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 McNamara[1], Hylke E. Beck[7], Joschka Thurner[1], Alexandra Nauditt[1], Lars Ribbe[1], and Nguyen Xuan Thinh[2]

[1]Institute for Technology and Resources Management in the Tropics and Subtropics (ITT), TH Köln, Cologne, Germany
[2]Faculty of Spatial Planning, TU Dortmund University, Dortmund, Germany
[3]Department of Civil Engineering, Universidad de la Frontera, Temuco, Chile
[4]Center for Climate and Resilience Research, Universidad de Chile, Santiago, Chile
[5]Department of Civil Engineering, Universidad de Chile, Santiago, Chile
[6]Advanced Mining Technology Center (AMTC), Universidad de Chile, Santiago, Chile
[7]GloH2O, Almere, the Netherlands

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

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 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; Beck et al., 2016; Carrillo et al., 2011). 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 Alfredsen, 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 time 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 4,229 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 remains 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?

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 one.
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.
Zelelew and Alfredsen (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 2–6 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 of performance from calibration/verification to regionalisation ($\sim$10%) when evaluated during flash flood events. Using an ensemble of 2–4 donor catchments yielded the best the 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 USA	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 / 1,113; 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 reduced in heavily regulated catchments. Further evaluation against complementary datasets showed the best agreement for ET, followed by TWS, and the lowest for SM.
Swain and Patra (2017)	India	32 / Leave-one-out cross-validation	SP, FS, and PR	SP (both kriging and 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	4,229 / Ten-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.

2 Study area and selection of catchments

Our study domain is continental Chile (Figure 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°S–56.0°S) and on average 180 km of longitudinal extension (76.0°W–66.0°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 (17.50–26.00°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). P increases with altitude and latitude (in the southern direction) ranging from almost zero in the Atacama Desert to $\sim$6000 mm yr^{-1} in the surroundings of Puerto Cardenas ($\sim$43.2°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 accessed 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 (big_dam $= 0$).
3. Less than 10% of Q allocated to consumptive uses (interv_degree < 0.1).
4. Not dominated by glaciers (lc_glacier $< 5\%$).
5. Less than 5% of the area defined as urban (imp_frac $< 5\%$).
6. Absence of substantial irrigation abstractions (crop_frac $< 20\%$).
7. Less than 20% of the area covered by forest plantations (fp_frac $< 20\%$).

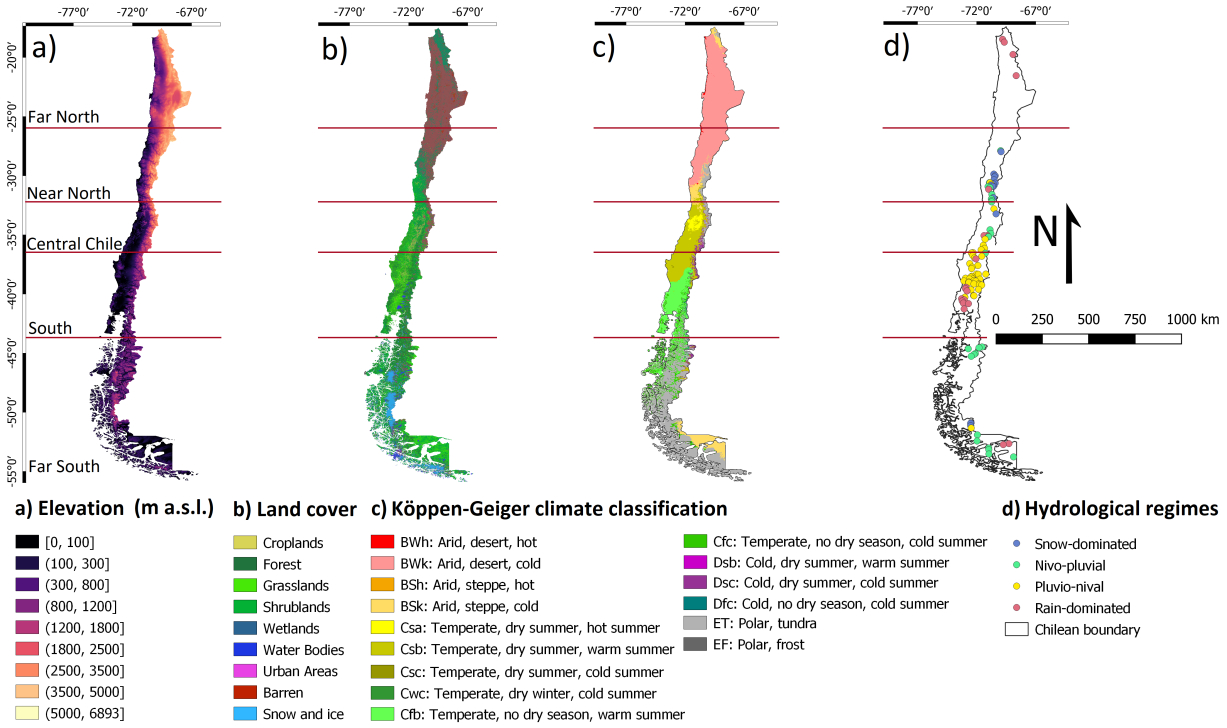

Figure 1. Study area: a) elevation (SRTMv4.1; Jarvis et al., 2008); b) land cover classification (Zhao et al., 2016); c) Köppen-Geiger climate classification (Beck et al., 2018); and d) hydrological regimes of the selected catchments over the five major macroclimatic zones according to Zambrano-Bigiarini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

8. No signs of artificial regulation in the hydrograph (10 excluded in total).

The drainage area of the selected catchments (100) ranges from 35 to 11,137 km^2, with a median value of 645 km^2. 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 Figure 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).

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)

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 reanalysis 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 Figure S1 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/), 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–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; $iii)$ parametrisation of microphysics; and $iv)$ representation of tropical variability (Hersbach et al., 2020). Although ERA5 also assimilates NCEP Stage IV P estimates over the conterminous USA, 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 compared to the remaining 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 Figure 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 Figure 2b shows boxplots of the mean monthly P averaged over catchments located within each macroclimatic zone. All P products show relatively similar patterns of spatial variability across continental 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 mm, 902 mm, and 2172 mm, respectively), when considering only our case study catchments (Figure 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 mm 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 (3,018 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 (Figure 3). First, to account for days without rain ($P < 1$ mm), we used the consecutive dry days index (CDD; Figure 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 less 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; Figure 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 days, followed by MSWEPv2.8 (0 to 15 days). 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 days. 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 (Figure 3c) and R95pTOT (Figure 3d) indices, which represent the maximum P accumulated over five 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 (Figure 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.

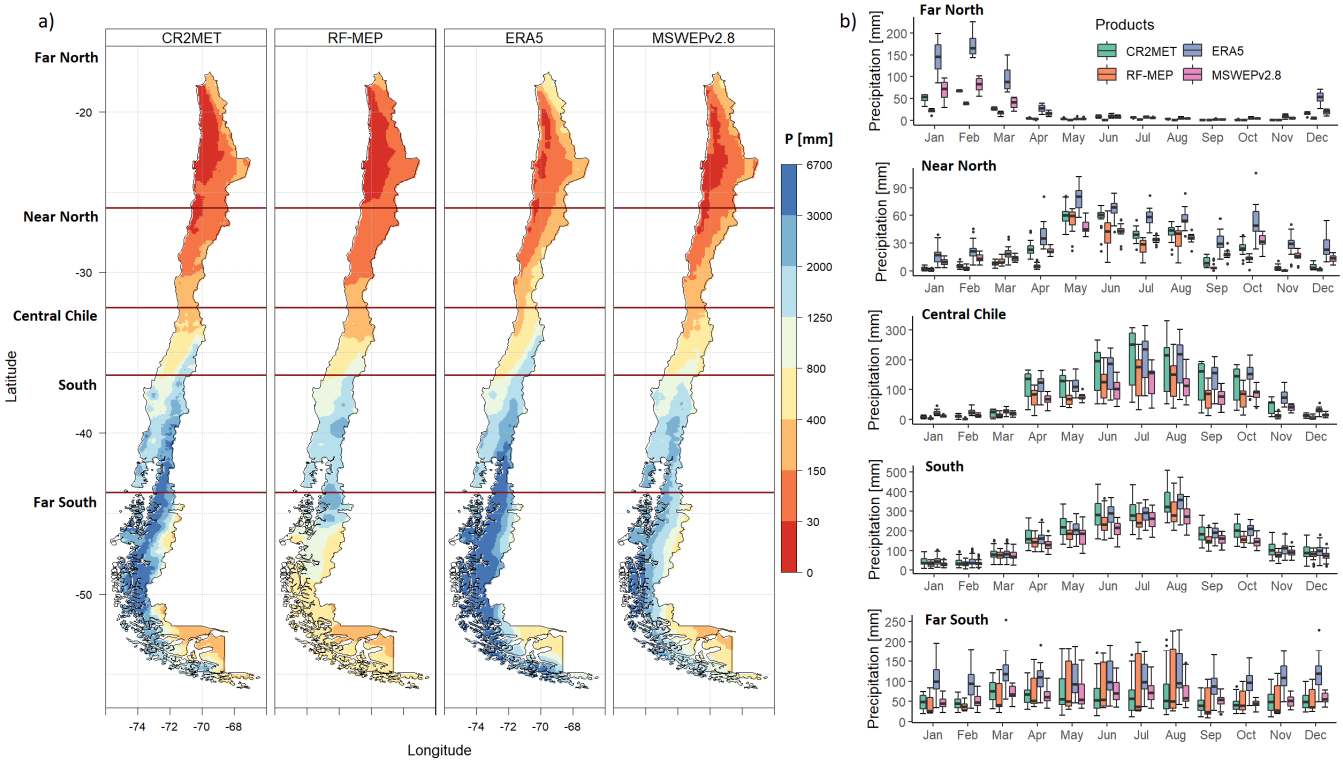

Figure 2. Comparison of P products over 1990–2018 (full time period): a) mean annual P for each product resampled to a 0.05° spatial resolution using the nearest neighbour method. The dark red horizontal lines represent the limits of each major macroclimatic zone; and b) mean monthly P averaged over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

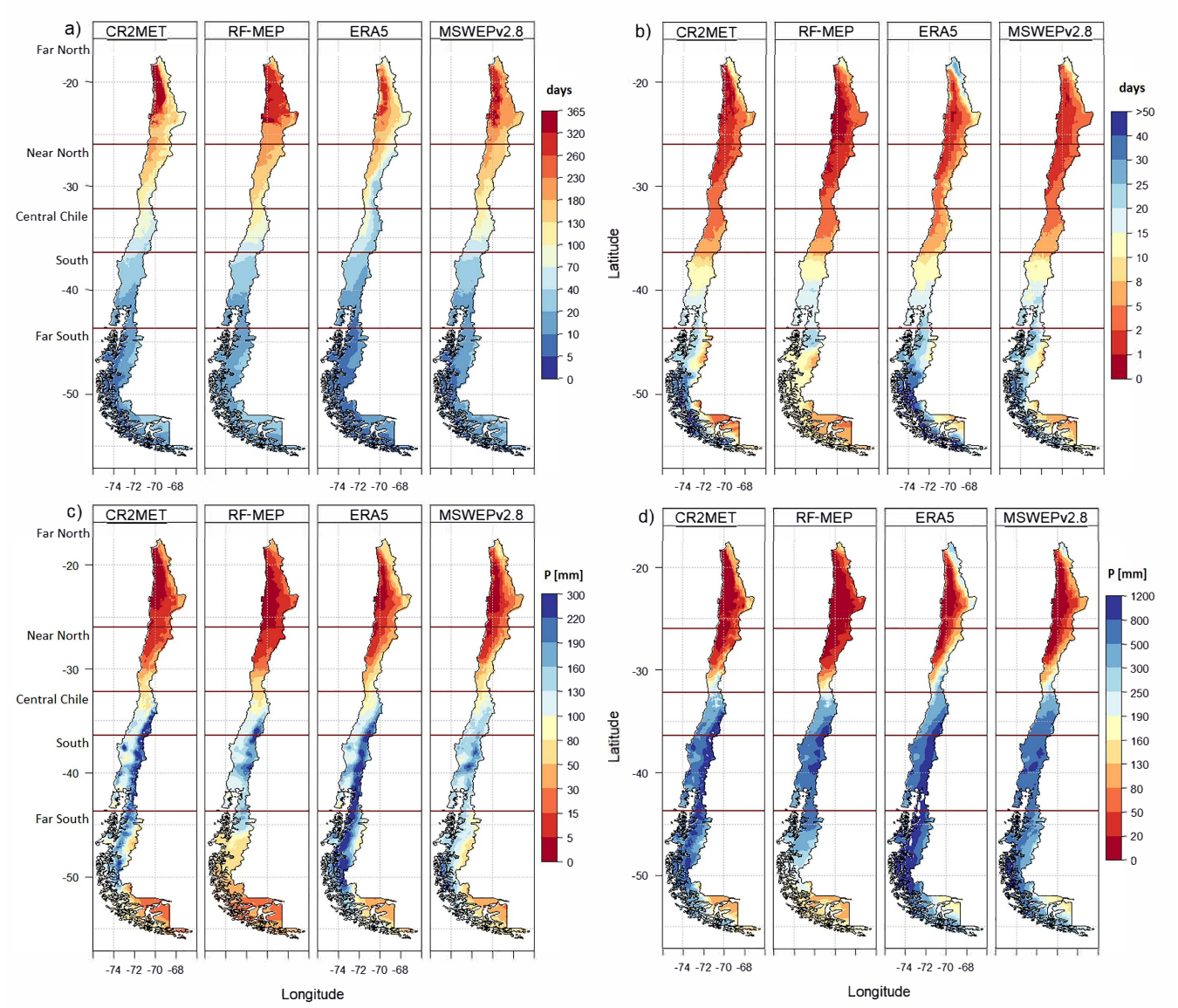

Figure 3. Median annual values of four Climdex indices over 1990–2018 (full-period): a) number of consecutive dry days (CDD); b) number of consecutive wet days (CWD); c) maximum P over five consecutive days (RX5day); and d) annual P that is above the 95th percentile of P accumulated for events that are above the 95th percentile of the daily P for wet days (R95pTOT). The dark red horizontal lines represent the limits of each macroclimatic zone.

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 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 (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; Sleziak 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 EZ equal-area elevation bands 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.

3.3 Independent catchment calibration and verification

The simulation period used for this study was 1990–2018. For calibration purposes, we used the first ten years as a conservative warm-up period to initialise the model stores, 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, we used an 8-year warm up period due to P product availability. We replicated Figures 2 and 3 for these three periods to analyse the differences between the selected P products (see the supplement, Figures S2-S7).

Table 3. 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

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 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 one, 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).

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

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

$$\beta = \frac{\mu_s}{\mu_o} \tag{3}$$

$$\gamma = \frac{CV_s}{CV_o} = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} \tag{4}$$

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 *spso2011* 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 = $1E-10$), 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., 2014a, b; 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 (Section 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):

$$\mathrm{ED}_{ij} = \sqrt{\sum_{k=1}^{n} (x_{k,i} - x_{k,j})^2} \tag{5}$$

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 redundant 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 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:

$$Z_f = \frac{x_f - x_{min}}{x_{max} - x_{min}} \tag{6}$$

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:

$$S_{i,j} = \sum_{m=1}^{n} \mid Z_{i,m} - Z_{j,m} \mid \tag{7}$$

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 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, 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 TUWmodel parameter, using all thirteen independent catchment characteristics listed in Table 4 as covariates. Our experimental setup used an ensemble of 2000 regression trees, a minimum of five

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.

N°	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-day precipitation	P product	Related to extreme P events.

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 cutoff 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 Section 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), 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.

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 Section 3.1.1), 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 Figure 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 Figure 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

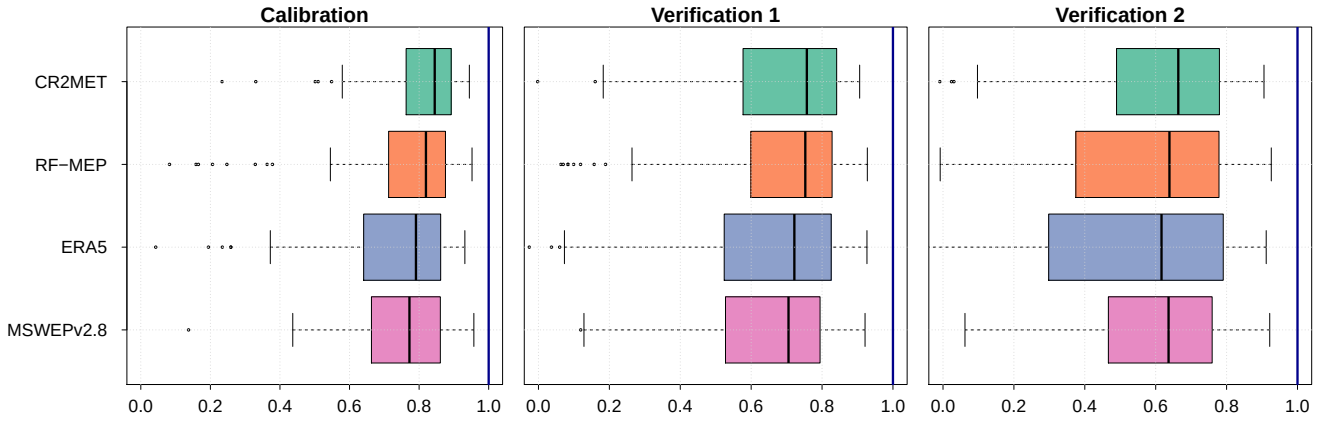

Figure 4. Performance of TUWmodel during the calibration (2000–2014), Verification 1 (1990–1999) and Verification 2 (2015–2018), prior to any regionalisation, using the modified Kling-Gupta efficiency (KGE'). The solid line represents the median value, the edges of the boxes represent the first and third 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'.

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).

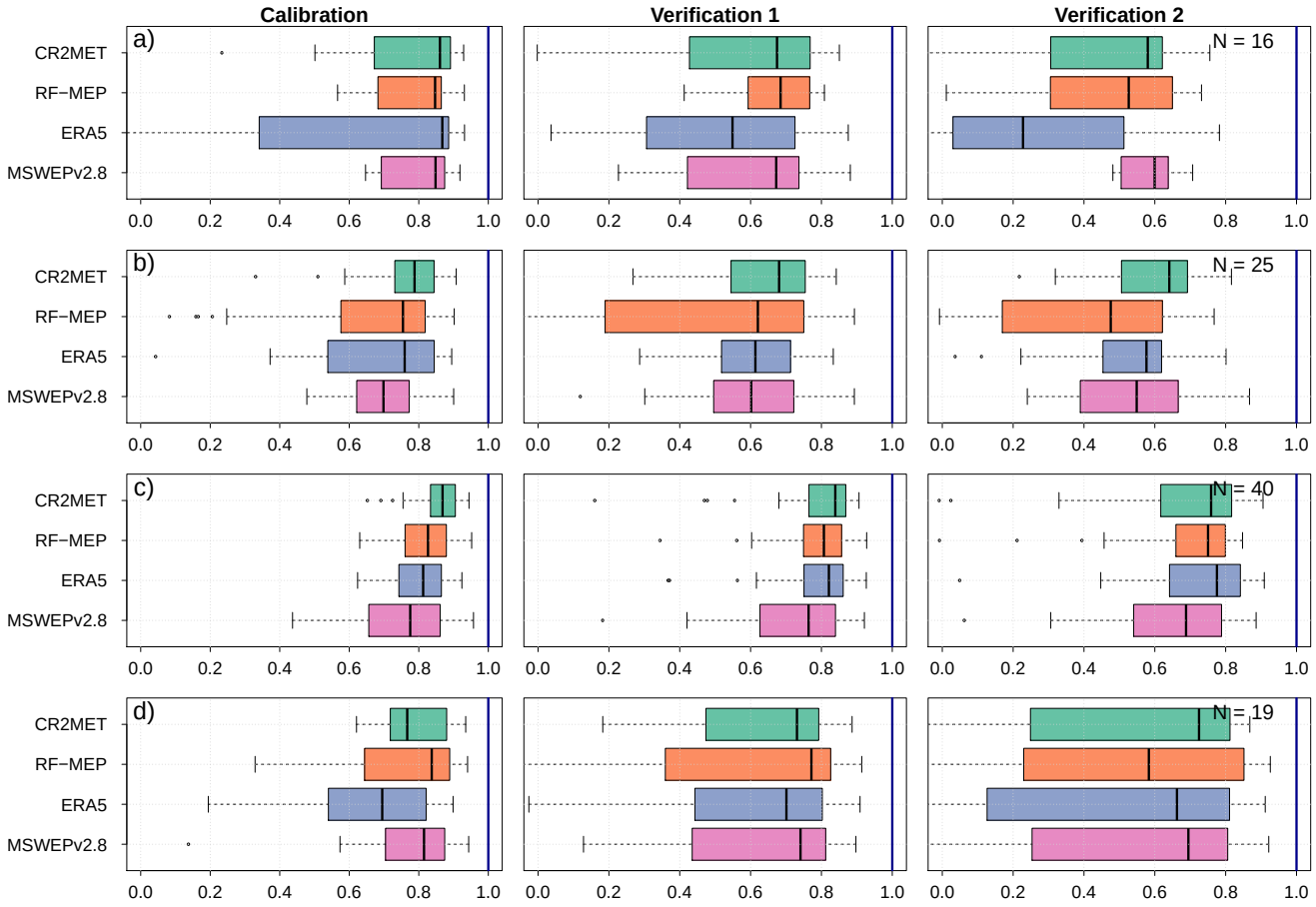

Figure 5. Performance of TUWmodel during calibration (2000–2014), Verification 1 (1990–1999) and Verification 2 (2015–2018), prior to any regionalisation over catchments with different hydrological regimes: a) snow-dominated; b) nivo-pluvial; c) pluvio-nival; and d) rain-dominated.

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 (Figure 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).

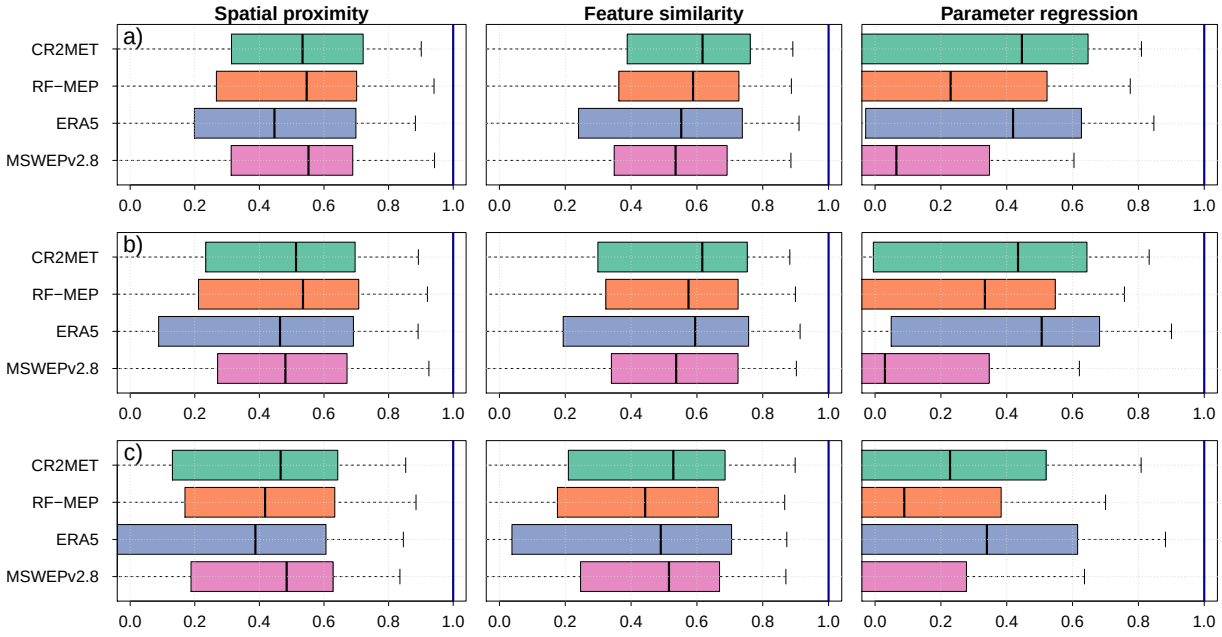

Figure 6. Leave-one-out cross-validation results for the three regionalisation methods applied with different P products during the: a) calibration (2000–2014); b) Verification 1 (1990–1999); and c) Verification 2 (2015–2018) periods.

For each regionalisation technique, Figure 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 evaluated 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 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 occur-

ring 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 amount of potential donor catchments (eleven for latitudes $> 49°$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 to catchments throughout the rest of the country (Alvarez-Garreton et al., 2018, their Figure 9). This may affect the hydrological simulations when model parameters from catchments located $< 49°$S are transferred to these far southern catchments.

4.2 Evaluation of regionalisation techniques

4.2.1 Overall performance

For each P product, Figure 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 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 were substantially lower, especially for spatial proximity and parameter regression.

These results are in agreement with the lower panels located below each map in Figure 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 Figures 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 Section 3.6). In general, the highest median performance is obtained when using 4 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 0.39 to 0.54). For all three periods, feature similarity provided better performance considering the distribution of the KGE' values.

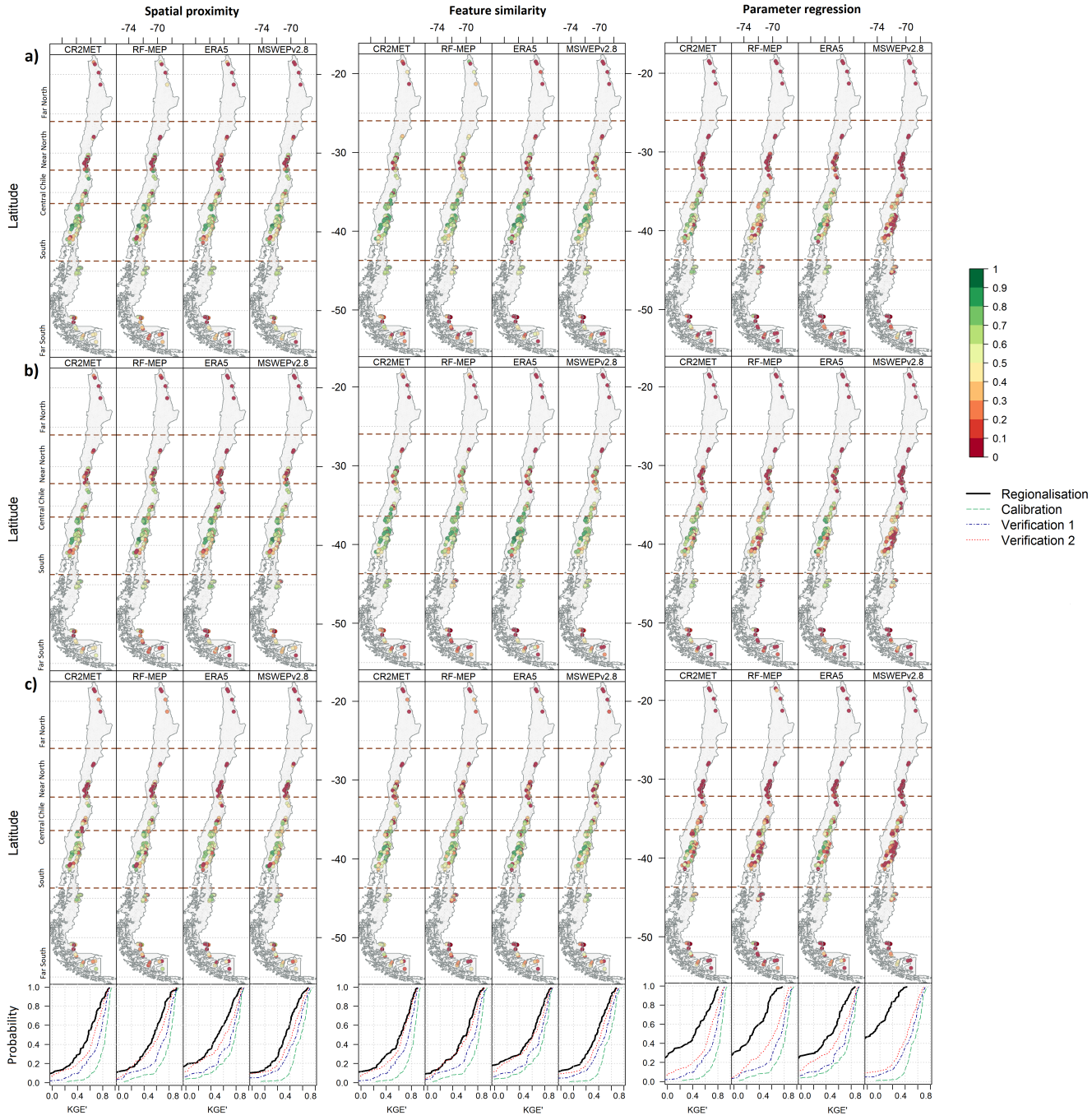

Figure 7. Spatial performance of the leave-one-out cross-validation results for the three regionalisation methods according to P product used to force TUWmodel. Results are presented for the: a) calibration (2000–2014); b)Verification 1 (1990–1999); and c) Verification 2 (2015–2018) periods. The panels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

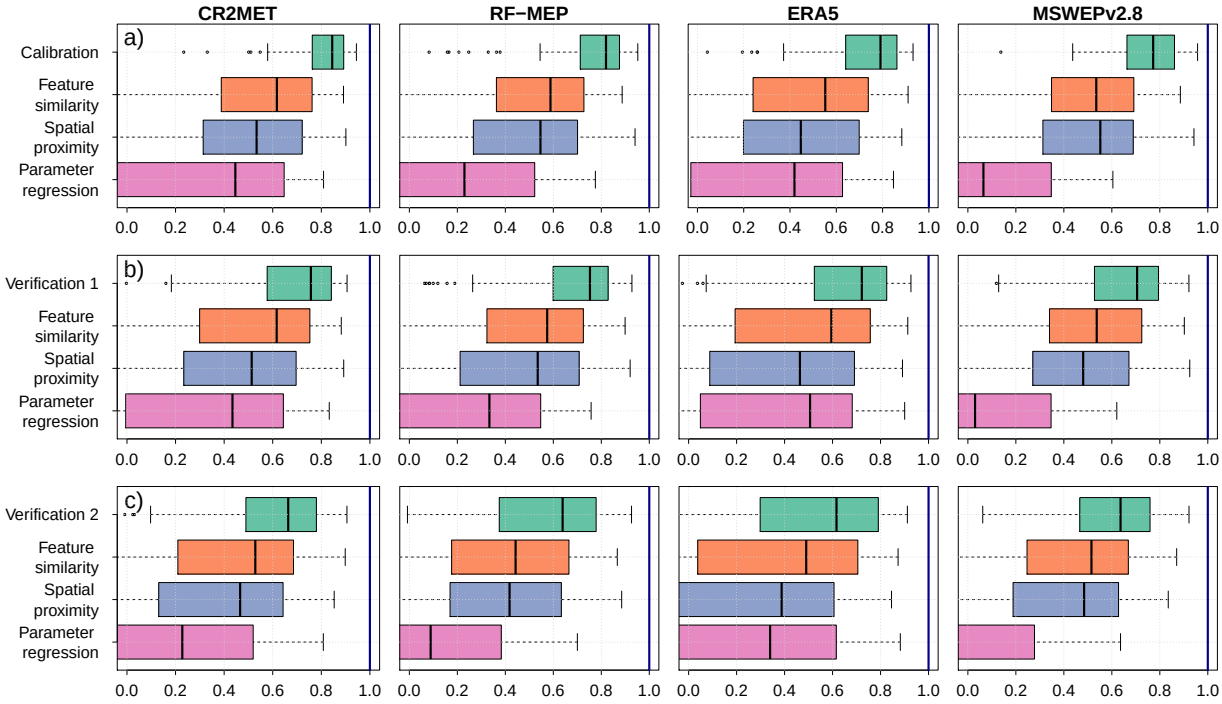

Figure 8. Performance of the regionalisation methods during the: a) calibration (2000–2014); b) Verification 1 (1990–1999); and c) Verification 2 (2015–2018) periods.

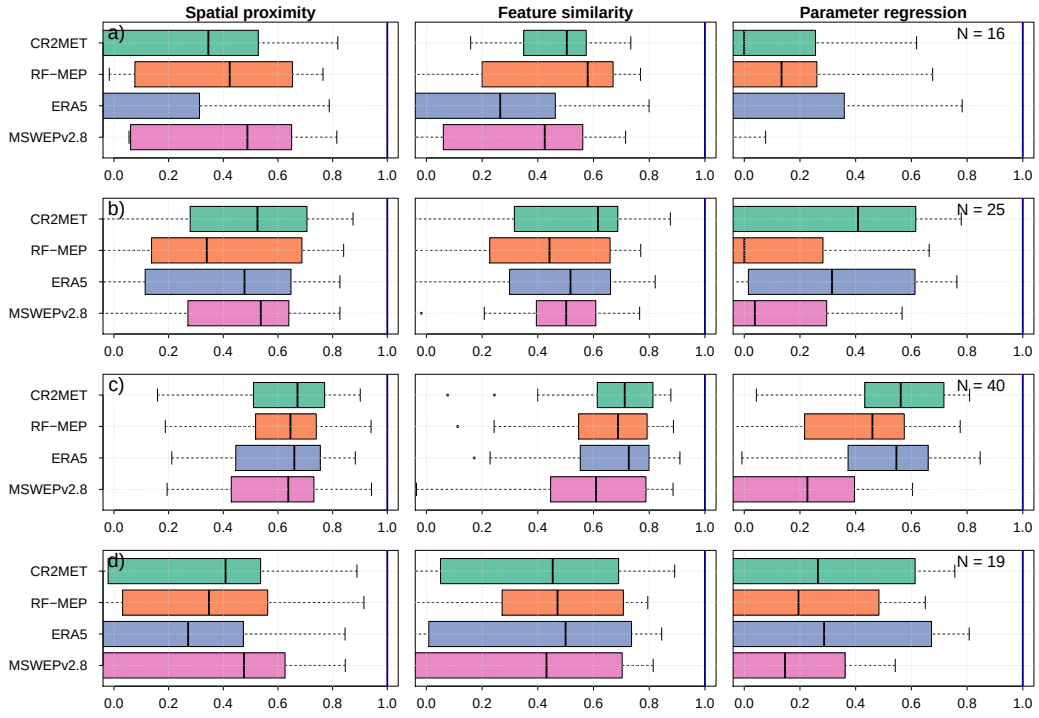

Figure 9. Performance of regionalisation methods for calibration (2000–2014) according to the hydrological regime: a) snow-dominated; b) nivo-pluvial; c) pluvio-nival; and d) rain-dominated. N denotes the number of catchments per hydrological regime.

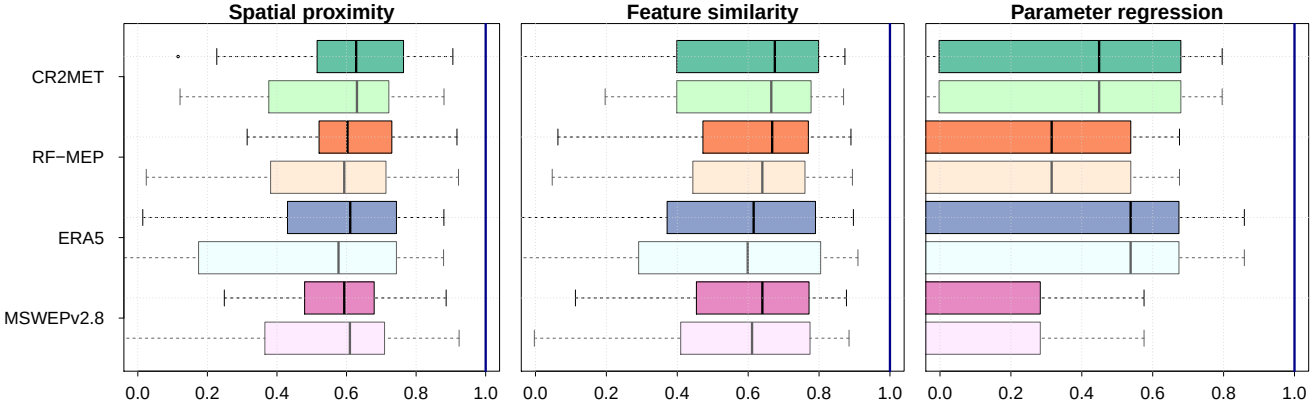

Figure 10. Comparison of regionalisation performance using all catchments as potential donors (dark colours) against the performance when nested catchments are excluded as potential donors (light colours).

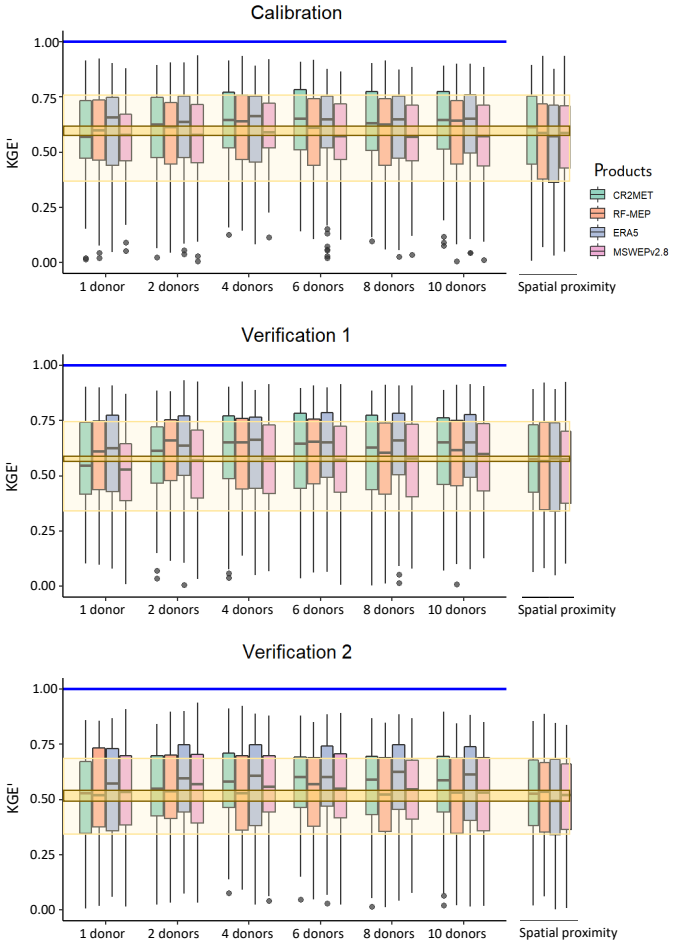

Figure 11. Influence of the number of donors used for feature similarity for calibration (2000—2014); Verification 1 (1990–1999); and Verification 2 (2015–2018). The results from spatial proximity are included on the right of each panel for comparison purposes. The dark yellow box denotes the upper and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

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 Figure 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 Figure 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 necessarily translate into improved hydrological model performance, which may be attributed to the: i) lack of P rain gauges in the Andes Mountains; ii) 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) 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 Figures 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 Figures 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 Figure 7), the TUWmodel appears to be flexible enough to compensate, 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 western USA 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 the bias of seven P products used to force the fully-distributed LISFLOOD model in four catchments in southern Africa.

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°; ∼25 km^2) than MSWEPv2.8 (∼0.10°; ∼100 km^2) and ERA5 (∼0.28°; ∼625 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 km^2) was similar to that obtained with products with a higher spatial resolution (Figure 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 through model calibration.

5.2 How does the calibration of 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 Figures 2 and 3) among the four products. Using the example of the nivo-pluvial catchments, Figure 12 illustrates how TUWmodel parameters compensate for differences between the P forcings used in calibration, while Figure 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 (Figures 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 (Figures 12 and S12). These catchments are primarily located in the arid Near North region (see Figure 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 Figure S1).

ERA5 presented relatively low SCF values over nivo-pluvial catchments compared to the other P products (Figure 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 (Figure 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 (Figure 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 Figure 3c–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 (Figure S16). This behaviour was more pronounced with RF-MEP and MSWEPv2.8, which exhibited the lowest P amounts and longer dry spells (Figure 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 Figure 3)) are similar for all four P products.

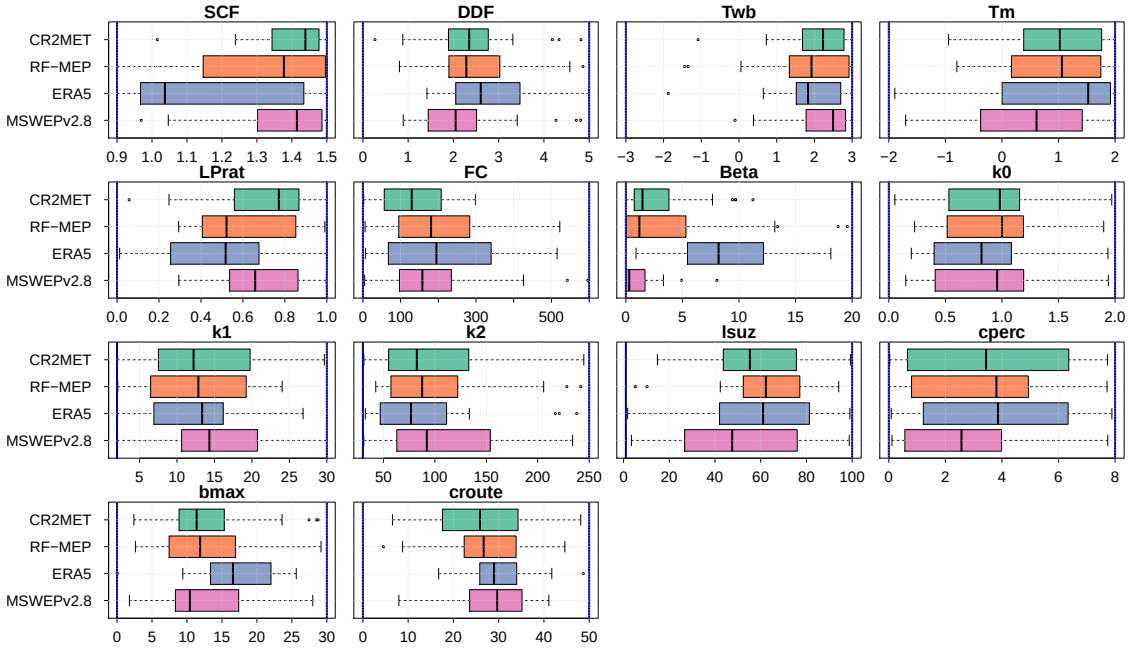

Figure 12. Model parameters obtained through calibration in nivo-pluvial catchments. The vertical blue lines indicate the upper and lower limits of the parameter ranges.

5.3 Evaluation of regionalisation techniques

The compensation due to the flexibility of the TUWmodel observed during the independent calibration and verification (see Section 5.2) also influences the regionalisation performance. Feature similarity provided the best performance when the TUW-

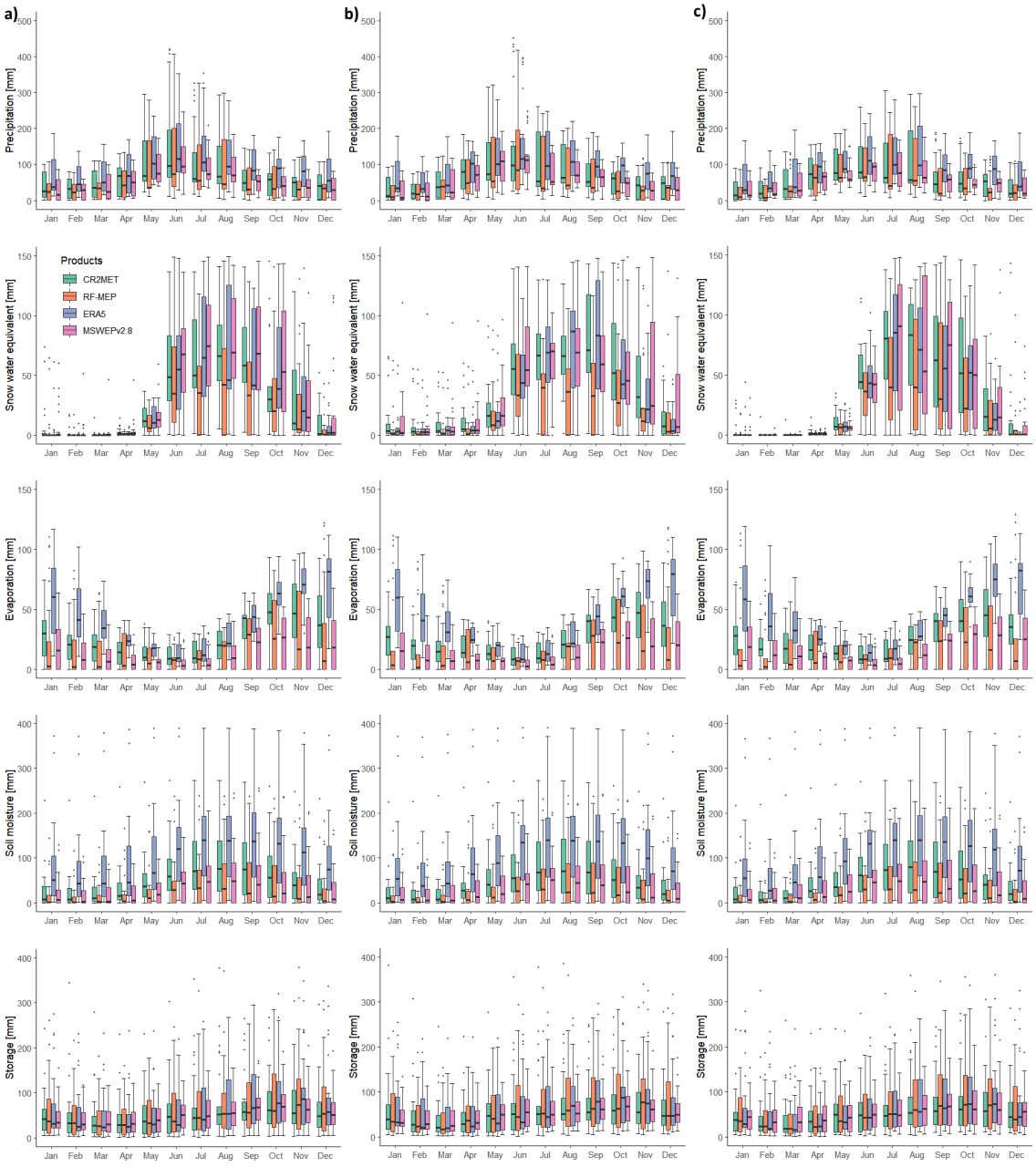

Figure 13. Mean monthly water balance components over nivo-pluvial catchments, obtained by forcing the TUW model with different P products for the: a) calibration (2000–2014); b) Verification 1 (1990–1999); and c) Verification 2 (2015–2018) periods. Mean monthly P was added for comparison purposes.

model was forced with all P products (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 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 Figure 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 (Figures 6 and 8); however, Figure 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 Figure 1d) for the calibration period. Comparing Figures 5 and 9, it is evident that the snow-dominated catchments performed substantially worse than in the independent performance during the same period (Figure 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 Figure 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).

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 4 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 time scale, 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 yielded the best (worst) results during regionalisation.

2. The P products corrected with daily gauge observations did not necessarily yielded 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.

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 of 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.

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 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.

Appendix A: Conceptual figure of hydrological regimes

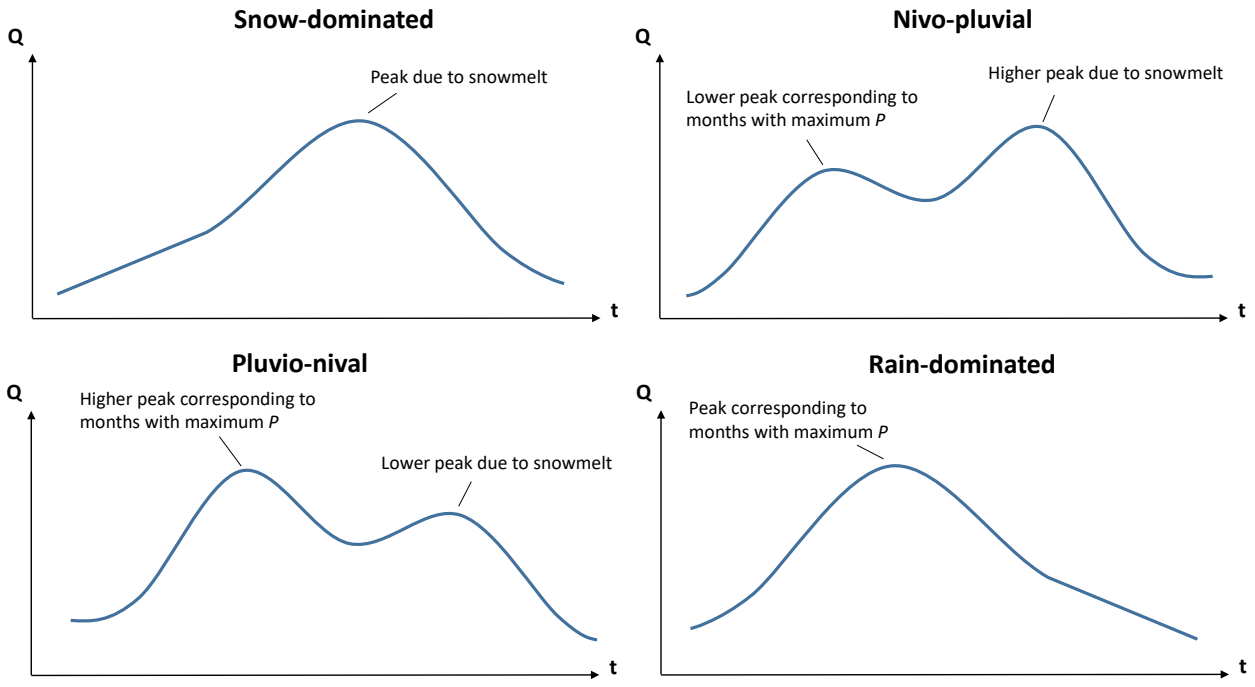

Figure A1. Conceptual illustration of the hydrological regimes used to classify the 100 near-natural catchments used in this study.

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.

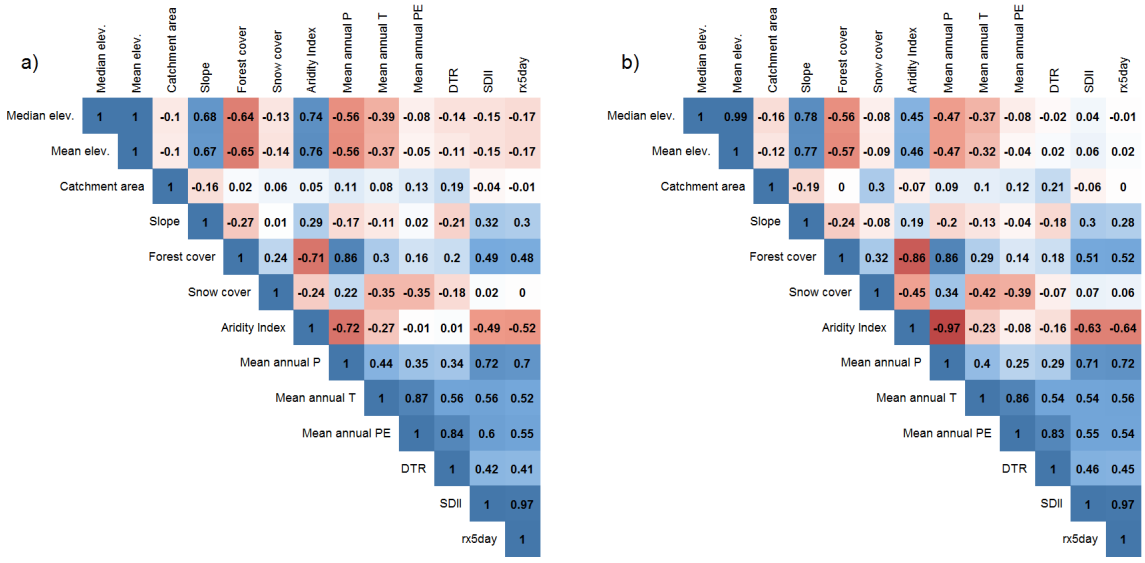

Figure B1. Correlation matrices of the catchment characteristics described in Table 4 using CR2MET as the P product for: a) the Pearson correlation, to evaluate linear correlation; and b) the Spearman correlation to evaluate the monotonic correlation.

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 (r)	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 catchment.

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

Author contributions.

Competing interests. No competing interests

Acknowledgements. We particularly thank Jim Freer, Juraj Parajka, Elena Toth, and one anonymous reviewer, whose constructive comments helped to improve the quality of the final manuscript. We would also like to thank the HESS editorial team for their support; the Centers for Natural Resources and Development (CNRD) PhD program for their financial support to the main author; the CAMELS-CL dataset (http://camels.cr2.cl/); Camila Álvarez-Garretón for providing an initial dataset of catchment that could be considered as *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. Dr. Zambrano-Bigiarini thanks Conicyt-Fondecyt 11150861 "Understanding the relationship between the spatio-temporal characteristics of meteorological drought and the availability of water resources, by using satellite-based rainfall and snow-cover data. A case study in a data-scarce Andean Chilean catchment" for the financial support from 2016 to 2018. Pablo Mendoza received support from Fondecyt Project 11200142. The authors are also grateful to the active R community for unselfish and prompt support, in particular to Robert J. Hijmans, and Alberto Viglione / Juraj Parajka for developing and maintaining the `raster` and `TUWmodel` R packages, respectively.

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, Geoscientific Model Development, 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, Hydrology and Earth System Sciences, 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 Resources Research, 54, 8792–8812, 2018.

Adhikary, S. K., Yilmaz, A. G., and Muttil, N.: Optimal design of rain gauge network in the Middle Yarra River catchment, Australia, Hydrological processes, 29, 2582–2599, 2015.

Alvarez-Garreton, 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, Hydrology and Earth System Sciences, 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 Resources Research, 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, Natural Hazards, 43, 167, https://doi.org/10.1007/s11069-007-9121-6, 2007.

Astagneau, P. C., Thirel, G., Delaigue, O., Guillaume, J. H., Parajka, J., Brauer, C. C., Viglione, A., Buytaert, W., and Beven, K. J.: Hydrology modelling R packages–a unified analysis of models and practicalities from a user perspective, Hydrology and Earth System Sciences, 25, 3937–3973, 2021.

Athira, P., Sudheer, K., Cibin, R., and Chaubey, I.: 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, 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, Atmospheric Research, 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 Sensing of Environment, 239, 111 606, 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, 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, Journal of Hydrology, 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 Resources Research, 52, 3599–3622, 2016.

Beck, H. E., Van Dijk, A. I., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., and Roo, A. d.: 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, 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, Hydrology and Earth System Sciences, 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, Scientific data, 5, 180 214, 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, Bulletin of the American Meteorological Society, 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, Journal of Geophysical Research: Atmospheres, 125, e2019JD031 485, 2020a.

Beck, H. E., Wood, E. F., McVicar, T. R., Zambrano-Bigiarini, M., Alvarez-Garreton, 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, Journal of Climate, 33, 1299–1315, 2020b.

Bergström, S.: Development and application of a conceptual runoff model for Scandinavian catchments, 1976.

Bergström, S.: The HBV model, Computer models of watershed hydrology, 1995.

Beven, K. J.: Changing ideas in hydrology - The case of physically-based models, Journal of Hydrology, 105, 157–172, https://doi.org/10.1016/0022-1694(89)90101-7, 1989.

Beven, K. J.: Uniqueness of place and process representations in hydrological modelling, Hydrology And Earth System Sciences, 4, 203–213, 2000.

Beven, K. J.: A manifesto for the equifinality thesis, Journal of Hydrology, 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, Journal of Hydrology: 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, Journal of Hydrometeorology, 13, 338–350, https://doi.org/10.1175/2011JHM1292.1, 2012.

Bivand, R. and Rundel, C.: rgeos: Interface to Geometry Engine - Open Source ('GEOS'), https://CRAN.R-project.org/package=rgeos, r package version 0.5-3, 2020.

Bivand, R., Keitt, T., and Rowlingson, B.: rgdal: Bindings for the 'Geospatial' Data Abstraction Library, https://CRAN.R-project.org/package=rgdal, r package version 1.5-12, 2020.

Boisier, J. P., Rondanelli, R., Garreaud, R. D., and Muñoz, F.: Anthropogenic and natural contributions to the Southeast Pacific precipitation decline and recent megadrought in central Chile, Geophysical Research Letters, 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 Cabauw polder, Hydrology and Earth System Sciences Discussions, Volume 11, Issue 2, 2014, pp.2091-2148, 11, 2091–2148, https://doi.org/10.5194/hessd-11-2091-2014, 2014a.

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, Hydrology and Earth System Sciences, Volume 18, Issue 10, 2014, pp.4007-4028, 18, 4007–4028, https://doi.org/10.5194/hess-18-4007-2014, 2014b.

Breiman, L.: Random Forests, Machine Learning, 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, Hydrology and Earth System Sciences, 15, 3411–3430, 2011.

Ceola, S., Arheimer, B., Baratti, E., Blöschl, G., Capell, R., Castellarin, A., Freer, J., Han, D., Hrachowitz, M., Hundecha, Y., et al.: Virtual laboratories: new opportunities for collaborative water science, Hydrology and Earth System Sciences, 19, 2101–2117, 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, International journal of applied earth observation and geoinformation, 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, Journal of Hydrometeorology, 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 Resources Research, 44, 2008.

Clerc, M.: From theory to practice in particle swarm optimization, in: Handbook of Swarm Intelligence, pp. 3–36, Springer, 2011a.

Clerc, M.: Standard particle swarm optimisation from 2006 to 2011, Particle Swarm Central, 253, 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, Hydrology and Earth System Sciences, 20, 3549–3560, https://doi.org/10.5194/hess-20-3549-2016, 2016.

Dallery, D., Squividant, H., De Lavenne, A., Launay, J., and Cudennec, C.: An end-user-friendly hydrological Web Service for hydrograph prediction in ungauged basins, Hydrological Sciences Journal, pp. 1–9, 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 resources research, 56, 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, https://snia.mop.gob.cl/sad/ADM600v1.pdf, 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, https://snia.mop.gob.cl/sad/CUH2087.pdf, 1999.

DGA: Cuenca Quebrada de Tarapacá, Tech. rep., Dirección General de Aguas, Santiago, https://mma.gob.cl/wp-content/uploads/2017/12/Tarapaca.pdf, 2004a.

DGA: Cuenca Río Loa, Tech. rep., Dirección General de Aguas, Santiago, https://mma.gob.cl/wp-content/uploads/2017/12/Loa.pdf, 2004b.

DGA: Cuenca del Río Elqui, Tech. rep., Dirección General de Aguas, Santiago, https://mma.gob.cl/wp-content/uploads/2017/12/Elqui.pdf%0A, 2004c.

DGA: Evaluación de los recursos hídricos superficales de las cuencas de los ríos Petorca y La Ligua V Región, Tech. rep., Dirección General de Aguas, Santiago, https://snia.mop.gob.cl/sad/SUP4496.pdf, 2006.

DGA: Análisis integral de soluciones a la escasez hídrica, región de Arica y Parinacota : informe final, Tech. rep., Dirección General de Aguas, Santiago, https://snia.mop.gob.cl/sad/REH5720.pdf, 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, https://snia.mop.gob.cl/sad/SUB5698.pdf, 2016b.

DGA: Herramientas de gestión y actualización de los modelos numéricos del acuífero de Copiapó : informe final, Tech. rep., Dirección General de Aguas, Santiago, https://snia.mop.gob.cl/sad/SUB5851v1.pdf, 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, Hydrological Processes, 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, pp. 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?, Journal of Hydrometeorology, 15, 1384–1403, 2014.

Fernandez, W., Vogel, R., and Sankarasubramanian, A.: Regional calibration of a watershed model, Hydrological sciences journal, 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, Journal of Hydrology, 595, 126 047, 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, Journal of Hydrology, 525, 383–399, 2015.

Garcia, F., Folton, N., and Oudin, L.: Which objective function to calibrate rainfall–runoff models for low-flow index simulations?, Hydrological sciences journal, 62, 1149–1166, 2017.

Garreaud, R. D., Alvarez-Garreton, C., Barichivich, J., Pablo Boisier, J., 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, 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 Journal of Climatology, 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, p. e1487, 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, Journal of Hydrology, 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, Journal of Natural Resources and Development, 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). no. fiche no. 85-2517., 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, 2018.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, 2020.

Hijmans, R. J.: raster: Geographic Data Analysis and Modeling, https://CRAN.R-project.org/package=raster, 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, Climate dynamics, 35, 841–858, 2010.

Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., Arheimer, B., Blume, T., Clark, M., Ehret, U., et al.: A decade of Predictions in Ungauged Basins (PUB)—a review, Hydrological sciences journal, 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, Journal of hydrology, 577, 123 585, 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 Resources Research, 57, e2020WR029 143, 2021.

Jarvis, A., Reuter, H. I., Nelson, A., Guevara, E., et al.: Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database (http://srtm. csi. cgiar. org), 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, Hydrology and Earth System Sciences, 24, 1081–1100, 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, pp. 3–7, Springer, 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, Journal of Hydrology, 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, vol. 4, pp. 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, Journal of Hydrology, 424, 264–277, https://doi.org/10.1016/j.jhydrol.2012.01.011, 2012.

Knoben, W. J., Freer, J. E., and Woods, R. A.: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, Hydrology and Earth System Sciences, 23, 4323–4331, 2019.

Koffler, D., Gauster, T., and Laaha, G.: lfstat: Calculation of Low Flow Statistics for Daily Stream Flow Data, https://CRAN.R-project.org/package=lfstat, r package version 0.9.4, 2016.

Kuentz, A., Arheimer, B., Hundecha, Y., and Wagener, T.: Understanding hydrologic variability across Europe through catchment classification, Hydrology and Earth System Sciences, 21, 2863–2879, 2017.

Kundu, D., Vervoort, R. W., and van Ogtrop, F. F.: The value of remotely sensed surface soil moisture for model calibration using SWAT, Hydrological Processes, 31, 2764–2780, https://doi.org/10.1002/hyp.11219, 2017.

Lagos, M., Mendoza, P., Rondanellu, 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, 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, Journal of Geophysical Research: Atmospheres, 99, 14 415–14 428, 1994.

Liaw, A. and Wiener, M.: Classification and Regression by randomForest, R News, 2, 18–22, https://CRAN.R-project.org/doc/Rnews/, 2002.

Lindström, G.: A simple automatic calibration routine for the HBV model, Hydrology Research, 28, 153–168, 1997.

Maggioni, V. and Massari, C.: On the performance of satellite precipitation products in riverine flood modeling: A review, Journal of Hydrology, 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, Journal of Hydrometeorology, 14, 1194–1211, 2013.

McIntyre, N., Lee, H., Wheater, H., Young, A., and Wagener, T.: Ensemble predictions of runoff in ungauged catchments, Water Resources Research, 41, 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, Hydrology and Earth System Sciences, 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?, Hydrological Processes, 30, 1071–1095, 2016.

Merz, R. and Blöschl, G.: Regionalisation of catchment model parameters, Journal of hydrology, 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, 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, Hydrology and Earth System Sciences, 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, Journal of Hydrometeorology, 14, 171–185, 2013.

Ollivier, C., Mazzilli, N., Olioso, A., Chalikakis, K., Carrière, S. D., Danquigny, C., and Emblanch, C.: Karst recharge-discharge semi distributed model to assess spatial variability of flows, Science of the Total Environment, 703, 134 368, 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 ungaged catchments: A comparison of regionalization approaches based on 913 French catchments, Water Resources Research, 44, 2008.

Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalisation methods for catchment model parameters, 2005.

Parajka, J., Merz, R., and Blöschl, G.: Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments, Hydrological Processes: An International Journal, 21, 435–446, 2007.

Parajka, J., Viglione, A., Rogger, M., Salinas, J., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins–Part 1: Runoff-hydrograph studies, Hydrology and Earth System Sciences, 17, 1783–1795, 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, Hydrology and Earth System Sciences, 20, 2085, 2016.

Perpiñán, O. and Hijmans, R.: rasterVis, https://oscarperpinan.github.io/rastervis/, 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, Journal of Hydrology, 418, 49–60, 2012.

Pool, S., Vis, M. J., Knight, R. R., and Seibert, J.: Streamflow characteristics from modeled runoff time series–importance of calibration criteria selection, Hydrology and Earth System Sciences, 21, 5443–5457, 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, https://www.R-project.org/, 2020.

Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., Attinger, S., Schäfer, D., Schrön, M., and Samaniego, L.: Multiscale and multivariate evaluation of water fluxes and states over European river basins, Journal of Hydrometeorology, 17, 287–307, 2016.

Ren, Z. and Li, M.: Errors and correction of precipitation measurements in China, Advances in Atmospheric Sciences, Volume 24, Issue 3, pp.449-458, 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, 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 Environmental 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, 2019.

Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resources Research, 46, 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, Hydrology and Earth System Sciences, 15, 2895–2911, 2011.

Sevruk, B., Ondrás, M., and Chvíla, B.: The WMO precipitation measurement intercomparisons, Atmospheric Research, 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 Resources Research, 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, Hydrology and Earth System Sciences, 22, 1831–1849, 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, Hydrological Processes, 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, Journal of Hydrology, p. 124639, 2020.

Stisen, S. and Sandholt, I.: Evaluation of remote-sensing-based rainfall products through predictive capability in hydrological runoff modelling, Hydrological Processes, 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, Reviews of Geophysics, 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, Journal of Hydrologic Engineering, 22, 04017 010, 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 Resources Research, p. e2019WR026153, 2020.

Tarek, M., Brissette, F. P., and Arsenault, R.: Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America, Hydrology and Earth System Sciences, 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, Journal of Hydrology, 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, Hydrological Sciences Journal, 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, Hydrological Sciences Journal, 63, 1133–1149, 2018.

Vandewiele, G. and Elias, A.: Monthly water balance of ungauged catchments obtained by geographical regionalization, Journal of hydrology, 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, pp. 189–208, Springer, 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, Journal of Applied Meteorology and Climatology, 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 System Dynamics, 6, 2015.

Viglione, A. and Parajka, J.: TUWmodel: Lumped/Semi-Distributed Hydrological Model for Education Purposes, https://CRAN.R-project.org/package=TUWmodel, 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 Resources Research, 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?", Stochastic Environmental Research and Risk Assessment, 23, 1061–1062, https://doi.org/10.1007/s00477-008-0284-9, 2009.

Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V., and Sorooshian, S.: A framework for development and application of hydrological models, Hydrology and Earth System Sciences, 5, 13–26, 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, Journal of Hydrology, 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), International Journal of Climatology, 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?, Journal of Hydrology, 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 Resources Research, 55, 7784–7803, 2019.

Yapo, P. O., Gupta, H. V., and Sorooshian, S.: Multi-objective global optimization for hydrologic models, Journal of Hydrology, 204, 83–97, 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, Journal of Hydrology, 320, 155–172, 2006.

Zambrano-Bigiarini, M.: hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series, https://doi.org/10.5281/zenodo.839854, https://github.com/hzambran/hydroGOF, r package version 0.4-0, 2020a.

Zambrano-Bigiarini, M.: hydroTSM: Time Series Management, Analysis and Interpolation for Hydrological Modelling, https://github.com/hzambran/hydroTSM, r package version 0.6-0 . doi: https://doi.org/10.5281/zenodo.83964, 2020b.

Zambrano-Bigiarini, M. and Baez-Villanueva, O.: Tutorial for using hydroPSO to calibrate TUWmodel, https://doi.org/10.5281/zenodo.3772176, https://doi.org/10.5281/zenodo.3772176, 2020.

Zambrano-Bigiarini, M. and Rojas, R.: A model-independent Particle Swarm Optimisation software for model calibration, Environmental Modelling & Software, 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, Hydrology and Earth System Sciences, 21, 1295, 2017.

Zambrano-Bigiarini, M., Baez-Villanueva, O. M., and Giraldo-Osorio, J.: RFmerge: Merging of Satellite Datasets with Ground Observations using Random Forests, https://CRAN.R-project.org/package=RFmerge, r package version 0.1-10 . doi:10.5281/zenodo.3581515, 2020.

Zelelew, M. B. and Alfredsen, K.: Transferability of hydrological model parameter spaces in the estimation of runoff in ungauged catchments, Hydrological Sciences Journal, 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, Science of The Total Environment, 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, Journal of Hydrology, 594, 125 969, 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 Resources Research, 45, 2009.

Zhang, Y. and Wang, K.: Global precipitation system size, Environmental Research Letters, 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, Journal of Hydrology, 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 Sensing of Environment, 183, 170–185, https://doi.org/10.1016/j.rse.2016.05.016, 2016.