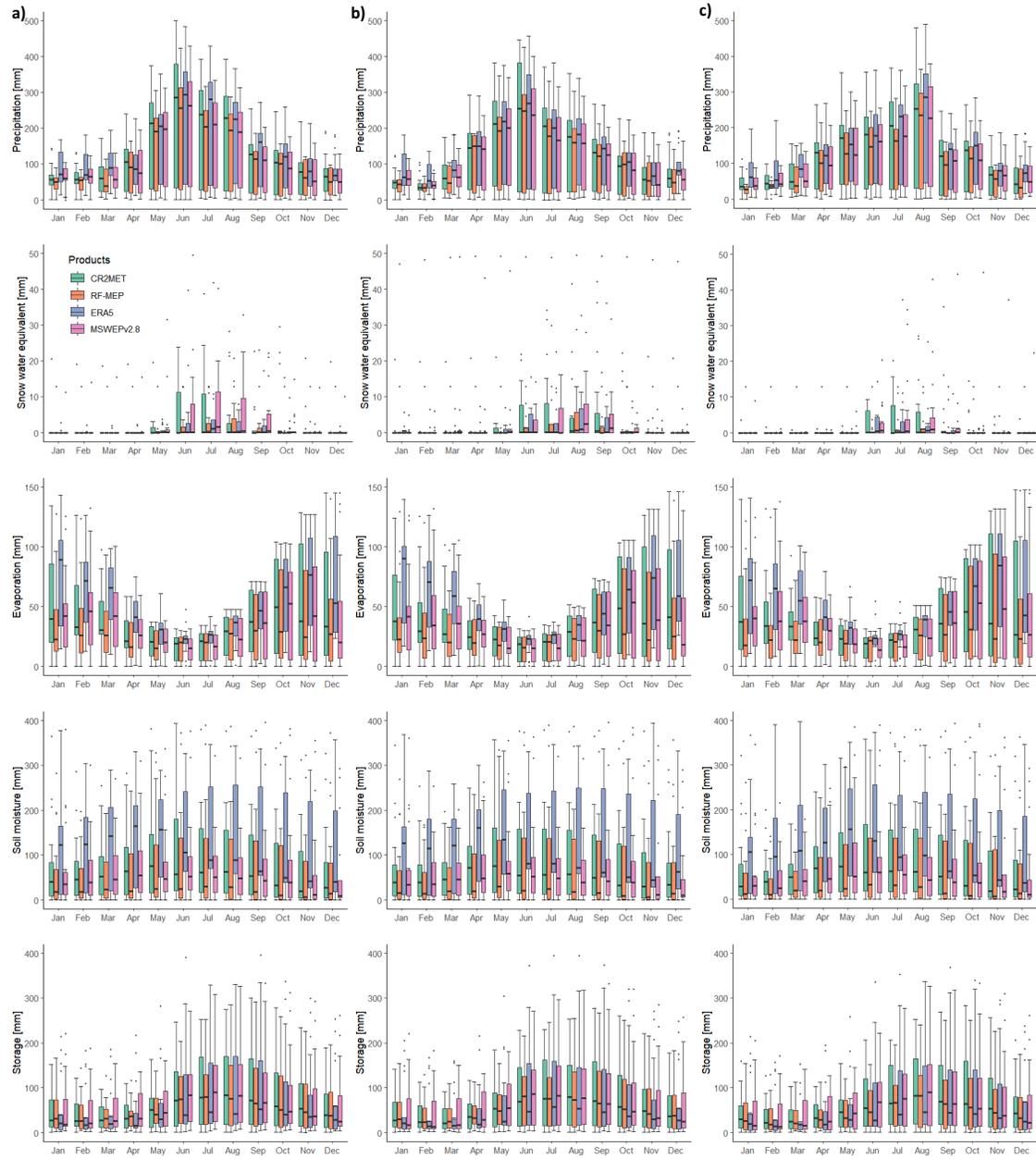


Figure S17: Mean monthly water balance components over rain-dominated catchments, obtained by forcing the TUW model with different P products for the: *a*) calibration (2000–2014); *b*) Verification 1 (1990–1999); and *c*) Verification 2 (2015–2018) periods. The mean monthly P was added for comparison purposes.

References

- Abdelaziz, R., Merkel, B.J., Zambrano-Bigiarini, M., Nair, S., 2019. Particle swarm optimization for the estimation of surface complexation constants with the geochemical model PHREEQC-3.1.2. *Geoscientific Model Development* 12, 167–177.
- Adam, J.C., Lettenmaier, D.P., 2003. Adjustment of global gridded precipitation for systematic bias. *Journal of Geophysical Research: Atmospheres* 108.
- Addor, N., Nearing, G., Prieto, C., Newman, A., Le Vine, N., Clark, M.P., 2018. A ranking of hydrological signatures based on their predictability in space. *Water Resources Research* 54, 8792–8812.
- Adhikary, S.K., Yilmaz, A.G., Muttil, N., 2015. Optimal design of rain gauge network in the Middle Yarra River catchment, Australia. *Hydrological Processes* 29, 2582–2599.
- Ahmad, S.K., Hossain, F., 2019. A generic data-driven technique for forecasting of reservoir inflow: Application for hydropower maximization. *Environmental Modelling & Software* 119, 147–165.
- Alvarez-Garreton, C., Mendoza, P.A., Boisier, J.P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Cortes, G., Garreaud, R., McPhee, J., et al., 2018. The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies-Chile dataset. *Hydrology and Earth System Sciences* 22, 5817–5846.
- Andres, L., Boateng, K., Borja-Vega, C., Thomas, E., 2018. A review of in-situ and remote sensing technologies to monitor water and sanitation interventions. *Water* 10, 756.
- Arsenault, R., Brissette, F.P., 2014. Continuous streamflow prediction in ungauged basins: The effects of equifinality and parameter set selection on uncertainty in regionalization approaches. *Water Resources Research* 50, 6135–6153.

REFERENCES

- Arshad, M., Ma, X., Yin, J., Ullah, W., Ali, G., Ullah, S., Liu, M., Shahzaman, M., Ullah, I., 2021. Evaluation of GPM-IMERG and TRMM-3B42 precipitation products over Pakistan. *Atmospheric Research* 249, 105341.
- Artan, G., Gadain, H., Smith, J.L., Asante, K., Bandaragoda, C.J., Verdin, J.P., 2007. Adequacy of satellite derived rainfall data for stream flow modeling. *Natural Hazards* 43, 167.
- Ashouri, H., Hsu, K.L., Sorooshian, S., Braithwaite, D.K., Knapp, K.R., Cecil, L.D., Nelson, B.R., Prat, O.P., 2015. PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bulletin of the American Meteorological Society* 96, 69–83.
- Astagneau, P.C., Thirel, G., Delaigue, O., Guillaume, J.H., Parajka, J., Brauer, C.C., Viglione, A., Buytaert, W., Beven, K.J., 2021. Hydrology modelling R packages—a unified analysis of models and practicalities from a user perspective. *Hydrology and Earth System Sciences* 25, 3937–3973.
- Athira, P., Sudheer, K., Cibir, R., Chaubey, I., 2016. Predictions in ungauged basins: An approach for regionalization of hydrological models considering the probability distribution of model parameters. *Stochastic environmental research and risk assessment* 30, 1131–1149.
- Atmoko, R., Riantini, R., Hasin, M., 2017. IoT real time data acquisition using MQTT protocol, in: *Journal of Physics: Conference Series*, IOP Publishing. p. 012003.
- Baez-Villanueva, O.M., Zambrano-Bigiarini, M., Beck, H.E., McNamara, I., Ribbe, L., Nauditt, A., Birkel, C., Verbist, K., Giraldo-Osorio, J.D., Thinh, N.X., 2020. RF-MEP: A novel random forest method for merging gridded precipitation products and ground-based measurements. *Remote Sensing of Environment* 239, 111606.
- Baez-Villanueva, O.M., Zambrano-Bigiarini, M., Mendoza, P.A., McNamara, I., Beck, H.E., Thurner, J., Nauditt, A., Ribbe, L., Thinh, N.X., 2021. On the selection of precipitation products for the regionalisation of hydrological model parameters. *Hydrology and Earth System Sciences* 25, 5805–5837.
- Baez-Villanueva, O.M., Zambrano-Bigiarini, M., Ribbe, L., Nauditt, A., Giraldo-Osorio, J.D., Thinh, N.X., 2018. Temporal and spatial evaluation of satellite rainfall estimates over different regions in Latin-America. *Atmospheric Research* 213, 34–50.

REFERENCES

- Bakker, K., Ritts, M., 2018. Smart earth: A meta-review and implications for environmental governance. *Global environmental change* 52, 201–211.
- Bambach, N., Bustos, E., Meza, F., Morales, D., Suarez, F., na, V., 2018. Aplicación de la metodología de actualización del balance hídrico nacional en las cuencas de la Macrozona Norte y Centro.
- Bao, Z., Zhang, J., Liu, J., Fu, G., Wang, G., He, R., Yan, X., Jin, J., Liu, H., 2012. Comparison of regionalization approaches based on regression and similarity for predictions in ungauged catchments under multiple hydro-climatic conditions. *Journal of Hydrology* 466, 37–46.
- Barrios-Ordóñez, J.E., Salinas-Rodríguez, S.A., Martínez-Pacheco, A., López-Pérez, M., Villón-Bracamonte, R.A., Rosales-Ángeles, F., Guerra Gilbert, A., Sánchez Navarro, R., 2015. National Water Reserves Program in Mexico. Experiences with environmental flows and the allocation of water for the environment. Technical Report. Technical Note BID-TN-864.
- Beal, C.D., Flynn, J., 2015. Toward the digital water age: Survey and case studies of Australian water utility smart-metering programs. *Utilities Policy* 32, 29–37.
- Beck, H.E., van Dijk, A.I., De Roo, A., Miralles, D.G., McVicar, T.R., Schellekens, J., Bruijnzeel, L.A., 2016. Global-scale regionalization of hydrologic model parameters. *Water Resources Research* 52, 3599–3622.
- Beck, H.E., van Dijk, A.I., Larraondo, P.R., McVicar, T.R., Pan, M., Dutra, E., Miralles, D.G., 2021a. MSWX: Global 3-hourly 0.1° bias-corrected meteorological data including near real-time updates and forecast ensembles. *Bulletin of the American Meteorological Society* , 1–55.
- Beck, H.E., van Dijk, A.I.J.M., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B., de Roo, A., 2017a. MSWEP: 3-hourly 0.25° global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences* 21, 589–615.
- Beck, H.E., Pan, M., Lin, P., Seibert, J., van Dijk, A.I., Wood, E.F., 2020a. Global fully distributed parameter regionalization based on observed streamflow from 4,229 headwater catchments. *Journal of Geophysical Research: Atmospheres* 125, e2019JD031485.
- Beck, H.E., Pan, M., Miralles, D.G., Reichle, R.H., Dorigo, W.A., Hahn, S., Sheffield, J., Karthikeyan, L., Balsamo, G., Parinussa, R.M., et al., 2021b.

REFERENCES

- Evaluation of 18 satellite-and model-based soil moisture products using in situ measurements from 826 sensors. *Hydrology and Earth System Sciences* 25, 17–40.
- Beck, H.E., Pan, M., Roy, T., Weedon, G.P., Pappenberger, F., Van Dijk, A.I., Huffman, G.J., Adler, R.F., Wood, E.F., 2019a. Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrology and Earth System Sciences* 23, 207–224.
- Beck, H.E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A.I.J.M., Weedon, G.P., Brocca, L., Pappenberger, F., Huffman, G.J., Wood, E.F., 2017b. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences* 21, 6201–6217.
- Beck, H.E., Wood, E.F., McVicar, T.R., Zambrano-Bigiarini, M., Alvarez-Garretón, C., Baez-Villanueva, O.M., Sheffield, J., Karger, D.N., 2020b. Bias correction of global high-resolution precipitation climatologies using streamflow observations from 9372 catchments. *Journal of Climate* 33, 1299–1315.
- Beck, H.E., Wood, E.F., Pan, M., Fisher, C.K., Miralles, D.G., Van Dijk, A.I., McVicar, T.R., Adler, R.F., 2019b. MSWEP V2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment. *Bulletin of the American Meteorological Society* 100, 473–500.
- Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific data* 5, 180214.
- Behrens, T., Schmidt, K., Viscarra Rossel, R.A., Gries, P., Scholten, T., MacMillan, R.A., 2018. Spatial modelling with Euclidean distance fields and machine learning. *European Journal of Soil Science* 69.
- Bello, O.M., Aina, Y.A., 2014. Satellite remote sensing as a tool in disaster management and sustainable development: towards a synergistic approach. *Procedia-Social and Behavioral Sciences* 120, 365–373.
- Bergström, S., 1976. Development and application of a conceptual runoff model for Scandinavian catchments.
- Bergström, S., 1995. The HBV model. *Computer models of watershed hydrology Highlands Ranch*, 443–476.

REFERENCES

- Biau, G., Scornet, E., 2016. A random forest guided tour. *TEST* 25, 197.
- Bisselink, B., Zambrano-Bigiarini, M., Burek, P., de Roo, A., 2016. Assessing the role of uncertain precipitation estimates on the robustness of hydrological model parameters under highly variable climate conditions. *Journal of Hydrology: Regional Studies* 8, 112–129.
- Bitew, M.M., Gebremichael, M., Ghebremichael, L.T., Bayissa, Y.A., 2012. Evaluation of high-resolution satellite rainfall products through streamflow simulation in a hydrological modeling of a small mountainous watershed in Ethiopia. *Journal of Hydrometeorology* 13, 338–350.
- Bivand, R., Keitt, T., Rowlingson, B., 2020. *rgdal: Bindings for the 'Geospatial' Data Abstraction Library*. R package version 1.5-12.
- Bivand, R., Rundel, C., 2020. *rgeos: Interface to Geometry Engine - Open Source ('GEOS')*. R package version 0.5-3.
- Blöschl, G., Blöschl, G., Sivapalan, M., Wagener, T., Savenije, H., Viglione, A., 2013. *Runoff prediction in ungauged basins: Synthesis across processes, places and scales*. Cambridge University Press.
- Blöschl, G., Gaál, L., Hall, J., Kiss, A., Komma, J., Nester, T., Parajka, J., Perdigão, R.A., Plavcová, L., Rogger, M., et al., 2015. Increasing river floods: Fiction or reality? *Wiley Interdisciplinary Reviews: Water* 2, 329–344.
- Blumenstock, J., 2018. Don't forget people in the use of big data for development. *Nature Publishing Group* 561, 170–172.
- Boisier, J.P., Alvarez-Garretón, C., Cepeda, J., Osses, A., Vásquez, N., Rondanelli, R., 2018. CR2MET: A high-resolution precipitation and temperature dataset for hydroclimatic research in Chile, p. 19739.
- Boisier, J.P., Rondanelli, R., Garreaud, R.D., Muñoz, F., 2016. Anthropogenic and natural contributions to the Southeast Pacific precipitation decline and recent megadrought in Central Chile. *Geophysical Research Letters* 43, 413–421.
- Bond, N.R., Burrows, R.M., Kennard, M.J., Bunn, S.E., 2019. Water scarcity as a driver of multiple stressor effects, in: *Multiple stressors in river ecosystems*. Elsevier, pp. 111–129.

REFERENCES

- Brauer, C.C., Torfs, P.J.J.F., Teuling, A.J., Uijlenhoet, R., 2014. The wageningen lowland runoff simulator (walrus): application to the hupsel brook catchment and the cabauw polder. *Hydrology and Earth System Sciences* 18, 4007–4028.
- Brauman, K.A., Richter, B.D., Postel, S., Malsy, M., Flörke, M., Blum, J.D., 2016. Water depletion: An improved metric for incorporating seasonal and dry-year water scarcity into water risk assessments. *Elementa: Science of the Anthropocene* 4.
- Breiman, L., 2001. Random Forests. *Machine Learning* 45, 5–32.
- Brocca, L., Ciabatta, L., Massari, C., Moramarco, T., Hahn, S., Hasenauer, S., Kidd, R., Dorigo, W., Wagner, W., Levizzani, V., 2014. Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data. *Journal of Geophysical Research: Atmospheres* 119, 5128–5141.
- Brown, M., Racoviteanu, A., Tarboton, D.G., Gupta, A.S., Nigro, J., Policelli, F., Habib, S., Tokay, M., Shrestha, M., Bajracharya, S., et al., 2014. An integrated modeling system for estimating glacier and snow melt driven streamflow from remote sensing and earth system data products in the Himalayas. *Journal of Hydrology* 519, 1859–1869.
- Brunetti, M., Melillo, M., Peruccacci, S., Ciabatta, L., Brocca, L., 2018. How far are we from the use of satellite rainfall products in landslide forecasting? *Remote sensing of environment* 210, 65–75.
- Calera, A., Campos, I., Osann, A., D’Urso, G., Menenti, M., 2017. Remote sensing for crop water management: From ET modelling to services for the end users. *Sensors* 17, 1104.
- Carrillo, G., Troch, P.A., Sivapalan, M., Wagener, T., Harman, C., Sawicz, K., 2011. Catchment classification: Hydrological analysis of catchment behavior through process-based modeling along a climate gradient. *Hydrology and Earth System Sciences* 15, 3411–3430.
- Ceola, S., Arheimer, B., Baratti, E., Blöschl, G., Capell, R., Castellarin, A., Freer, J., Han, D., Hrachowitz, M., Hundecha, Y., et al., 2015. Virtual laboratories: New opportunities for collaborative water science. *Hydrology and Earth System Sciences* 19, 2101–2117.
- Chen, C., Hu, B., Li, Y., 2021a. Easy-to-use spatial random-forest-based downscaling-calibration method for producing precipitation data with high resolution and high accuracy. *Hydrology and Earth System Sciences* 25, 5667–5682.

REFERENCES

- Chen, M., Shi, W., Xie, P., Silva, V.B., Kousky, V.E., Wayne Higgins, R., Janowiak, J.E., 2008. Assessing objective techniques for gauge-based analyses of global daily precipitation. *Journal of Geophysical Research: Atmospheres* 113.
- Chen, X., Liang, S., He, L., Yang, Y., Yin, C., 2021b. A temporally consistent 8-day 0.05 degree gap-free snow cover extent dataset over the northern hemisphere for the period 1981–2019. *Earth System Science Data Discussions* 1, 1–30.
- Chen, Y., Xia, J., Liang, S., Feng, J., Fisher, J.B., Li, X., Li, X., Liu, S., Ma, Z., Miyata, A., et al., 2014. Comparison of satellite-based evapotranspiration models over terrestrial ecosystems in China. *Remote Sensing of Environment* 140, 279–293.
- Cheong, S.M., Choi, G.W., Lee, H.S., 2016. Barriers and solutions to smart water grid development. *Environmental management* 57, 509–515.
- Ciabatta, L., Brocca, L., Massari, C., Moramarco, T., Gabellani, S., Puca, S., Wagner, W., 2016. Rainfall-runoff modelling by using SM2RAIN-derived and state-of-the-art satellite rainfall products over Italy. *International journal of applied earth observation and geoinformation* 48, 163–173.
- Ciabatta, L., Brocca, L., Massari, C., Moramarco, T., Puca, S., Rinollo, A., Gabellani, S., Wagner, W., 2015. Integration of satellite soil moisture and rainfall observations over the Italian territory. *Journal of Hydrometeorology* 16, 1341–1355.
- Clark, M.P., Slater, A.G., Rupp, D.E., Woods, R.A., Vrugt, J.A., Gupta, H.V., Wagener, T., Hay, L.E., 2008. Framework for understanding structural errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research* 44.
- Clerc, M., 2011a. From theory to practice in particle swarm optimization, in: *Handbook of Swarm Intelligence*. Springer, pp. 3–36.
- Clerc, M., 2011b. Standard particle swarm optimisation from 2006 to 2011. *Particle Swarm Central* 253.
- Cominola, A., Giuliani, M., Piga, D., Castelletti, A., Rizzoli, A.E., 2015. Benefits and challenges of using smart meters for advancing residential water demand modeling and management: A review. *Environmental Modelling & Software* 72, 198–214.

REFERENCES

- Connor, R., Renata, A., Ortigara, C., Koncagül, E., Uhlenbrook, S., Lamizana-Diallo, B.M., Zadeh, S.M., Qadir, M., Kjellén, M., Sjödin, J., et al., 2017. The united nations world water development report 2017. Wastewater: The untapped resource. The United Nations World Water Development Report .
- Corral, C., Berenguer, M., Sempere-Torres, D., Poletti, L., Silvestro, F., Rebora, N., 2019. Comparison of two early warning systems for regional flash flood hazard forecasting. *Journal of Hydrology* 572, 603–619.
- Cosgrove, W.J., Loucks, D.P., 2015. Water management: Current and future challenges and research directions. *Water Resources Research* 51, 4823–4839.
- Coughlan, E., van den Hurk, B., van Aalst, M.K., Amuron, I., Bamanya, D., Hauser, T., Jongma, B., Lopez, A., Mason, S., Mendler de Suarez, J., Pappenberger, F., Rueth, A., Stephens, E., Suarez, P., Wagemaker, J., Zsoter, E., 2016. Action-based flood forecasting for triggering humanitarian action. *Hydrology and Earth System Sciences* 20, 3549–3560.
- CR2, 2022. Center for Climate and Resilience Research: Base de Datos, <https://www.cr2.cl/bases-de-datos/>, last accessed on 2022-03-14.
- Crow, W., van Den Berg, M., Huffman, G., Pellarin, T., 2011. Correcting rainfall using satellite-based surface soil moisture retrievals: The Soil Moisture Analysis Rainfall Tool (SMART). *Water Resources Research* 47.
- Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Källberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.N., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* 137, 553–597.
- Dembélé, M., Ceperley, N., Zwart, S.J., Salvatore, E., Mariethoz, G., Schaefli, B., 2020a. Potential of satellite and reanalysis evaporation datasets for hydrological modelling under various model calibration strategies. *Advances in Water Resources* 143, 103667.
- Dembélé, M., Hrachowitz, M., Savenije, H.H., Mariéthoz, G., Schaefli, B., 2020b. Improving the predictive skill of a distributed hydrological model by

REFERENCES

- calibration on spatial patterns with multiple satellite data sets. *Water resources research* 56.
- Derin, Y., Yilmaz, K.K., 2014. Evaluation of multiple satellite-based precipitation products over complex topography. *Journal of Hydrometeorology* 15, 1498–1516.
- Díaz-Uriarte, R., Alvarez de Andrés, S., 2006. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics* 7, 3.
- Ding, J., Wallner, M., Müller, H., Haberlandt, U., 2016. Estimation of instantaneous peak flows from maximum mean daily flows using the HBV hydrological model. *Hydrological Processes* 30, 1431–1448.
- Dinku, T., Ceccato, P., Connor, S.J., 2011. Challenges of satellite rainfall estimation over mountainous and arid parts of east Africa. *International journal of remote sensing* 32, 5965–5979.
- Dinku, T., Ceccato, P., Grover-Kopec, E., Lemma, M., Connor, S., Ropelewski, C., 2007. Validation of satellite rainfall products over East Africa’s complex topography. *International Journal of Remote Sensing* 28, 1503–1526.
- Dinku, T., Ruiz, F., Connor, S.J., Ceccato, P., 2010. Validation and inter-comparison of satellite rainfall estimates over Colombia. *Journal of Applied Meteorology and Climatology* 49, 1004–1014.
- Doolan, J., Keary, J., Bouilly, L., Langford, J., Claydon, G., Slatyer, T., 2016. *The Australian water reform journey: An overview of three decades of policy, management and institutional transformation.* eWater Limited.
- Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., et al., 2017. ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sensing of Environment* 203, 185–215.
- Dos Reis, J.B.C., Rennó, C.D., Lopes, E.S.S., 2017. Validation of satellite rainfall products over a mountainous watershed in a humid subtropical climate region of Brazil. *Remote Sensing* 9, 1240.
- Eberhart, R., Kennedy, J., 1995. A new optimizer using particle swarm theory, in: *Micro Machine and Human Science, 1995. MHS '95., Proceedings of the Sixth International Symposium on*, pp. 39–43.

REFERENCES

- ECMWF, 2021. Climate reanalysis. <https://www.ecmwf.int/en/research/climate-reanalysis>. Accessed: 2021-10-08.
- Eggimann, S., Mutzner, L., Wani, O., Schneider, M.Y., Spuhler, D., Moy de Vitry, M., Beutler, P., Maurer, M., 2017. The potential of knowing more: A review of data-driven urban water management. *Environmental science & technology* 51, 2538–2553.
- Elsner, M.M., Gangopadhyay, S., Pruitt, T., Brekke, L.D., Mizukami, N., Clark, M.P., 2014. How does the choice of distributed meteorological data affect hydrologic model calibration and streamflow simulations? *Journal of Hydrometeorology* 15, 1384–1403.
- Entekhabi, D., Njoku, E.G., O’Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., Johnson, J., et al., 2010. The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE* 98, 704–716.
- Espinoza Villar, J.C., Ronchail, J., Guyot, J.L., Cochonneau, G., Naziano, F., Lavado, W., De Oliveira, E., Pombosa, R., Vauchel, P., 2009. Spatio-temporal rainfall variability in the Amazon Basin countries (Brazil, Peru, Bolivia, Colombia, and Ecuador). *International Journal of Climatology: A Journal of the Royal Meteorological Society* 29, 1574–1594.
- Fan, Z., Li, W., Jiang, Q., Sun, W., Wen, J., Gao, J., 2021. A comparative study of four merging approaches for regional precipitation estimation. *IEEE Access* 9, 33625–33637.
- Fay, M., Andres, L.A., Fox, C., Narloch, U., Slawson, M., 2017. Rethinking infrastructure in Latin America and the Caribbean: Spending better to achieve more. World Bank Publications.
- Fernandez, W., Vogel, R., Sankarasubramanian, A., 2000. Regional calibration of a watershed model. *Hydrological sciences journal* 45, 689–707.
- Flörke, M., Schneider, C., McDonald, R.I., 2018. Water competition between cities and agriculture driven by climate change and urban growth. *Nature Sustainability* 1, 51–58.
- Fu, Y., Xia, J., Yuan, W., Xu, B., Wu, X., Chen, Y., Zhang, H., 2016. Assessment of multiple precipitation products over major river basins of China. *Theoretical and applied climatology* 123, 11–22.

REFERENCES

- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data* 2, 150066.
- Gallant, J., Read, A., Dowling, T., 2012. Removal of tree offsets from SRTM and other digital surface models. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 39, 275–280.
- Gao, Y.C., Liu, M.F., 2013. Evaluation of high-resolution satellite precipitation products using rain gauge observations over the Tibetan Plateau. *Hydrology and Earth System Sciences* 17, 837–849.
- Garambois, P.A., Roux, H., Larnier, K., Labat, D., Dartus, D., 2015. Parameter regionalization for a process-oriented distributed model dedicated to flash floods. *Journal of Hydrology* 525, 383–399.
- Garcia, F., Folton, N., Oudin, L., 2017. Which objective function to calibrate rainfall–runoff models for low-flow index simulations? *Hydrological sciences journal* 62, 1149–1166.
- García, L., Rodríguez, D., Wijnen, M., Pakulski, I., 2016. Earth observation for water resources management: Current use and future opportunities for the water sector. World Bank Publications.
- Garreaud, R.D., Alvarez-Garreton, C., Barichivich, J., Pablo Boisier, J., Christie, D., Galleguillos, M., LeQuesne, C., McPhee, J., Zambrano-Bigiarini, M., 2017. The 2010–2015 megadrought in Central Chile: Impacts on regional hydroclimate and vegetation. *Hydrology and Earth System Sciences* 21, 6307–6327.
- Garreaud, R.D., Boisier, J.P., Rondanelli, R., Montecinos, A., Sepúlveda, H.H., Veloso-Aguila, D., 2020. The Central Chile Mega Drought (2010–2018): A climate dynamics perspective. *International Journal of Climatology* 40, 421–439.
- Gerlitz, L., Vorogushyn, S., Gafurov, A., 2020. Climate informed seasonal forecast of water availability in Central Asia: State-of-the-art and decision making context. *Water Security* 10, 100061.
- Grafton, R.Q., Pittock, J., Davis, R., Williams, J., Fu, G., Warburton, M., Udall, B., McKenzie, R., Yu, X., Che, N., et al., 2013. Global insights into

REFERENCES

- water resources, climate change and governance. *Nature Climate Change* 3, 315–321.
- Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology* 377, 80–91.
- Hall, D.K., Riggs, G.A., 2007. Accuracy assessment of the MODIS snow products. *Hydrological Processes: An International Journal* 21, 1534–1547.
- Hann, H., Nauditt, A., Zambrano-Bigiarini, M., Thurner, J., McNamara, I., Ribbe, L., 2021. Combining satellite-based rainfall data with rainfall-runoff modelling to simulate low flows in a Southern Andean catchment. *Journal of Natural Resources and Development* 11, 1–19.
- Hargreaves, G.H., Samani, Z.A., 1985. Reference crop evapotranspiration from ambient air temperature. American Society of Agricultural Engineers (fiche no. 85-2517). (Microfiche collection)(USA). no. fiche no. 85-2517.
- Hellegers, P.J., Soppe, R., Perry, C.J., Bastiaanssen, W.G., 2010. Remote sensing and economic indicators for supporting water resources management decisions. *Water resources management* 24, 2419–2436.
- Hengl, T., 2019. GSIF: Global Soil Information Facilities. R package version 0.5-5.
- Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B., Gräler, B., 2018. Random Forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ* 6, e5518.
- Henriksen, H.J., Roberts, M.J., van der Keur, P., Harjanne, A., Egilson, D., Alfonso, L., 2018. Participatory early warning and monitoring systems: A Nordic framework for web-based flood risk management. *International journal of disaster risk reduction* 31, 1295–1306.
- Herman, M.R., Nejadhashemi, A.P., Abouali, M., Hernandez-Suarez, J.S., Daneshvar, F., Zhang, Z., Anderson, M.C., Sadeghi, A.M., Hain, C.R., Sharifi, A., 2018. Evaluating the role of evapotranspiration remote sensing data in improving hydrological modeling predictability. *Journal of Hydrology* 556, 39–49.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al., 2020. The ERA5

REFERENCES

- global reanalysis. *Quarterly Journal of the Royal Meteorological Society* 146, 1999–2049.
- Hijmans, R.J., 2018. raster: Geographic Data Analysis and Modeling. R package version 2.8-4.
- Hijmans, R.J., 2020. raster: Geographic Data Analysis and Modeling. R package version 3.3-13.
- Hirpa, F.A., Gebremichael, M., Hopson, T., 2010. Evaluation of high-resolution satellite precipitation products over very complex terrain in Ethiopia. *Journal of Applied Meteorology and Climatology* 49, 1044–1051.
- Hirpa, F.A., Hopson, T.M., De Groeve, T., Brakenridge, G.R., Gebremichael, M., Restrepo, P.J., 2013. Upstream satellite remote sensing for river discharge forecasting: Application to major rivers in South Asia. *Remote Sensing of Environment* 131, 140–151.
- Hoffmann, L., Günther, G., Li, D., Stein, O., Wu, X., Griessbach, S., Heng, Y., Konopka, P., Müller, R., Vogel, B., et al., 2019. From ERA-Interim to ERA5: the considerable impact of ECMWF’s next-generation reanalysis on lagrangian transport simulations. *Atmospheric Chemistry and Physics* 19, 3097–3124.
- Hofmann, J., Schüttrumpf, H., 2019. Risk-based early warning system for pluvial flash floods: Approaches and foundations. *Geosciences* 9, 127.
- Hong, Y., Adler, R.F., Huffman, G., 2007a. An experimental global prediction system for rainfall-triggered landslides using satellite remote sensing and geospatial datasets. *IEEE Transactions on Geoscience and Remote Sensing* 45, 1671–1680.
- Hong, Y., Gochis, D., Cheng, J.t., Hsu, K.l., Sorooshian, S., 2007b. Evaluation of PERSIANN-CCS rainfall measurement using the name event rain gauge network. *Journal of Hydrometeorology* 8, 469–482.
- Hossain, F., 2006. Towards formulation of a space-borne system for early warning of floods: Can cost-effectiveness outweigh prediction uncertainty? *Natural Hazards* 37, 263–276.
- Houck, O.A., 2002. The clean water act TMDL program: Law, policy, and implementation. Environmental Law Institute.

REFERENCES

- Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., Arheimer, B., Blume, T., Clark, M., Ehret, U., et al., 2013. A decade of Predictions in Ungauged basins (PUB)—a review. *Hydrological sciences journal* 58, 1198–1255.
- Huang, S., Eisner, S., Magnusson, J.O., Lussana, C., Yang, X., Beldring, S., 2019. Improvements of the spatially distributed hydrological modelling using the HBV model at 1 km resolution for Norway. *Journal of hydrology* 577, 123585.
- Huffman, G.J., Adler, R.F., Bolvin, D.T., Nelkin, E.J., 2010. The TRMM Multi-satellite Precipitation Analysis (TMPA). Chapter 1 in *Satellite Rainfall Applications for Surface Hydrology*, F. Hossain and M. Gebremichael, Eds.
- Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Xie, P., Yoo, S.H., 2015. NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG). Algorithm Theoretical Basis Document (ATBD) Version 4, 26.
- Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Wolff, D.B., Adler, R.F., Gu, G., Hong, Y., Bowman, K.P., Stocker, E.F., 2007. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology* 8, 38.
- IDEAM, 2001. Estudio ambiental de la cuenca Magdalena-Cauca y elementos para su ordenamiento territorial. Technical report and Arcinfo database.
- Jansen, K.F., Teuling, A.J., Craig, J.R., Dal Molin, M., Knoben, W.J., Parajka, J., Vis, M., Melsen, L.A., 2021. Mimicry of a conceptual hydrological model (HBV): What’s in a name? *Water Resources Research* 57, 29–143.
- Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E., et al., 2008. Hole-filled SRTM for the globe Version 4. Available from the CGIAR-CSI SRTM 90 m Database (<http://srtm.csi.cgiar.org>) 15, 25–54.
- Jewell, S.A., Gaussiat, N., 2015. An assessment of kriging-based rain-gauge-radar merging techniques. *Quarterly Journal of the Royal Meteorological Society* 141, 2300–2313.
- Jiang, S., Ren, L., Hong, Y., Yong, B., Yang, X., Yuan, F., Ma, M., 2012. Comprehensive evaluation of multi-satellite precipitation products with a dense rain gauge network and optimally merging their simulated hydrological flows

REFERENCES

- using the Bayesian model averaging method. *Journal of Hydrology* 452, 213–225.
- Joyce, R.J., Janowiak, J.E., Arkin, P.A., Xie, P., 2004. CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology* 5, 487–503.
- Jurečka, F., Fischer, M., Hlavinka, P., Balek, J., Semerádová, D., Bláhová, M., Anderson, M.C., Hain, C., Žalud, Z., Trnka, M., 2021. Potential of water balance and remote sensing-based evapotranspiration models to predict yields of spring barley and winter wheat in the Czech Republic. *Agricultural Water Management* 256, 107064.
- Kallis, G., Butler, D., 2001. The EU water framework directive: measures and implications. *Water policy* 3, 125–142.
- Kearney, M.R., Maino, J.L., 2018. Can next-generation soil data products improve soil moisture modelling at the continental scale? An assessment using a new microclimate package for the R programming environment. *Journal of Hydrology* 561, 662–673.
- Kennedy, J., Eberhart, R., 1995. Particle swarm optimization, in: *Neural Networks, 1995. Proceedings., IEEE International Conference on*, pp. 1942–1948.
- Kerr, Y.H., Waldteufel, P., Wigneron, J.P., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M.J., Font, J., Reul, N., Gruhier, C., et al., 2010. The SMOS mission: New tool for monitoring key elements of the global water cycle. *Proceedings of the IEEE* 98, 666–687.
- Kidd, C., Becker, A., Huffman, G.J., Muller, C.L., Joe, P., Skofronick-Jackson, G., Kirschbaum, D.B., 2017. So, how much of the Earth’s surface is covered by rain gauges? *Bulletin of the American Meteorological Society* 98, 69–78.
- Kiguchi, M., Shen, Y., Kanae, S., Oki, T., 2015. Re-evaluation of future water stress due to socio-economic and climate factors under a warming climate. *Hydrological Sciences Journal* 60, 14–29.
- Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper Danube Basin under an ensemble of climate change scenarios. *Journal of Hydrology* 424, 264–277.

REFERENCES

- Knoben, W.J., Freer, J.E., Woods, R.A., 2019. Inherent benchmark or not? comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences* 23, 4323–4331.
- Kobayashi, S., Ota, Y., Harada, Y., Ebata, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., et al., 2015. The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II* 93, 5–48.
- Koffler, D., Gauster, T., Laaha, G., 2016. *lfstat: Calculation of Low Flow Statistics for Daily Stream Flow Data*. R package version 0.9.4.
- Kollat, J., Reed, P., Wagener, T., 2012. When are multiobjective calibration trade-offs in hydrologic models meaningful? *Water Resources Research* 48.
- Konapala, G., Mishra, A.K., Wada, Y., Mann, M.E., 2020. Climate change will affect global water availability through compounding changes in seasonal precipitation and evaporation. *Nature Communications* 11, 1–10.
- Koppa, A., Gebremichael, M., Hopson, T.M., Riddle, E., Boehnert, J., Broman, D.P., 2021. A scalable Earth Observations-based decision support system for hydropower planning in Africa. *JAWRA Journal of the American Water Resources Association* .
- Kornfeld, R.P., Arnold, B.W., Gross, M.A., Dahya, N.T., Klipstein, W.M., Gath, P.F., Bettadpur, S., 2019. GRACE-FO: the gravity recovery and climate experiment follow-on mission. *Journal of spacecraft and rockets* 56, 931–951.
- Kummerow, C., 1998. Beamfilling errors in passive microwave rainfall retrievals. *Journal of Applied Meteorology* 37, 356–370.
- Kundu, D., Vervoort, R.W., van Ogtrop, F.F., 2017. The value of remotely sensed surface soil moisture for model calibration using SWAT. *Hydrological Processes* 31, 2764–2780.
- Kundzewicz, Z.W., Krysanova, V., Benestad, R., Hov, Ø., Piniewski, M., Otto, I.M., 2018. Uncertainty in climate change impacts on water resources. *Environmental Science & Policy* 79, 1–8.
- Lagos, M., Mendoza, P., Rondanellu, R., Daniele, D., Tomaás, G., 2019. Aplicación de la metodología de actualización del balance hídrico nacional en las cuencas de la Macrozona Sur y parte de la Macrozona Austral.

REFERENCES

- Li, M., Shao, Q., 2010. An improved statistical approach to merge satellite rainfall estimates and raingauge data. *Journal of Hydrology* 385, 51–64.
- Li, Z., Yang, D., Hong, Y., 2013. Multi-scale evaluation of high-resolution multi-sensor blended global precipitation products over the yangtze river. *Journal of Hydrology* 500, 157–169.
- 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: Atmospheres* 99, 14415–14428.
- Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. *R News* 2, 18–22.
- Lievens, H., Tomer, S.K., Al Bitar, A., De Lannoy, G.J., Drusch, M., Dumedah, G., Franssen, H.J.H., Kerr, Y.H., Martens, B., Pan, M., et al., 2015. SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia. *Remote Sensing of Environment* 168, 146–162.
- Lindström, G., 1997. A simple automatic calibration routine for the HBV model. *Hydrology Research* 28, 153–168.
- Liu, C., Guo, L., Ye, L., Zhang, S., Zhao, Y., Song, T., 2018. A review of advances in China’s flash flood early-warning system. *Natural hazards* 92, 619–634.
- Liu, Y., Hu, J., Snell-Feikema, I., VanBemmel, M.S., Lamsal, A., Wimberly, M.C., 2015. Software to facilitate remote sensing data access for disease early warning systems. *Environmental Modelling & Software* 74, 247–257.
- Lorenz, C., Kunstmann, H., 2012. The hydrological cycle in three state-of-the-art reanalyses: Intercomparison and performance analysis. *Journal of Hydrometeorology* 13, 1397–1420.
- Ly, S., Charles, C., Degre, A., 2011. Geostatistical interpolation of daily rainfall at catchment scale: The use of several variogram models in the Ourthe and Ambleve catchments, Belgium. *Hydrology and Earth System Sciences* 15, 2259–2274.
- Ma, Y., Hong, Y., Chen, Y., Yang, Y., Tang, G., Yao, Y., Long, D., Li, C., Han, Z., Liu, R., 2018. Performance of optimally merged multisatellite precipitation products using the dynamic bayesian model averaging scheme over the Tibetan Plateau. *Journal of Geophysical Research: Atmospheres* 123, 814–834.

REFERENCES

- Ma, Y., Zhang, Y., Yang, D., Farhan, S.B., 2015. Precipitation bias variability versus various gauges under different climatic conditions over the Third Pole Environment (TPE) region. *International Journal of Climatology* 35, 1201–1211.
- Madsen, H., Lawrence, D., Lang, M., Martinkova, M., Kjeldsen, T., 2014. Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. *Journal of Hydrology* 519, 3634–3650.
- Maggioni, V., Massari, C., 2018. On the performance of satellite precipitation products in riverine flood modeling: A review. *Journal of Hydrology* 558, 214–224.
- Maggioni, V., Vergara, H.J., Anagnostou, E.N., Gourley, J.J., Hong, Y., Stamoulis, D., 2013. Investigating the applicability of error correction ensembles of satellite rainfall products in river flow simulations. *Journal of Hydrometeorology* 14, 1194–1211.
- Manz, B., Buytaert, W., Zulkaffi, Z., Lavado, W., Willems, B., Robles, L.A., Rodríguez-Sánchez, J.P., 2016. High-resolution satellite-gauge merged precipitation climatologies of the Tropical Andes. *Journal of Geophysical Research: Atmospheres* 121, 1190–1207.
- Martens, B., Gonzalez Miralles, D., Lievens, H., Van Der Schalie, R., De Jeu, R.A., Fernández-Prieto, D., Beck, H.E., Dorigo, W., Verhoest, N., 2017. GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development* 10, 1903–1925.
- Mazdiyasni, O., AghaKouchak, A., 2015. Substantial increase in concurrent droughts and heatwaves in the United States. *Proceedings of the National Academy of Sciences* 112, 11484–11489.
- McCabe, M.F., Rodell, M., Alsdorf, D.E., Miralles, D.G., Uijlenhoet, R., Wagner, W., Lucieer, A., Houborg, R., Verhoest, N.E., Franz, T.E., et al., 2017. The future of Earth observation in hydrology. *Hydrology and earth system sciences* 21, 3879–3914.
- McIntyre, N., Lee, H., Wheeler, H., Young, A., Wagener, T., 2005. Ensemble predictions of runoff in ungauged catchments. *Water Resources Research* 41.
- McNamara, I., Baez-Villanueva, O.M., Zomorodian, A., Ayyad, S., Zambrano-Bigiarini, M., Zaroug, M., Mersha, A., Nauditt, A., Mbuliro, M., Wamala, S., et al., 2021. How well do gridded precipitation and actual evapotranspiration

REFERENCES

- products represent the key water balance components in the Nile Basin? *Journal of Hydrology: Regional Studies* 37, 100884.
- McVicar, T.R., Körner, C., 2013. On the use of elevation, altitude, and height in the ecological and climatological literature. *Oecologia* 171, 335–337.
- Mekonnen, K., Melesse, A.M., Woldesenbet, T.A., 2021. Spatial evaluation of satellite-retrieved extreme rainfall rates in the Upper Awash River Basin, Ethiopia. *Atmospheric Research* 249, 105297.
- Melo, D.d.C., Xavier, A.C., Bianchi, T., Oliveira, P.T., Scanlon, B.R., Lucas, M.C., Wendland, E., 2015. Performance evaluation of rainfall estimates by TRMM Multi-satellite Precipitation Analysis 3B42V6 and V7 over Brazil. *Journal of Geophysical Research: Atmospheres* 120, 9426–9436.
- Melsen, L.A., Addor, N., Mizukami, N., Newman, A.J., Torfs, P.J.J.F., Clark, M.P., Uijlenhoet, R., Teuling, A.J., 2018. Mapping (dis)agreement in hydrologic projections. *Hydrology and Earth System Sciences* 22, 1775–1791.
- 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* 30, 1071–1095.
- Menne, M.J., Durre, I., Vose, R.S., Gleason, B.E., Houston, T.G., 2012. An overview of the global historical climatology network-daily database. *Journal of atmospheric and oceanic technology* 29, 897–910.
- Merz, R., Blöschl, G., 2004. Regionalisation of catchment model parameters. *Journal of hydrology* 287, 95–123.
- Mizukami, N., Rakovec, O., Newman, A.J., Clark, M.P., Wood, A.W., Gupta, H.V., Kumar, R., 2019. On the choice of calibration metrics for “high-flow” estimation using hydrologic models. *Hydrology and Earth System Sciences* 23, 2601–2614.
- Mohammadi, B., Ahmadi, F., Mehdizadeh, S., Guan, Y., Pham, Q.B., Linh, N.T.T., Tri, D.Q., 2020. Developing novel robust models to improve the accuracy of daily streamflow modeling. *Water Resources Management* 34, 3387–3409.
- Montecinos, A., Aceituno, P., 2003. Seasonality of the ENSO-related rainfall variability in Central Chile and associated circulation anomalies. *Journal of Climate* 16, 281–296.

REFERENCES

- Mostafaie, A., Forootan, E., Safari, A., Schumacher, M., 2018. Comparing multi-objective optimization techniques to calibrate a conceptual hydrological model using in situ runoff and daily grace data. *Computational Geosciences* 22, 789–814.
- MoUD, 2008. National urban sanitation policy. ministry of urban development (moud).
- Mu, Q., Zhao, M., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment* 115, 1781–1800.
- Neri, M., Parajka, J., Toth, E., 2020. Importance of the informative content in the study area when regionalising rainfall-runoff model parameters: the role of nested catchments and gauging station density. *Hydrology and Earth System Sciences* 24, 5149–5171.
- Nie, S., Luo, Y., Wu, T., Shi, X., Wang, Z., 2015. A merging scheme for constructing daily precipitation analyses based on objective bias-correction and error estimation techniques. *Journal of Geophysical Research: Atmospheres* 120, 8671–8692.
- Nikolopoulos, E.I., Anagnostou, E.N., Borga, M., 2013. Using high-resolution satellite rainfall products to simulate a major flash flood event in northern Italy. *Journal of Hydrometeorology* 14, 171–185.
- Niu, W.j., Feng, Z.k., 2021. Evaluating the performances of several artificial intelligence methods in forecasting daily streamflow time series for sustainable water resources management. *Sustainable Cities and Society* 64, 102562.
- Oki, T., Kanae, S., 2006. Global hydrological cycles and world water resources. *Science* 313, 1068–1072.
- Oliveira, R., Maggioni, V., Vila, D., Morales, C., 2016. Characteristics and diurnal cycle of GPM rainfall estimates over the central Amazon region. *Remote Sensing* 8, 544.
- Oliver, M., Webster, R., 2014. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena* 113, 56–69.
- Ollivier, C., Mazzilli, N., Oliosio, A., Chalikakis, K., Carrière, S.D., Danquigny, C., Emblanch, C., 2020. Karst recharge-discharge semi distributed model to assess spatial variability of flows. *Science of the Total Environment* 703, 134368.

REFERENCES

- Oudin, L., Andréassian, V., Perrin, C., Michel, C., Le Moine, N., 2008. Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resources Research* 44.
- Parajka, J., Blaschke, A.P., Blöschl, G., Haslinger, K., Hepp, G., Laaha, G., Schöner, W., Trautvetter, H., Viglione, A., Zessner, M., 2016. Uncertainty contributions to low-flow projections in Austria. *Hydrology and Earth System Sciences* 20, 2085.
- Parajka, J., Merz, R., Blöschl, G., 2005. A comparison of regionalisation methods for catchment model parameters. *Hydrology and Earth System Sciences* 9, 157–171.
- Parajka, J., Merz, R., Blöschl, G., 2007. Uncertainty and multiple objective calibration in regional water balance modelling: Case study in 320 Austrian catchments. *Hydrological Processes: An International Journal* 21, 435–446.
- Perpiñán, O., Hijmans, R., 2020. rasterVis. R package version 0.49.
- Peterson, T.C., Vose, R.S., 1997. An overview of the global historical climatology network temperature database. *Bulletin of the American Meteorological Society* 78, 2837–2850.
- Pool, S., Vis, M.J., Knight, R.R., Seibert, J., 2017. Streamflow characteristics from modeled runoff time series—importance of calibration criteria selection. *Hydrology and Earth System Sciences* 21, 5443–5457.
- Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: Bagging and Random Forests for ecological prediction. *Ecosystems* 9, 181.
- Prein, A.F., Gobiet, A., 2017. Impacts of uncertainties in European gridded precipitation observations on regional climate analysis. *International Journal of Climatology* 37, 305–327.
- Prudhomme, C., Giuntoli, I., Robinson, E.L., Clark, D.B., Arnell, N.W., Dankers, R., Fekete, B.M., Franssen, W., Gerten, D., Gosling, S.N., et al., 2014. Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. *Proceedings of the National Academy of Sciences* 111, 3262–3267.

REFERENCES

- Qin, Y., Mueller, N.D., Siebert, S., Jackson, R.B., AghaKouchak, A., Zimmerman, J.B., Tong, D., Hong, C., Davis, S.J., 2019. Flexibility and intensity of global water use. *Nature Sustainability* 2, 515–523.
- Quesada-Montano, B., Wetterhall, F., Westerberg, I.K., Hidalgo, H.G., Halldin, S., 2019. Characterising droughts in Central America with uncertain hydro-meteorological data. *Theoretical and Applied Climatology* 137, 2125–2138.
- R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.
- Rabiei, E., Haberlandt, U., 2015. Applying bias correction for merging rain gauge and radar data. *Journal of Hydrology* 522, 544–557.
- Rahman, K.U., Shang, S., Zohaib, M., 2021. Assessment of merged satellite precipitation datasets in monitoring meteorological drought over Pakistan. *Remote Sensing* 13, 1662.
- Rajulapati, C.R., Papalexiou, S.M., Clark, M.P., Razavi, S., Tang, G., Pomeroy, J.W., 2020. Assessment of extremes in global precipitation products: How reliable are they? *Journal of Hydrometeorology* 21, 2855–2873.
- Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., Attinger, S., Schäfer, D., Schrön, M., Samaniego, L., 2016. Multiscale and multivariate evaluation of water fluxes and states over European river basins. *Journal of Hydrometeorology* 17, 287–307.
- Rana, S., McGregor, J., Renwick, J., 2015. Precipitation seasonality over the Indian subcontinent: An evaluation of gauge, reanalyses, and satellite retrievals. *Journal of Hydrometeorology* 16, 631–651.
- Restrepo, J.D., Kjerfve, B., Hermelin, M., Restrepo, J.C., 2006. Factors controlling sediment yield in a major South American drainage basin: the Magdalena River, Colombia. *Journal of Hydrology* 316, 213–232.
- Richey, A.S., Thomas, B.F., Lo, M.H., Famiglietti, J.S., Swenson, S., Rodell, M., 2015. Uncertainty in global groundwater storage estimates in a total groundwater stress framework. *Water resources research* 51, 5198–5216.
- Rios Gaona, M., Overeem, A., Leijnse, H., Uijlenhoet, R., 2015. Measurement and interpolation uncertainties in rainfall maps from cellular communication networks. *Hydrology and Earth System Sciences* 19, 3571–3584.

REFERENCES

- Rivera, N., Encina, F., Muñoz-Pedrerros, A., Mejias, P., 2004. La calidad de las aguas en los ríos Cautín e Imperial, IX Región-Chile. *Información tecnológica* 15, 89–101.
- Robertson, A.W., Baethgen, W., Block, P., Lall, U., Sankarasubramanian, A., de Souza Filho, F.d.A., Verbist, K.M., 2014. Climate risk management for water in semi-arid regions. *Earth Perspectives* 1, 12.
- Rockström, J., Falkenmark, M., Lannerstad, M., Karlberg, L., 2012. The planetary water drama: Dual task of feeding humanity and curbing climate change. *Geophysical Research Letters* 39.
- Rodriguez-Castiblanco, C., Garcia-Echeverri, C., Gonzales-Murillo, C., Zamora, D., 2021. Efectos de las condiciones climáticas cambiantes en la simulación de la esorrentía en la cuenca del río Sogamoso. *Comision Tecnica Permanente de Ingenieria y Recursos Hidricos*, Barranquilla, Colombia.
- Rojas, R., Feyen, L., Watkiss, P., 2013. Climate change and river floods in the European Union: Socio-economic consequences and the costs and benefits of adaptation. *Global Environmental Change* 23, 1737–1751.
- Román-Cascón, C., Pellarin, T., Gibon, F., Brocca, L., Cosme, E., Crow, W., Fernández-Prieto, D., Kerr, Y.H., Massari, C., 2017. Correcting satellite-based precipitation products through SMOS soil moisture data assimilation in two land-surface models of different complexity: API and SURFEX. *Remote sensing of environment* 200, 295–310.
- Rossa, A.M., Del Guerra, F.L., Borga, M., Zanon, F., Settin, T., Leuenberger, D., 2010. Radar-driven high-resolution hydro-meteorological forecasts of the 26 september 2007 Venice flash flood. *Journal of Hydrology* 394, 230–244.
- Roudier, P., Andersson, J.C., Donnelly, C., Feyen, L., Greuell, W., Ludwig, F., 2016. Projections of future floods and hydrological droughts in Europe under a+2°C global warming. *Climatic change* 135, 341–355.
- Roy, M.H., Larocque, D., 2012. Robustness of random forests for regression. *Journal of Nonparametric Statistics* 24, 993–1006.
- Rozante, J.R., Moreira, D.S., de Goncalves, L.G.G., Vila, D.A., 2010. Combining TRMM and surface observations of precipitation: Technique and validation over South America. *Weather and forecasting* 25, 885–894.
- Saadi, M., Oudin, L., Ribstein, P., 2019. Random forest ability in regionalizing hourly hydrological model parameters. *Water* 11, 1540.

REFERENCES

- Muñoz Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D.G., Piles, M., Rodríguez-Fernández, N.J., Zsoter, E., Buontempo, C., Thépaut, J.N., 2021. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data* 13, 4349–4383.
- Sabater, S., Bregoli, F., Acuña, V., Barceló, D., Elosegí, A., Ginebreda, A., Marcé, R., Muñoz, I., Sabater-Liesa, L., Ferreira, V., 2018. Effects of human-driven water stress on river ecosystems: A meta-analysis. *Scientific Reports* 8, 1–11.
- Salio, P., Hobouchian, M.P., Skabar, Y.G., Vila, D., 2015. Evaluation of high-resolution satellite precipitation estimates over southern South America using a dense rain gauge network. *Atmospheric Research* 163, 146–161.
- Salvia, M.M., Sánchez, N., Piles, M., Ruscica, R.C., González-Zamora, Á., Roitberg, E., Martínez-Fernández, J., 2021. The added-value of remotely-sensed soil moisture data for agricultural drought detection in Argentina. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14, 6487–6500.
- Samaniego, L., Kumar, R., Attinger, S., 2010. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research* 46.
- Scanlon, B.R., Zhang, Z., Save, H., Wiese, D.N., Landerer, F.W., Long, D., Longuevergne, L., Chen, J., 2016. Global evaluation of new GRACE mascon products for hydrologic applications. *Water Resources Research* 52, 9412–9429.
- Schlösser, C.A., Strzepek, K., Gao, X., Fant, C., Blanc, É., Paltsev, S., Jacoby, H., Reilly, J., Gueneau, A., 2014. The future of global water stress: An integrated assessment. *Earth's Future* 2, 341–361.
- Schneider, U., Fuchs, T., Meyer-Christoffer, A., Rudolf, B., 2008. Global precipitation analysis products of the GPCC. Global Precipitation Climatology Centre (GPCC), DWD: Internet Publication 112.
- Senay, G., Velpuri, N.M., Bohms, S., Budde, M., Young, C., Rowland, J., Verdin, J., 2015. Drought monitoring and assessment: Remote sensing and modeling approaches for the famine early warning systems network, in: *Hydro-meteorological hazards, risks and disasters*. Elsevier, pp. 233–262.

REFERENCES

- Senay, G.B., Bohms, S., Singh, R.K., Gowda, P.H., Velpuri, N.M., Alemu, H., Verdin, J.P., 2013. Operational evapotranspiration mapping using remote sensing and weather datasets: A new parameterization for the SSEB approach. *JAWRA Journal of the American Water Resources Association* 49, 577–591.
- Serrat-Capdevila, A., Valdes, J.B., Stakhiv, E.Z., 2014. Water management applications for satellite precipitation products: Synthesis and recommendations. *JAWRA Journal of the American Water Resources Association* 50, 509–525.
- Sharma, S., Siddique, R., Reed, S., Ahnert, P., Mendoza, P., Mejia, A., 2018. Relative effects of statistical preprocessing and postprocessing on a regional hydrological ensemble prediction system. *Hydrology and Earth System Sciences* 22, 1831–1849.
- Sheffield, J., Wood, E.F., Pan, M., Beck, H., Coccia, G., Serrat-Capdevila, A., Verbist, K., 2018. Satellite remote sensing for water resources management: Potential for supporting sustainable development in data-poor regions. *Water Resources Research* 54, 9724–9758.
- Shen, Y., Xiong, A., Hong, Y., Yu, J., Pan, Y., Chen, Z., Saharia, M., 2014. Uncertainty analysis of five satellite-based precipitation products and evaluation of three optimally merged multi-algorithm products over the Tibetan Plateau. *International journal of remote sensing* 35, 6843–6858.
- Shi, H., Chen, J., Li, T., Wang, G., 2017. A new method for estimation of spatially distributed rainfall through merging satellite observations, raingauge records, and terrain digital elevation model data. *Journal of Hydro-environment Research* 28, 1–14.
- Silal, S.P., Little, F., Barnes, K.I., White, L.J., 2015. Predicting the impact of border control on malaria transmission: A simulated focal screen and treat campaign. *Malar J* 14, 268.
- Silva, W.M., Simões, S.J.C., 2014. Spatial intra-annual variability of precipitation based on geostatistics. A case study for the Paraíba do Sul Basin, South-eastern Brazil. *International Journal of Geosciences* , 408–417.
- Simoës, S.J., Barros, A.P., 2007. Regional hydroclimatic variability and Brazil's 2001 energy crisis. *Management of Environmental Quality: An International Journal* 18, 263–273.

REFERENCES

- Singer, M.B., Asfaw, D.T., Rosolem, R., Cuthbert, M.O., Miralles, D.G., MacLeod, D., Quichimbo, E.A., Michaelides, K., 2021. Hourly potential evapotranspiration at 0.1° resolution for the global land surface from 1981–present. *Scientific Data* 8, 1–13.
- Singh, S.K., Bárdossy, A., Göttinger, J., Sudheer, K., 2012. Effect of spatial resolution on regionalization of hydrological model parameters. *Hydrological Processes* 26, 3499–3509.
- Sleziak, P., Szolgay, J., Hlavčová, K., Danko, M., Parajka, J., 2020. The effect of the snow weighting on the temporal stability of hydrologic model efficiency and parameters. *Journal of Hydrology* 582, 124639.
- Smith, K.A., Barker, L.J., Tanguy, M., Parry, S., Harrigan, S., Legg, T.P., Prudhomme, C., Hannaford, J., 2019. A multi-objective ensemble approach to hydrological modelling in the UK: An application to historic drought reconstruction. *Hydrology and Earth System Sciences* 23, 3247–3268.
- Soares, P.V., Pereira, S.Y., Simoes, S.J.C., de Paula Bernardes, G., Barbosa, S.A., Trannin, I.C.B., 2012. The definition of potential infiltration areas in Guaratinguetá watershed, Paraíba do Sul Basin, Southeastern Brazil: An integrated approach using physical and land-use elements. *Environmental Earth Sciences* 67, 1685–1694.
- Sorooshian, S., Hsu, K.L., Gao, X., Gupta, H.V., Imam, B., Braithwaite, D., 2000. Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bulletin of the American Meteorological Society* 81, 2035–2046.
- Sreeparvathy, V., Srinivas, V., 2022. A bayesian fuzzy clustering approach for design of precipitation gauge network using merged remote sensing and ground-based precipitation products. *Water Resources Research* 58, e2021WR030612.
- Stisen, S., Sandholt, I., 2010. Evaluation of remote-sensing-based rainfall products through predictive capability in hydrological runoff modelling. *Hydrological Processes* 24, 879–891.
- Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., Hsu, K.L., 2018. A review of global precipitation data sets: Data sources, estimation, and inter-comparisons. *Reviews of Geophysics* 56, 79–107.
- Swain, J.B., Patra, K.C., 2017. Streamflow estimation in ungauged catchments using regional flow duration curve: Comparative study. *Journal of Hydrologic Engineering* 22, 04017010.

REFERENCES

- 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* 56, e2019WR026153.
- Tang, G., Behrangi, A., Long, D., Li, C., Hong, Y., 2018. Accounting for spatiotemporal errors of gauges: A critical step to evaluate gridded precipitation products. *Journal of hydrology* 559, 294–306.
- Tang, G., Clark, M.P., Papalexiou, S.M., 2021. The use of serially complete station data to improve the temporal continuity of gridded precipitation and temperature estimates. *Journal of Hydrometeorology* 22, 1553–1568.
- Tang, L., Tian, Y., Lin, X., 2014. Validation of precipitation retrievals over land from satellite-based passive microwave sensors. *Journal of Geophysical Research: Atmospheres* 119, 4546–4567.
- Tapiador, F.J., Turk, F.J., Petersen, W., Hou, A.Y., García-Ortega, E., Machado, L.A., Angelis, C.F., Salio, P., Kidd, C., Huffman, G.J., et al., 2012. Global precipitation measurement: Methods, datasets and applications. *Atmospheric Research* 104, 70–97.
- Tapley, B.D., Bettadpur, S., Ries, J.C., Thompson, P.F., Watkins, M.M., 2004. GRACE measurements of mass variability in the Earth system. *Science* 305, 503–505.
- Tarek, M., Brissette, F.P., Arsenault, R., 2020. Evaluation of the ERA5 re-analysis as a potential reference dataset for hydrological modelling over north america. *Hydrology and Earth System Sciences* 24, 2527–2544.
- Taszarek, M., Allen, J.T., Marchio, M., Brooks, H.E., 2021. Global climatology and trends in convective environments from ERA5 and rawinsonde data. *npj Climate and Atmospheric Science* 4, 1–11.
- Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., De Roo, A., 2013. Hydrological evaluation of satellite-based rainfall estimates over the Volta and Baro-Akobo Basin. *Journal of Hydrology* 499, 324–338.
- Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., Levizzani, V., De Roo, A., 2012. Validation of satellite-based precipitation products over sparsely gauged African river basins. *Journal of Hydrometeorology* 13, 1760–1783.

REFERENCES

- Tomsett, C., Leyland, J., 2019. Remote sensing of river corridors: A review of current trends and future directions. *River Research and Applications* 35, 779–803.
- UN, U.G.A., 2015. Transforming our world: the 2030 Agenda for Sustainable Development. Working Papers. eSocialSciences.
- Unduche, F., Tolossa, H., Senbeta, D., Zhu, E., 2018. Evaluation of four hydrological models for operational flood forecasting in a Canadian Prairie watershed. *Hydrological Sciences Journal* 63, 1133–1149.
- Valdés-Pineda, R., Pizarro, R., García-Chevesich, P., Valdés, J.B., Olivares, C., Vera, M., Balocchi, F., Pérez, F., Vallejos, C., Fuentes, R., Abarza, A., Helwig, B., 2014. Water governance in Chile: Availability, management and climate change. *Journal of Hydrology* 519, 2538–2567.
- Van Ginkel, K.C., Hoekstra, A.Y., Buurman, J., Hogeboom, R.J., 2018. Urban water security dashboard: Systems approach to characterizing the water security of cities. *Journal of water resources planning and management* 144, 04018075.
- Van Vliet, M., Van Beek, L., Eisner, S., Flörke, M., Wada, Y., Bierkens, M., 2016. Multi-model assessment of global hydropower and cooling water discharge potential under climate change. *Global Environmental Change* 40, 156–170.
- Vandewiele, G., Elias, A., 1995. Monthly water balance of ungauged catchments obtained by geographical regionalization. *Journal of hydrology* 170, 277–291.
- Vásquez, N., Cepeda, J., Gómez, T., Mendoza, P.A., Lagos, M., Boisier, J.P., Álvarez-Garretón, C., Vargas, X., 2021. Catchment-Scale Natural Water Balance in Chile, in: *Water Resources of Chile*. Springer, pp. 189–208.
- Verbist, K., Robertson, A.W., Cornelis, W.M., Gabriels, D., 2010. Seasonal predictability of daily rainfall characteristics in central northern Chile for dry-land management. *Journal of Applied Meteorology and Climatology* 49, 1938–1955.
- Verdin, A., Funk, C., Rajagopalan, B., Kleiber, W., 2016. Kriging and Local Polynomial methods for blending satellite-derived and gauge precipitation estimates to support hydrologic early warning systems. *IEEE Transactions on Geoscience and Remote Sensing* 54, 2552–2562.

REFERENCES

- Vetter, T., Huang, S., Aich, V., Yang, T., Wang, X., Krysanova, V., Hattermann, F., 2015. Multi-model climate impact assessment and intercomparison for three large-scale river basins on three continents. *Earth System Dynamics* 6.
- Vicente, G.A., Scofield, R.A., Menzel, W.P., 1998. The operational GOES infrared rainfall estimation technique. *Bulletin of the American Meteorological Society* 79, 1883–1898.
- Viglione, A., Parajka, J., 2020. TUWmodel: Lumped/Semi-Distributed Hydrological Model for Education Purposes. R package version 1.1-1.
- Villarini, G., Krajewski, W.F., 2008. Empirically-based modeling of spatial sampling uncertainties associated with rainfall measurements by rain gauges. *Advances in Water Resources* 31, 1015–1023.
- Viviroli, D., Kumm, M., Meybeck, M., Kallio, M., Wada, Y., 2020. Increasing dependence of lowland populations on mountain water resources. *Nature Sustainability* 3, 917–928.
- Vörösmarty, C.J., Green, P., Salisbury, J., Lammers, R.B., 2000. Global water resources: vulnerability from climate change and population growth. *Science* 289, 284–288.
- Wada, Y., Gleeson, T., Esnault, L., 2014. Wedge approach to water stress. *Nature Geoscience* 7, 615–617.
- Walker, S., 2000. The value of hydrometric information in water resources management and flood control. *Meteorological Applications* 7, 387–397.
- Wan, W., Zhao, J., Li, H.Y., Mishra, A., Ruby Leung, L., Hejazi, M., Wang, W., Lu, H., Deng, Z., Demissis, Y., et al., 2017. Hydrological drought in the anthropocene: Impacts of local water extraction and reservoir regulation in the us. *Journal of Geophysical Research: Atmospheres* 122, 11–313.
- Wang, C., Graham, R.M., Wang, K., Gerland, S., Granskog, M.A., 2019. Comparison of ERA5 and ERA-Interim near-surface air temperature, snowfall and precipitation over Arctic sea ice: Effects on sea ice thermodynamics and evolution. *The Cryosphere* 13, 1661–1679.
- Wei, L., Jiang, S., Ren, L., Wang, M., Zhang, L., Liu, Y., Yuan, F., Yang, X., 2021. Evaluation of seventeen satellite-, reanalysis-, and gauge-based precipitation products for drought monitoring across mainland China. *Atmospheric Research* 263, 105813.

REFERENCES

- Wei, X., Xun, Y., 2019. Evaluation of the effective forecast and decision horizon in optimal hydropower generation considering medium-range precipitation forecasts. *Water Supply* 19, 2147–2155.
- Weiss, L.L., Wilson, W.T., 1953. Evaluation of significance of slope changes in double-mass curves. *Transactions, American Geophysical Union* 34, 893–896.
- Whitfield, P.H., 2012. Floods in future climates: a review. *Journal of Flood Risk Management* 5, 336–365.
- Woldemeskel, F.M., Sivakumar, B., Sharma, A., 2013. Merging gauge and satellite rainfall with specification of associated uncertainty across Australia. *Journal of Hydrology* 499, 167–176.
- Wong, J.S., Yassin, F., Famiglietti, J.S., Pomeroy, J.W., 2021. A streamflow-oriented ranking-based methodological framework to combine multiple precipitation datasets across large river basins. *Journal of Hydrology* 603, 127174.
- Woolway, R.I., Kraemer, B.M., Lenters, J.D., Merchant, C.J., O’Reilly, C.M., Sharma, S., 2020. Global lake responses to climate change. *Nature Reviews Earth & Environment* 1, 388–403.
- Wright, D.B., Mantilla, R., Peters-Lidard, C.D., 2017. A remote sensing-based tool for assessing rainfall-driven hazards. *Environmental modelling & software* 90, 34–54.
- Xavier, A.C., King, C.W., Scanlon, B.R., 2016. Daily gridded meteorological variables in Brazil (1980–2013). *International Journal of Climatology* 36, 2644–2659.
- Xie, P., Joyce, R., Wu, S., Yoo, S.H., Yarosh, Y., Sun, F., Lin, R., 2017. Re-processed, Bias-Corrected CMORPH Global high-resolution precipitation estimates from 1998. *Journal of Hydrometeorology* 18, 1617–1641.
- Xie, P., Xiong, A.Y., 2011. A conceptual model for constructing high-resolution gauge-satellite merged precipitation analyses. *Journal of Geophysical Research: Atmospheres* 116, D21106.
- Xue, X., Hong, Y., Limaye, A.S., Gourley, J.J., Huffman, G.J., Khan, S.I., Dorji, C., Chen, S., 2013. Statistical and hydrological evaluation of TRMM-based Multi-satellite Precipitation Analysis over the Wangchu Basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins? *Journal of Hydrology* 499, 91–99.

REFERENCES

- Yang, Z., Hsu, K., Sorooshian, S., Xu, X., Braithwaite, D., Zhang, Y., Ver-bist, K.M.J., 2017. Merging high-resolution satellite-based precipitation fields and point-scale rain gauge measurements - A case study in Chile. *Journal of Geophysical Research: Atmospheres* 122, 5267–5284.
- Young, A.R., 2006. Stream flow simulation within UK ungauged catchments using a daily rainfall-runoff model. *Journal of Hydrology* 320, 155–172.
- Zambrano-Bigiarini, M., 2017a. hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series. R package version 0.3-10.
- Zambrano-Bigiarini, M., 2017b. hydroTSM: Time Series Management, Analysis and Interpolation for Hydrological Modelling. R package version 0.5-1 . doi: <https://doi.org/10.5281/zenodo.83964>.
- Zambrano-Bigiarini, M., 2018. Temporal and spatial evaluation of long-term satellite-based precipitation products across the complex topographical and climatic gradients of Chile, in: *Remote Sensing and Modeling of the Atmosphere, Oceans, and Interactions VII*, International Society for Optics and Photonics. p. 1078202.
- Zambrano-Bigiarini, M., 2020a. hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series. R package version 0.4-0.
- Zambrano-Bigiarini, M., 2020b. hydroTSM: Time Series Management, Analysis and Interpolation for Hydrological Modelling. R package version 0.6-0 . doi: <https://doi.org/10.5281/zenodo.83964>.
- Zambrano-Bigiarini, M., Baez-Villanueva, O.M., 2019. Characterizing meteorological droughts in data scarce regions using remote sensing estimates of precipitation, in: *Extreme Hydroclimatic Events and Multivariate Hazards in a Changing Environment*. Elsevier, pp. 221–246.
- Zambrano-Bigiarini, M., Baez-Villanueva, O., 2020. Tutorial for using hydroPSO to calibrate TUWmodel.
- Zambrano-Bigiarini, M., Baez-Villanueva, O.M., Tarantola, S., 2022. A model-independent and parallel-capable global sensitivity analysis software for selection of key model parameters. *Environmental Modelling and Software*, in prep.

REFERENCES

- Zambrano-Bigiarini, M., Nauditt, A., Birkel, C., Verbist, K., Ribbe, L., 2017. Temporal and spatial evaluation of satellite-based rainfall estimates across the complex topographical and climatic gradients of Chile. *Hydrology and Earth System Sciences* 21, 1295–1320.
- Zambrano-Bigiarini, M., Rojas, R., 2013. A model-independent Particle Swarm Optimisation software for model calibration. *Environmental Modelling & Software* 43, 5–25.
- Zegelew, M.B., Alfredsen, K., 2014. Transferability of hydrological model parameter spaces in the estimation of runoff in ungauged catchments. *Hydrological Sciences Journal* 59, 1470–1490.
- Zessner, M., Schönhart, M., Parajka, J., Trautvetter, H., Mitter, H., Kirchner, M., Hepp, G., Blaschke, A.P., Strenn, B., Schmid, E., 2017. A novel integrated modelling framework to assess the impacts of climate and socio-economic drivers on land use and water quality. *Science of The Total Environment* 579, 1137–1151.
- Zhang, C., Zhong, L., Fu, X., Wang, J., Wu, Z., 2016. Revealing water stress by the thermal power industry in China based on a high spatial resolution water withdrawal and consumption inventory. *Environmental science & technology* 50, 1642–1652.
- Zhang, L., Li, X., Zheng, D., Zhang, K., Ma, Q., Zhao, Y., Ge, Y., 2021. Merging multiple satellite-based precipitation products and gauge observations using a novel double machine learning approach. *Journal of Hydrology* 594, 125969.
- Zhang, Y., Chiew, F.H., 2009. Relative merits of different methods for runoff predictions in ungauged catchments. *Water Resources Research* 45, W07412.
- Zhang, Y., Kong, D., Gan, R., Chiew, F.H., McVicar, T.R., Zhang, Q., Yang, Y., 2019. Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017. *Remote Sensing of Environment* 222, 165–182.
- Zhang, Y., Vaze, J., Chiew, F.H., Li, M., 2015. Comparing flow duration curve and rainfall–runoff modelling for predicting daily runoff in ungauged catchments. *Journal of Hydrology* 525, 72–86.
- Zhang, Y., Wang, K., 2021. Global precipitation system size. *Environmental Research Letters* 16, 054005.

REFERENCES

- Zhao, T., Cai, X., Yang, D., 2011. Effect of streamflow forecast uncertainty on real-time reservoir operation. *Advances in water resources* 34, 495–504.
- Zhao, Y., Dong, N., Li, Z., Zhang, W., Yang, M., Wang, H., 2021. Future precipitation, hydrology and hydropower generation in the Yalong River Basin: Projections and analysis. *Journal of Hydrology* 602, 126738.
- Zhao, Y., Feng, D., Yu, L., Wang, X., Chen, Y., Bai, Y., Hernández, H.J., Galleguillos, M., Estades, C., Biging, G.S., Radke, J.D., Gong, P., 2016. Detailed dynamic land cover mapping of Chile: Accuracy improvement by integrating multi-temporal data. *Remote Sensing of Environment* 183, 170–185.

Eidesstattliche Versicherung

Dortmund 15.04.2022

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertationsschrift zum Thema:

Streamflow simulation in data-scarce regions using remote sensing data in combination with ground-based measurements

selbstständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Des Weiteren erkläre ich an Eides statt, dass diese Arbeit weder in gleicher noch in ähnlicher Fassung einer akademischen Prüfung vorgelegt wurde.

Oscar Manuel Baez Villanueva