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

Preprint · April 2021

DOI: 10.5194/hess-2021-156

9 authors, including:

Oscar Manuel Baez-Villanueva
Technische Hochschule Köln
14 PUBLICATIONS **228** CITATIONS
SEE PROFILE

Mauricio Zambrano-Bigiarini
Universidad de La Frontera
56 PUBLICATIONS **1,882** CITATIONS
SEE PROFILE

Pablo A. Mendoza
University of Chile
47 PUBLICATIONS **787** CITATIONS
SEE PROFILE

Ian McNamara
Forschungszentrum Jülich
8 PUBLICATIONS **75** CITATIONS
SEE PROFILE

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

Hydrology and Earth System Sciences
Discussions

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 years, 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 individual catchment calibration and verification, as well as the relative performance of three well-known regionalisation techniques (spatial proximity, feature similarity, and parameter regression) over 100 near-natural catchments with diverse hydrological regimes across Chile. We configured and calibrated a conceptual semi-distributed HBV-like hydrological model for each catchment, using four P products (ERA5, MSWEPv2.8, RF-MEPv2, and CR2MET), and two objective functions. The three regionalisation techniques were applied and evaluated for each combination of P product and objective function, using a leave-one-out cross-validation procedure. Despite differences in the spatio-temporal distribution of P quantities, all P 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 results (with median KGE's for the best method ranging from 0.56 to 0.63). Our results suggest that model calibration is able to compensate, to some extent, differences between forcing datasets, and that the spatial resolution of P products does not substantially affect the regionalisation performance. Overall, feature similarity provided the best results, followed closely by spatial proximity, while parameter regression performed the worst, thus reinforcing the importance of transferring complete parameter sets to ungauged catchments. Our results suggest that: i) merging P products and ground-based measurements does not necessarily translate into an improved hydrological modelling performance; ii) a P product that provides the best individual model performance during calibration and verification does not necessarily provide the best performance in terms of parameter regionalisation; and iii) the hydrological regime affects the performance of regionalisation methods, with rain-dominated catchments with a snow component performing the best over Chile for spatial proximity and feature similarity.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

Hydrology and Earth System Sciences
Discussions

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 (Duan et al., 1992; 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 are 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), therefore 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 with 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; Guo et al., 2021; Neri et al., 2020).

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021


Hydrology and Earth System Sciences
Discussions
Open Access
EGU

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 (Villarini and Krajewski, 2008; 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?

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

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 / Jack-knife 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 a 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 / Jack-knife 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 / Jack-knife 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 / Jack-knife cross-validation	SP and FS	FS outperformed SP. They reported only a small decrease of performance from calibration/verification to regionalisation (∼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 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 the median of KGE values 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.
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.
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.
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 hydrological regime.

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021


Hydrology and Earth System Sciences
Discussions
Open Access
EGU

2 Study area and selection of catchments

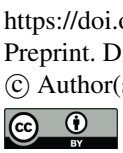
Our study domain is continental Chile (Figure 1), which is bounded to the west by the Pacific Ocean, to the north by Peru, and to the east by Bolivia and Argentina. The territory spans 4300 km of latitudinal extension (17.5°S–56.0°S) and on average 180 km of longitudinal extension (76.0°W–66.0°W), with elevation (Jarvis et al., 2008) ranging from 0 to 6892 m a.s.l. in the Andean Mountains. Figure 1 shows the elevation, land cover (Zhao et al., 2016), hydrological regime (classified through visual screening of the mean monthly Q), and the Köppen-Geiger climate classification (Beck et al., 2018). A large variety of climates are present across the country, transitioning from the (hyper)arid and semi-arid climates in the north, through temperate climates in central Chile, to more humid and polar climates in the south. P increases with altitude and latitude (in the southern direction) ranging from almost zero in the Atacama Desert to ∼6000 mm yr^{-1} in the surroundings of Puerto Cardenas (∼43.2°S). Similar to the P patterns, 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 (Garreaud et al., 2017; Boisier et al., 2016), 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, 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/) 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 the 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 $< 20\%$).
8. No signs of artificial regulation in the hydrograph (10 excluded in total).

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021



Hydrology and Earth System Sciences
Discussions
Open Access
EGU

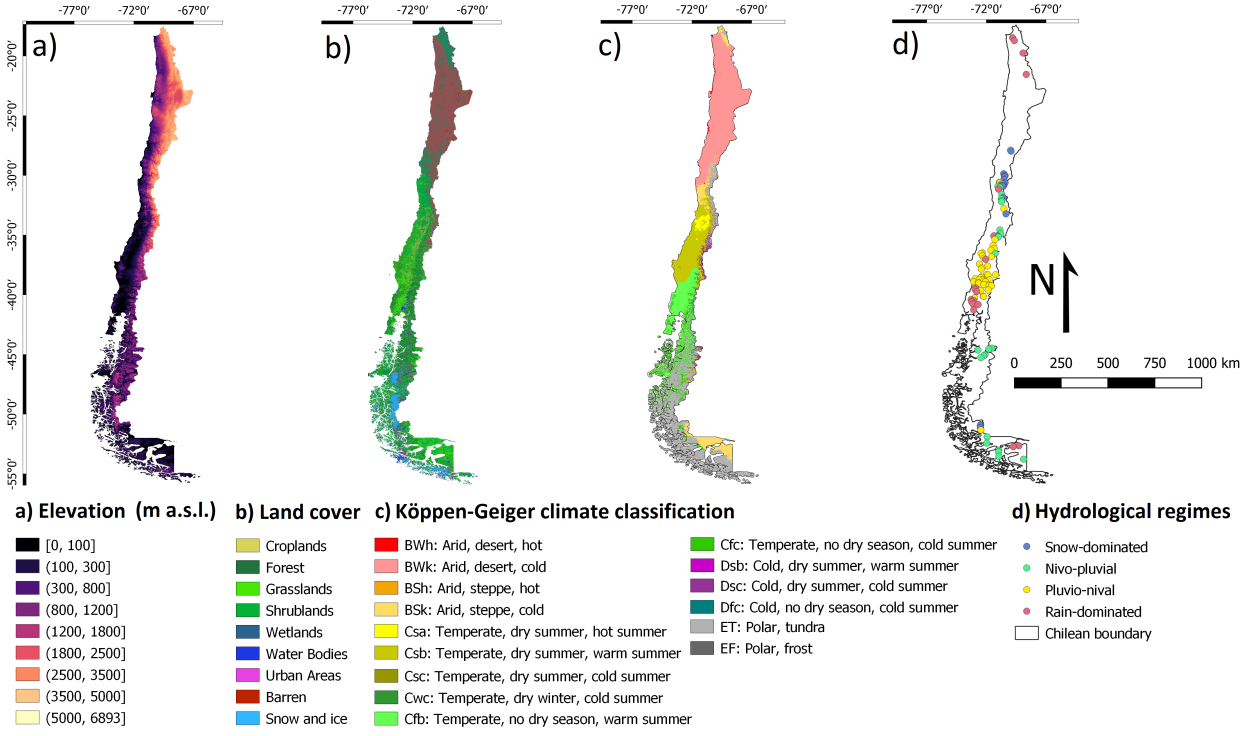


Figure 1. Study area: a) elevation (SRTMv4.1; Jarvis et al., 2008); b) land cover classification (Zhao et al., 2016); c) Köppen-Geiger climate classification (Beck et al., 2018); and d) hydrological regimes of the selected catchments.

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). Through visual screening, these selected catchments were classified by hydrological regimes as i) snow-dominated, ii) nivo-pluvial, i.e., snow-dominated with a lower rain component, iii) pluvio-nival, i.e, rain-dominated with a lower snow component, and iv) rain-dominated, as shown in Figure 1d.

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 with *in situ* measurements over continental Chile (Zambrano-Bigiarini et al., 2017; Boisier et al., 2018; Baez-Villanueva et al., 2018,

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021


Hydrology and Earth System Sciences
Discussions

2020). We included the coarse-resolution ERA5 product because i) Chile is dominated by large-scale, frontal systems (Zhang and Wang, 2021) and therefore, coarse-resolution products may perform well over small catchments; ii) reanalysis products tend perform well at high latitudes (Beck et al., 2017a); 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).

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	1990–2018	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 1979–2019. These estimates are produced by combining rain gauge observations with reanalysis data from ERA5 (Boisier et al., 2018).

The random forest merging technique (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 331 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). 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).

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 and accounts for their local reporting times (Beck et al., 2019). The 3-hourly MSWEPv2.8 estimates were also aggregated into daily P to account for the difference in the reporting times.

Figure 2 shows the mean annual P patterns for all products over 1990–2018. All P products show relatively similar patterns of spatial variation across continental Chile; however, there are some marked differences in the total P amounts. In general, P increases from the (hyper-arid) northern Chile to the south, as well as from the west coast into the Andes Mountains.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021


Hydrology and Earth System Sciences
Discussions

Over the arid region (17–30°S), ERA5 presents a mean annual P of 420 mm, which is four times the value of the second-highest product CR2MET (106 mm). Over this region, RF-MEP and MSWEPv2.8 present similar mean annual P values (61 and 66 mm, respectively). For central Chile (30–40°S), CR2MET has the highest mean annual P (889 mm), followed by ERA5 and RF-MEP (646 and 599 mm, respectively), while MSWEPv2.8 has the lowest mean value (446 mm). Finally, for latitudes lower than 40°S, CR2MET has the highest mean annual P (1,397 mm), followed by ERA5 (1,313 mm), MSWEPv2.8 (1,034 mm) and RF-MEP (727 mm).

3.1.2 Air temperature and potential evaporation

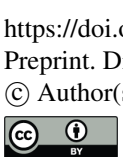
Maximum and minimum daily air temperature (T) at a spatial resolution of 0.05° were taken from CR2MET. T is estimated using multivariate regression from the Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature (LST) and ERA5 estimates as covariates (Alvarez-Garreton et al., 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., Ceola et al., 2015; Parajka et al., 2016; Sleziak et al., 2016; Zessner et al., 2017; Melsen et al., 2018; Nijzink et al., 2018; Sleziak et al., 2020; Széles 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; Jin et al., 2009; Singh et al., 2012; Wallner et al., 2013; Zelelew and Alfredsen, 2014; Beck et al., 2016; Neri et al., 2020); and ii) catchments with diverse hydroclimatic and geomorphological characteristics (Merz and Blöschl, 2004; Driessen et al., 2010; Samuel et al., 2011; 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, that were adopted from previous studies (Parajka et al., 2007; Ceola et al., 2015), which 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 (H_{max}) 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.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021



Hydrology and Earth System Sciences
Discussions
Open Access
EGU

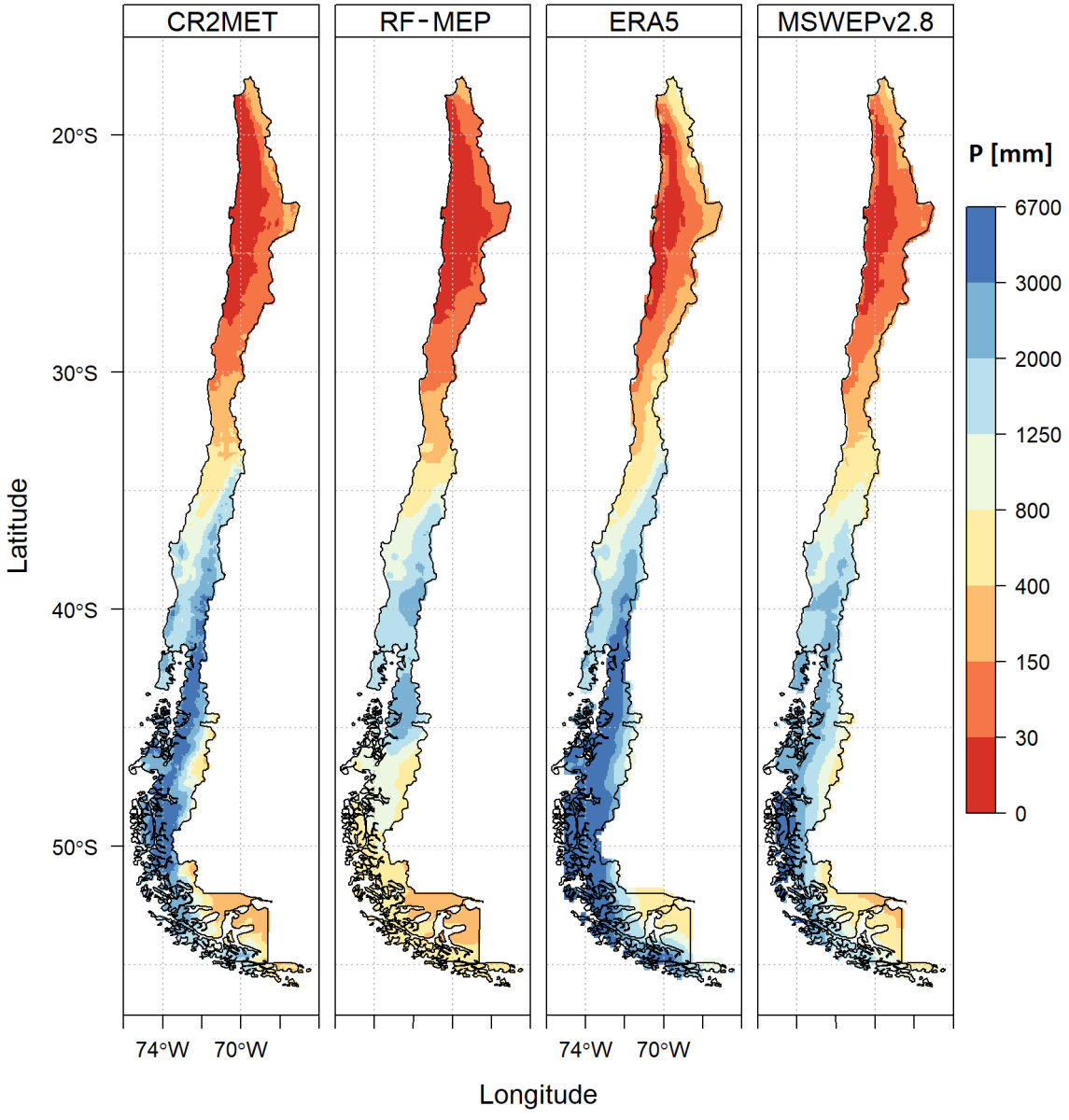


Figure 2. Mean annual P for each product over 1990–2018, resampled to a 0.05° spatial resolution using the nearest neighbour method.

3.3 Individual catchment calibration and verification

The simulation period used for this study was 1990–2018, using the first ten years as a warm-up period. The calibration period (2000–2014) includes near normal conditions and the beginning of the Chilean megadrought. The first evaluation period

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

Hydrology and Earth System Sciences
Discussions

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

(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 examine how the goodness-of-fit function used for calibration affected the regionalisation results, we considered two different objective functions to calibrate the TUWmodel. The first function was the modified Kling-Gupta efficiency (KGE', Eq. 1; Kling et al., 2012), 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).

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

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021


Hydrology and Earth System Sciences
Discussions

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

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

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

Based on previous research (Shafii and Tolson, 2015; Vis et al., 2015; Beck et al., 2016), we also implemented a new aggregated objective function (AOF) that combines Q signatures (AOF_{sig}) and classical goodness-of-fit indices (AOF_{gof}):

$$\text{AOF} = \frac{\text{AOF}_{gof} + \text{AOF}_{sig}}{2} \tag{5}$$

For the AOF_{gof} (Eq. 6), we used the average of the KGE' recommended by Garcia et al. (2017), which is computed for both the Q and its inverse (KGE'(Q) and KGE'(1/Q), respectively). The use of this function should lead to improved estimates of mean annual runoff, seasonality, and low flow indices. To avoid problems with zero Q values while computing KGE'(1/Q) (Santos et al., 2018), we added a small constant to all Q values, defined as one-hundredth of the mean daily Q (Pushpalatha et al., 2012; Garcia et al., 2017).

$$\text{AOF}_{gof} = \frac{\text{KGE'}(Q) + \text{KGE'}(1/Q)}{2} \tag{6}$$

Similar to Yilmaz et al. (2008) and Beck et al. (2016), we selected four Q signatures to calculate AOF_{sig} (Eq. 7): the difference between the observed and simulated baseflow indices ($eBFI$), which may be thought of as the proportion of river runoff deriving from stored sources; and the percent bias in the slope of the lower, medium, and higher segments of the flow duration curve ($pbiasFDC_{Low}$, $pbiasFDC_{Med}$, $pbiasFDC_{High}$, respectively). Equations 8a–d present the four aforementioned components of AOF_{sig}, where BFI represents the base flow index; sim and obs represent the simulated and observed Q values, respectively; and Q_{99}, Q_{70}, Q_{20}, Q_{10}, Q_1 represent streamflows with 99, 70, 20, 10 and 1% probability of exceedance, respectively.

$$AOF_{sig} = 1 - (eBFI + pbiasFDC_{Low} + pbiasFDC_{Med} + pbiasFDC_{High}) \tag{7}$$

$$eBFI = |BFI_{sim} - BFI_{obs}| \tag{8a}$$

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021


Hydrology and Earth System Sciences
Discussions

$$pbiasFDC_{Low} = 100 \cdot \left| \frac{log(Q_{70,s}) - log(Q_{99,s})}{log(Q_{70,o}) - log(Q_{99,o})} - 1 \right| \tag{8b}$$

$$pbiasFDC_{Med} = 100 \cdot \left| \frac{log(Q_{20,s}) - log(Q_{70,s})}{log(Q_{20,o}) - log(Q_{70,o})} - 1 \right| \tag{8c}$$

$$pbiasFDC_{High} = 100 \cdot \left| \frac{Q_{1,s} - Q_{10,s}}{Q_{1,o} - Q_{10,o}} - 1 \right| \tag{8d}$$

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). 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). The default hydroPSO settings were used for all calibrations (Zambrano-Bigiarini and Rojas, 2013).

3.4 Regionalisation techniques

After obtaining catchment-specific model parameters through individual catchment calibration using the KGE' and AOF separately, 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):

$$\mathrm{ED}_{ij} = \sqrt{\sum_{k=1}^{n} (x_{k,i} - x_{k,j})^2} \tag{9}$$

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

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021


Hydrology and Earth System Sciences
Discussions

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 A). 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. 10:

$$Z_f = \frac{x_f - x_{min}}{x_{max} - x_{min}} \tag{10}$$

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:

$$S_{i,j} = \sum_{m=1}^{n} \mid Z_{i,m} - Z_{j,m} \mid \tag{11}$$

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 (McIntyre et al., 2005; Oudin et al., 2008; Zhang and Chiew, 2009; Zhang et al., 2015; Beck et al., 2016). The full parameter sets obtained during the individual 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.

Finally, to quantify the overall uncertainty in model simulations when different objective functions are used for regionalisation, we adapted the P- and R-factor concepts described in Abbaspour et al. (2007); Schuol et al. (2008); Abbaspour et al. (2009) to be calculated using only 10 Q simulations. We defined the uncertainty band as the area between the maximum and minimum Q values for all time steps, obtained from the ensemble of the 10 simulations. Two indices were used to quantify the uncertainty of model simulations with respect to observations: $i)$ P_{reg}, which is defined as the percentage of Q observations bracketed by the uncertainty band; and $ii)$ R_{reg}, which represents the average width of the uncertainty band divided by the standard deviation of the observations. Ideally, we would capture all observations within the uncertainty band (i.e., $P_{reg} = 1$), with no uncertainty (i.e., $R_{reg} = 0$); however, a more realistic good result would be to achieve $P_{reg} \sim 1$ with the same spread as the observations (i.e., $R_{reg} \sim 1$).

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

Hydrology and Earth System Sciences
Discussions

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.

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

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 pre-

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021


Hydrology and Earth System Sciences
Discussions

dictand 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; Beck et al., 2020b; Baez-Villanueva et al., 2020; 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 the set of independent catchment characteristics listed in Table 4 as covariates. Our experimental setup used an ensemble of 2,000 regression trees, a minimum of five terminal nodes for each model, and $p/3$ variables randomly sampled as candidates at each split, where p represents the number of predictors. The trained RF models were then used to predict parameter values in pseudo-ungauged catchments, using its climatic and physiographic descriptors as covariates.

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.

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 are considered to be nested.

4 Results

4.1 Performance of P products

4.1.1 Calibration and verification

Figure 3 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. These results were obtained using the KGE' (Figure 3a) and the AOF (Figure 3b) for the calibration process. When the KGE' was used as the objective function, 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. However, when the AOF was used as the objective function to drive calibration, ERA5 (median KGE' of 0.72) outperformed CR2MET (0.71) during

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

Verification 1. Surprisingly, MSWEPv2.8 provided the poorest performance for calibration and Verification 1, while ERA5 and RF-MEP performed the worst during Verification 2, when using the KGE' and AOF, respectively. For all P products, the lowest performances were obtained during the (dry) Verification 2 period, emphasising the challenges of estimating Q over dry environments, as discussed by Maggioni et al. (2013) and Beck et al. (2016). Despite variations between P products, TUWmodel performed well for all products in the calibration, Verification 1 and Verification 2 periods, with median KGE' (AOF) values greater than 0.77, 0.71, and 0.62 (0.78, 0.70, and 0.61), respectively. The calibrated model parameters were within the selected parameter ranges in the large majority of the cases (see Figures S4 and S5 of the supplement material).

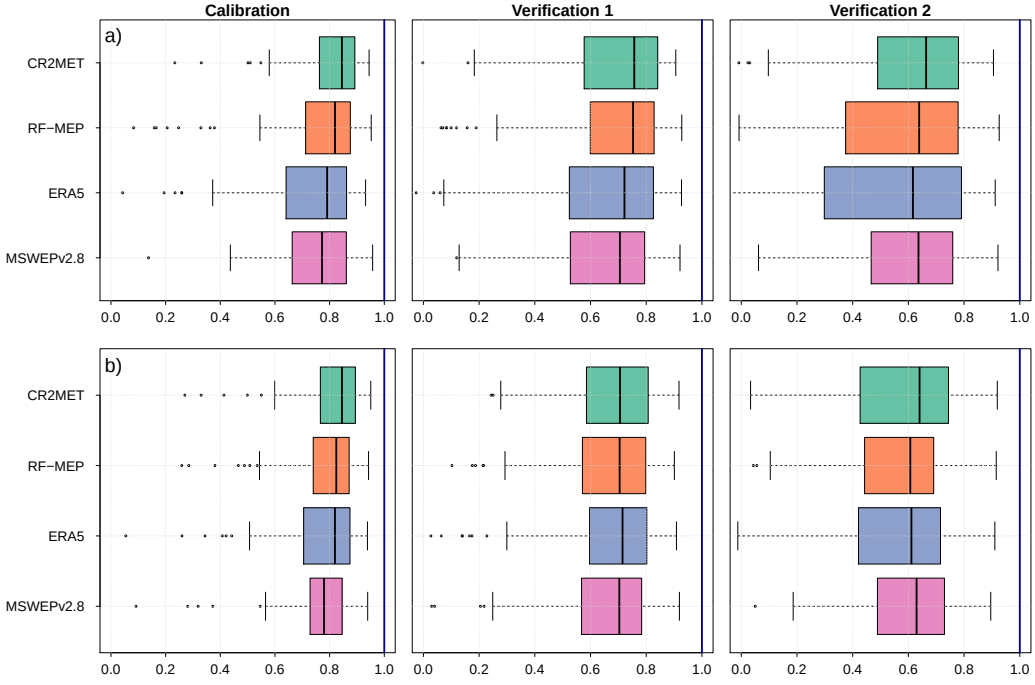


Figure 3. Performance of TUWmodel during the calibration (2000–2014), Verification 1 (1990–1999) and Verification 2 (2015–2018), prior to any regionalisation, using a) the modified Kling-Gupta efficiency (KGE'), and b) the aggregate objective function (AOF).

4.1.2 Performance during regionalisation

Figure 4 summarises the leave-one-out cross-validation results for the three tested regionalisation methods (spatial proximity, feature similarity, parameter regression), obtained when TUWmodel was forced with each P product over the entire simulation period (1990–2018). Overall, the median performance of all P products was the highest for feature similarity (KGE' and AOF $\sim$ 0.60), followed by spatial proximity (KGE' and AOF $\sim$ 0.55) and parameter regression (KGE' $\sim$ 0.35, AOF $\sim$ 0.25). In addition to exhibiting a considerably lower overall performance, parameter regression also resulted in a larger spread in performance.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

Hydrology and Earth System Sciences
Discussions

Focusing on only the two best-performing regionalisation methods and bearing in mind that KGE' and AOF have different theoretical grounds, the results presented on Figure 4 show that the overall performance obtained for feature similarity and spatial proximity are quite close for different P products, and for both objective functions. For example, all P products generate acceptable KGE' results (median KGE' > 0.56) for feature similarity, with the best model performance obtained by CR2METv2 (KGE' $= 0.63$), followed closely by RF-MEP (0.62). When using the AOF for feature similarity, the best model performance was again obtained by CR2METv2 (median AOF of 0.64), but this time followed closely by MSWEPv2.8 (0.62). MSWEPv2.8 slightly outperformed ERA5 (median KGE'=0.61) in feature similarity, which is consistent with the results obtained for spatial proximity. In the case of spatial proximity using the KGE', CR2MET performed the best, with a median KGE' of 0.57, followed by RF-MEP (0.56), while MSWEPv2.8 showed a slightly better KGE' performance than ERA5 (0.54 and 0.51, respectively). In the case of the AOF, MSWEPv2.8 showed a better AOF performance than RF-MEP (0.55 and 0.53, respectively).

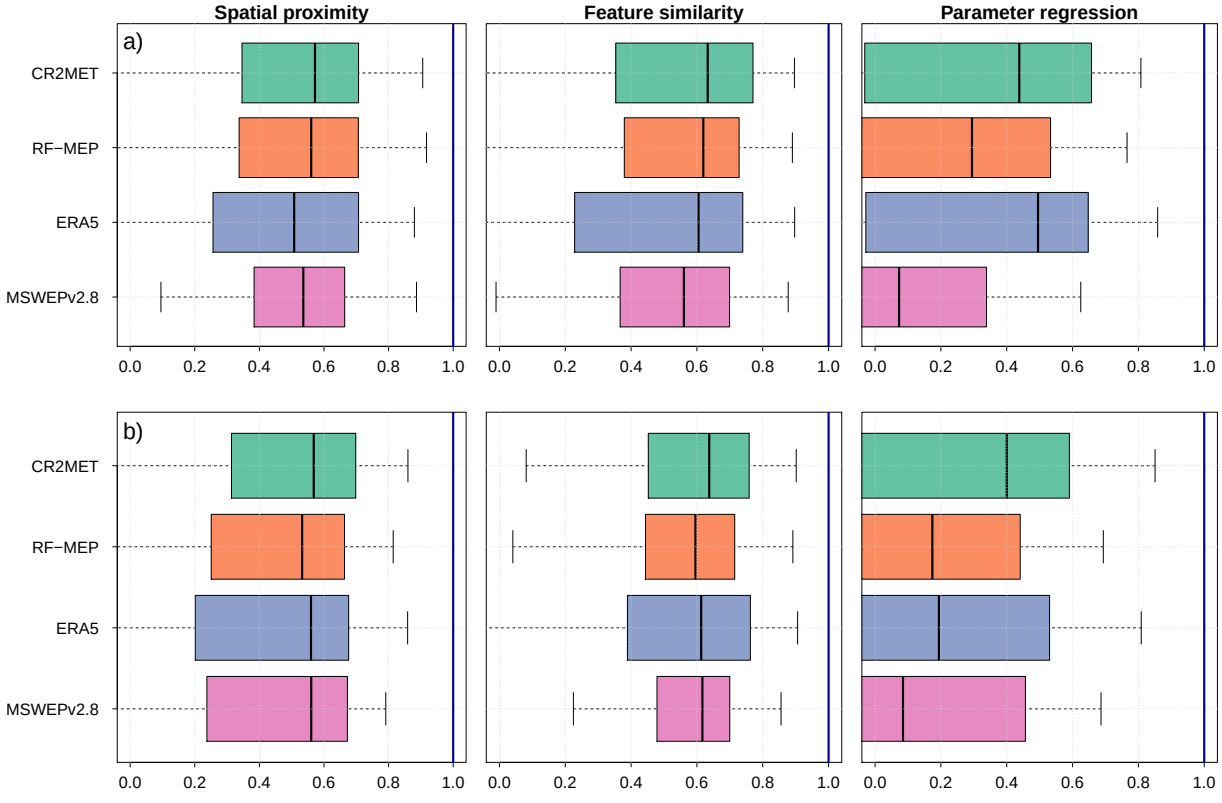


Figure 4. Leave-out-out cross-validation results for the three regionalisation methods according to P product used to force TUWmodel during 1990-2018, using a) the KGE', and b) the AOF.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021


Hydrology and Earth System Sciences
Discussions

For each regionalisation technique, Figure 5 summarises the spatial distribution of the performance of the four P products and both objective functions used in regionalisation. The spatial patterns obtained for all regionalisation methods were similar, independent of the P product or objective function, except for parameter regression using the AOF, which yielded worse results over high-elevation catchments. These results suggest that the choice of objective function used to drive calibration does not greatly impact the spatial performance of regionalisation methods that transfer complete sets of model parameters (feature similarity and spatial proximity).

All P products performed better in central (30-40°S) and southern Chile (40-56.5°S) than in the arid north (17-30°S). The low performance of regionalisation in the arid north is very likely due to the convective nature of storms occurring in the highlands of the Chilean Altiplano (elevations above 4,000 m a.s.l.), and the low density of Q stations over this area. Despite this general low performance, RF-MEP was the best performing P product over northern Chile (17-30°S) for both spatial proximity (median KGE' of 0.37) and feature similarity (median KGE' of 0.51), for the 10 most northern catchments (see Figure 1), suggesting that merging P products and ground-based observations helps improve, to some extent, the performance of hydrological modelling across arid regions. Conversely, all products outperformed RF-MEP over the far south. Figure 5 also highlights that spatial proximity provides the best performance over the far south, with median KGE' (AOF) values over the 10 most southern catchments of 0.47 (0.59), 0.45 (0.49), 0.25 (0.35), and 0.30 (0.35) for MSWEPv2.8, CR2MET, RF-MEP, and ERA5, respectively. The systematic lower performance of feature similarity compared to spatial proximity over the south could be attributed to lack of catchment characteristics that represent the hydrological behaviour of this complex area.

The panels located below each map in Figure 5 show the empirical cumulative distribution functions (ECDFs) of the performance of each regionalisation technique. These ECDFs compare the relative performance of each regionalisation method against those obtained from the individual calibration and verification of each catchment (used as benchmarks). As expected, all regionalisation methods presented a lower performance than the individual calibration and verification, with this reduction more pronounced for parameter regression. For feature similarity, all P products yielded a performance similar to that obtained for individual basin verification during the dry Verification 2 period.

4.2 Evaluation of regionalisation techniques

4.2.1 Overall performance

For each P product, Figure 6 compares the performances of the three regionalisation techniques using the AOF with those obtained in the individual calibration and verification periods. The individual 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). 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). Median AOFs obtained for feature similarity were the highest, followed closely by spatial proximity, regardless of the

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

Hydrology and Earth System Sciences
Discussions

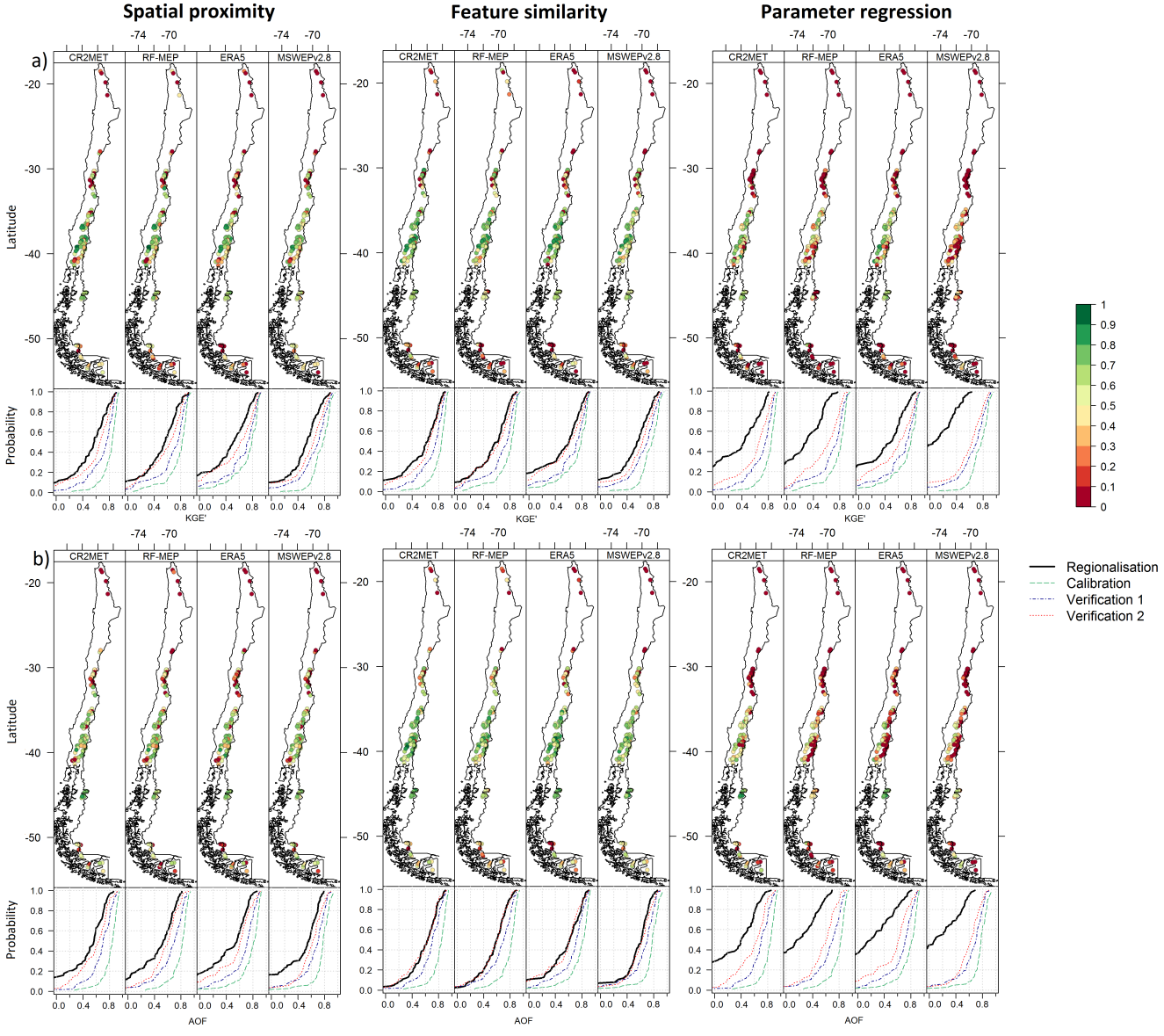
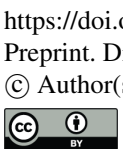


Figure 5. Spatial performance of the leave-one-out cross-validation results for the three regionalisation methods according to P product used to force TUWmodel (1990-2018) using a) the KGE' and b) the AOF. The panels beneath each map plot the ECDFs of the corresponding regionalisation technique and P product (black) against the performances during the calibration (blue), Verification 1 (purple), and Verification 2 (red) periods.

365 selected P product, but with a larger spread. Parameter regression had both the lowest median AOFs 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

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021



Hydrology and Earth System Sciences
Discussions
Open Access
EGU

for spatial proximity and parameter regression. Results obtained for the evaluation of regionalisation methods using the KGE' are presented in Figure S1 of the supplement material, and are similar to those for the AOF.

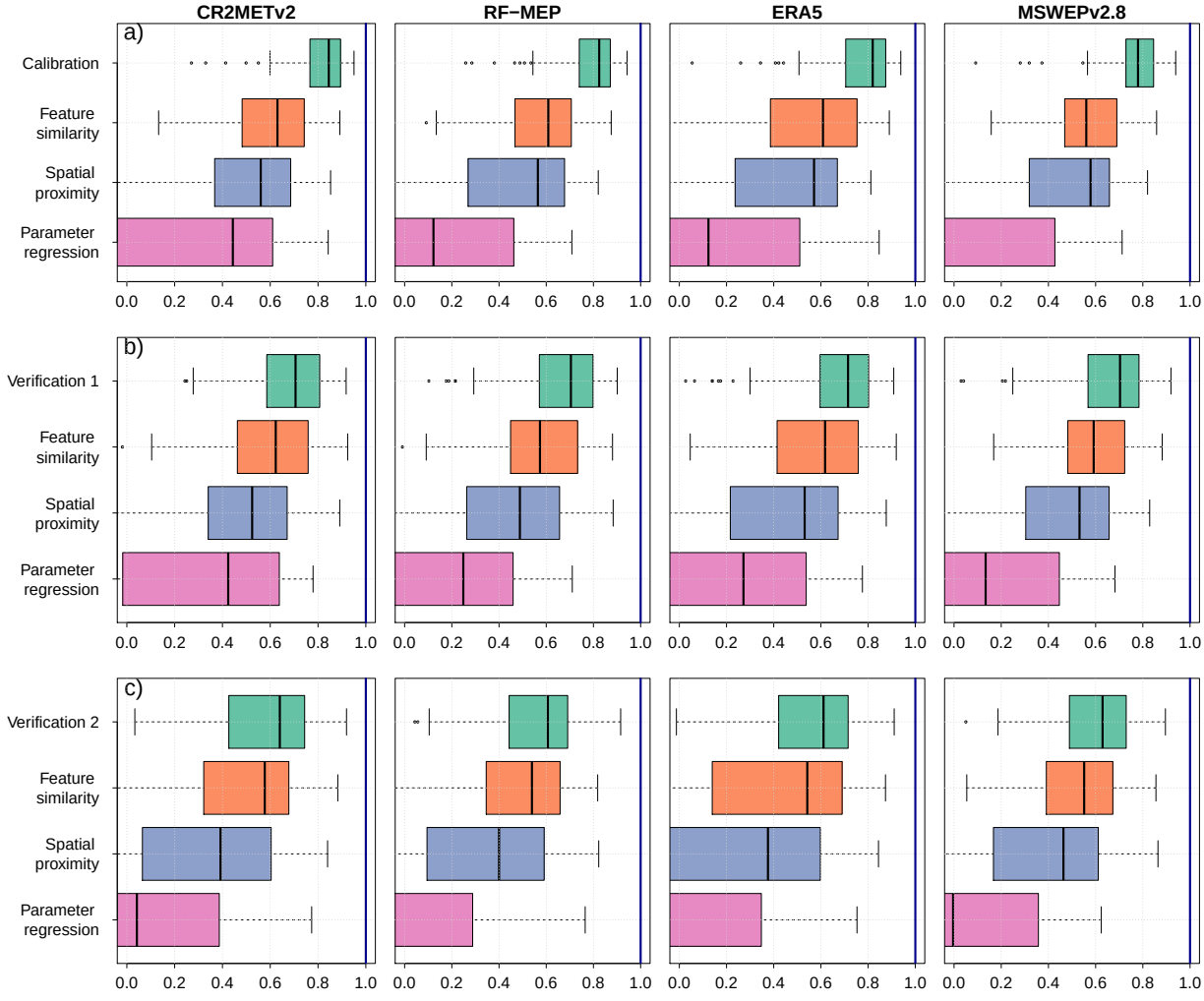


Figure 6. Performance of the regionalisation methods using the AOF objective function for: a) calibration (2000–2014); b) Verification 1 (1990–1999), and c) Verification 2 (2015–2018).

4.2.2 Impact of hydrological regimes

We also analysed the impact of the hydrological regime (see Figure 1d) on the performance of the three regionalisation methods. Figures 7 and 8 show the performance of the regionalisation techniques according to hydrological regime for all P products and the KGE' and AOF objective functions, respectively.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021


Hydrology and Earth System Sciences
Discussions

Figure 7 shows that when using the KGE', feature similarity provided the best median performance for all hydrological regimes and P products except snow-dominated catchments, where spatial proximity performed the best for CR2MET and ERA5. Similarly, Figure 8 shows that feature similarity performed better over all hydrological regimes, except for the case of snow-dominated catchments with CR2MET as the P product. 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, chosen regionalisation method, and objective function used to drive the hydrological model calibration. For feature similarity in snow-dominated catchments, RF-MEP performed the best using both the KGE' and AOF; while for nivo-pluvial catchments, CR2MET provided the best performance with the KGE', and all of MSWEPv2.8, ERA5 and CR2MET performed similarly well with the AOF. CR2MET performed the best in pluvio-nival catchments for the case of feature similarity when using the KGE', and CR2MET and RF-MEP with the AOF. Finally, across rainfall-dominated catchments, MSWEPv2.8 performed best for spatial proximity and ERA5 for feature similarity using the KGE'; and CR2MET for both regionalisation methods when the AOF was used.

Results obtained with the AOF were similar to those obtained with the KGE'. Although the two objective functions used are not directly comparable, results for the snow-dominated and the pluvio-nival catchments using the AOF for parameter regression were markedly lower than those obtained using the KGE', suggesting that the choice of the objective function has a greater influence in the performance of parameter regression results.

4.3 Impact of nested catchments in regionalisation performance

Finally, we evaluated the influence of the nested catchments in the regionalisation results. Figure 9 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 catchments (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 performance of parameter regression did not change substantially when evaluated with and without nested catchments.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

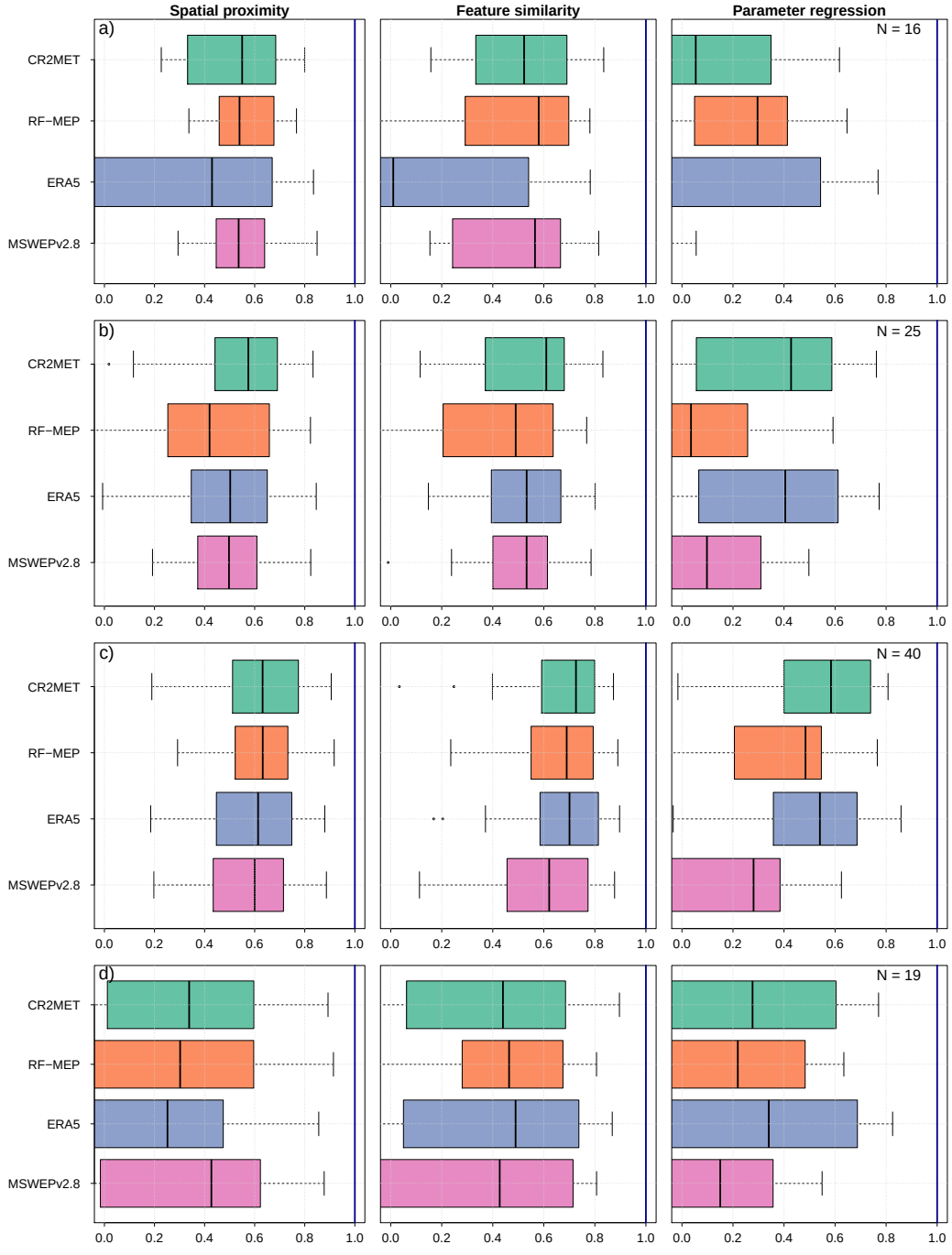


Figure 7. Performance of regionalisation methods using the KGE' objective function according to the hydrological regimes: a) snow-dominated, b) nivo-pluvial, c) pluvio-nival, and d) rain-dominated. N denotes the number of catchments per hydrological regime.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

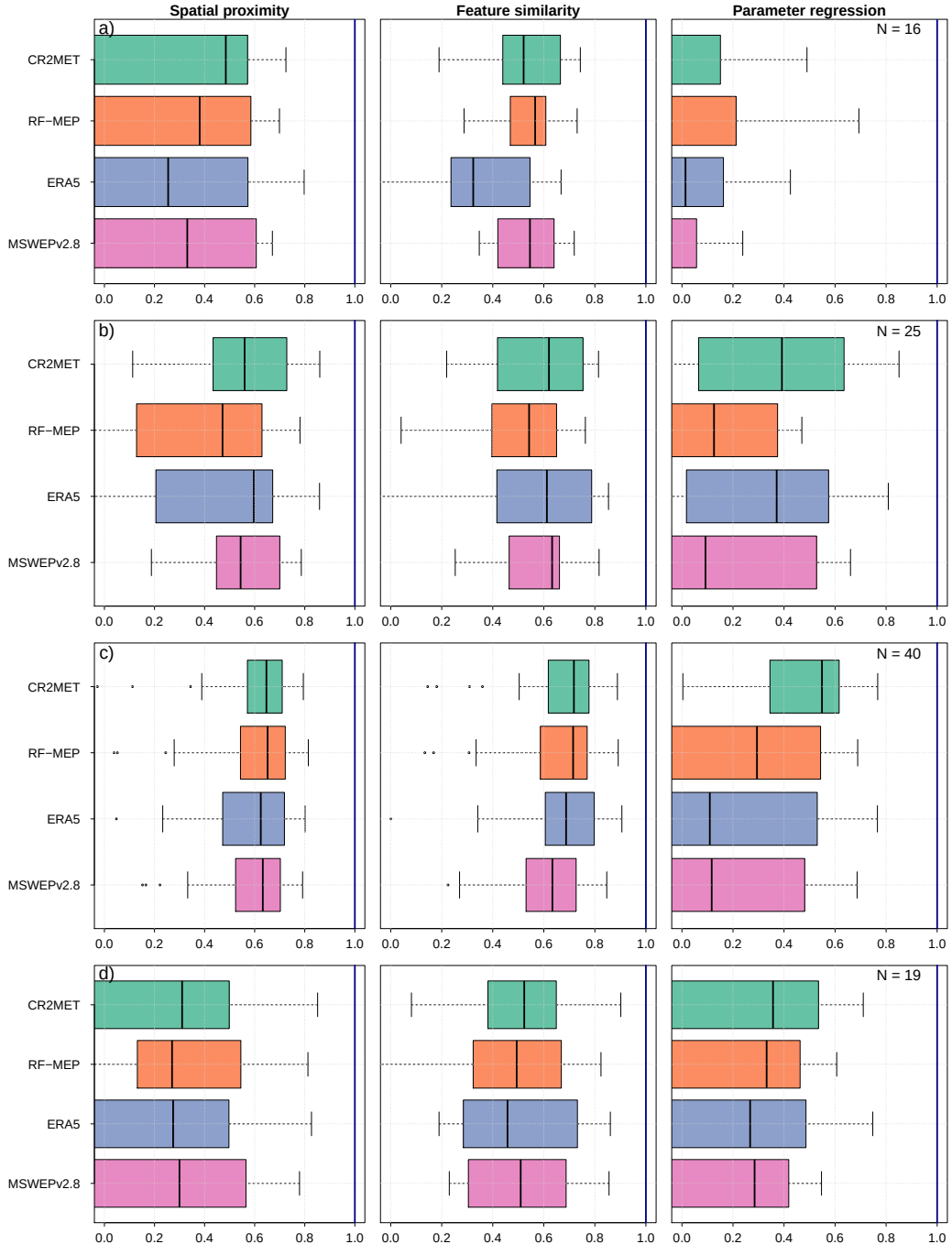
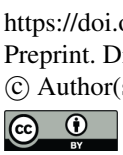


Figure 8. Performance of regionalisation methods using the AOF objective function according to the hydrological regimes: a) snow-dominated, b) nivo-pluvial, c) pluvio-nival, and d) rain-dominated. N denotes the number of catchments per hydrological regime.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021



Hydrology and Earth System Sciences
Discussions
Open Access
EGU

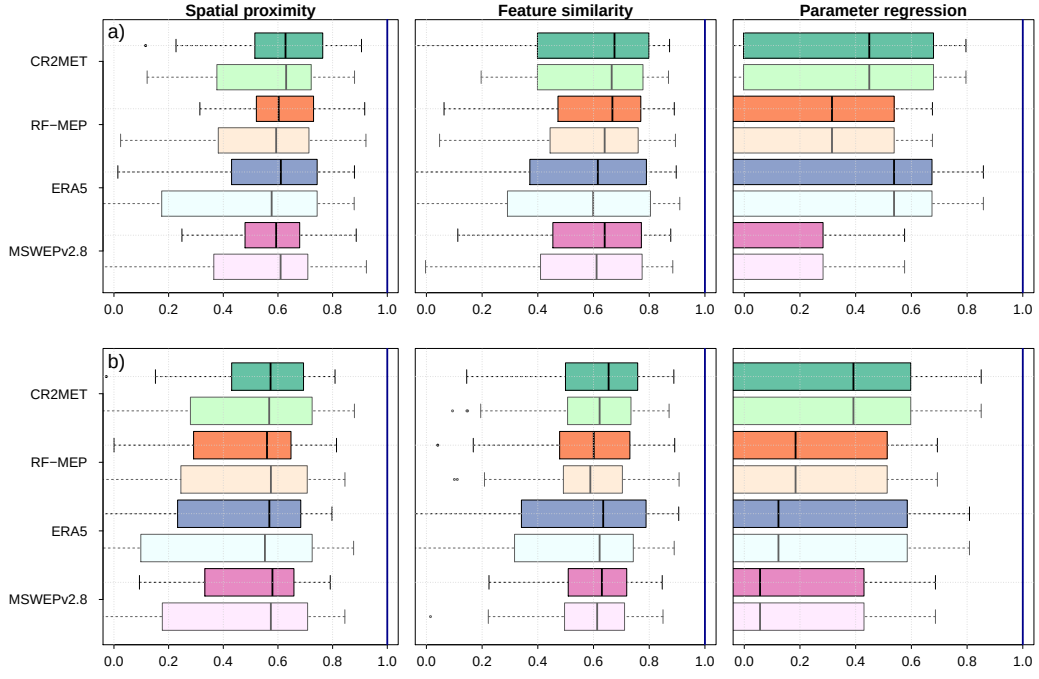


Figure 9. Comparison of regionalisation performance using all catchments as potential donors (dark colours) against the performance when catchments that share a common area with the pseudo-ungauged catchment are excluded as potential donors (light colours). Results are for a) the KGE', and b) the AOF.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021


Hydrology and Earth System Sciences
Discussions

5 Discussion

5.1 Performance of P products

During the individual catchment calibration (2000–2014) and two verification periods (1990–1999 and 2015–2018), good performances were obtained with all P products (see Figure 3). The results obtained with ERA5, which is a reanalysis-only product, were as good or even better than those obtained with the gauge-corrected products CR2MET, RF-MEP, and MSWEPv2.8 (e.g., see AOF results for Verification 1 in Figure 3). This 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 lack of P rain gauges in the Andes mountain range. 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 and for both objective functions. These results demonstrate that the calibration of hydrological model parameters smooths out, to some extent, the spatial differences among P products (see Figure 2), which is in agreement with previous studies that have demonstrated that model recalibration 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 systematic higher performance of CR2MET and RF-MEP for calibration and Verification 1 compared to MSWEPv2.8 (see Figure 3) can be attributed to the improved detection of P events of the merged products (regarding RF-MEP, see Baez-Villanueva et al., 2020). However, MSWEPv2.8 performed similar to these merged products under dry conditions (Verification 2).

Regarding the suitability of P products for parameter regionalisation, RF-MEP provided slightly better results in northern Chile (17-30°S) 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 low performance obtained with ERA5 in the north compared to the other P products can be attributed to the absence of ground-based meteorological stations in its development. The use of ground observations as they: $i)$ help to compensate 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). Despite the low performance of all P products in the north (median KGE' and AOF values $<$0.5, see Figure 5), 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 individual calibration of the hydrological model and the two verification periods. The performance of ERA5

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021


Hydrology and Earth System Sciences
Discussions

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 S3 of the supplement material). 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. These results 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 Evaluation of regionalisation techniques

Feature similarity provided the best performance when the TUWmodel was forced with all P products (Figure 6), reinforcing the idea that this conceptual hydrological model has a flexible enough structure to correct inter-forcing differences without losing predictive capability in space, as observed in the obtained P_{reg} values (only applied to the best performing regionalisation technique, i.e., feature similarity). Spatial proximity provided similar performance to feature similarity over southern-central Chile, 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 catchment spatial proximity. Parameter regression was the regionalisation method that provided the worst results (Figures 4 and 6); however, Figure 5 shows that this method generated good results over southern-central Chile (except for high-elevation areas), where there are many potential donor catchments. The equifinality of model parameters may also impact the relative performance of the regionalisation techniques by producing unrealisitic parameter sets, particularly for the case of parameter regression. 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 out of the scope of this study.

Figures 7 and 8 show the performances of the three regionalisation techniques according to hydrological regimes (see Figure 1). For both spatial proximity and feature similarity, the best and worst results were obtained for pluvio-nival catchments and rain-dominated catchments, respectively, regardless of the objective function. The relative performance of snow-dominated catchments and nivo-pluvial catchments was related to the selection of the objective function. A better performance was achieved over snow-dominated catchments when the KGE' was used as the objective function, while the opposite was observed for the AOF.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021


Hydrology and Earth System Sciences
Discussions

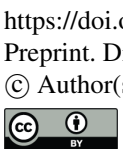
We calculated P_{reg} and R_{reg} to evaluate the ensemble spread in simulated Q resulting from the implementation of each of the 10 donor catchment parameter sets for feature similarity using both objective functions, as described in Section 3.4.2. Figure 10 shows the catchments according to whether P_{reg} was higher (indicating better performance) using the KGE' (green) or the AOF (orange), while catchments with P_{reg} differences of less than 0.05 were excluded (plotted as empty circles). The use of the AOF resulted in higher P_{reg} values over northern Chile, implying that a larger fraction of daily Q observations lie within the uncertainty bands, with median values for the P products ranging from 0.69 (MSWEPv2.8) to 0.74 (RF-MEP). On the other hand, the KGE' resulted in higher P_{reg} values over the humid south-central Chile, with median values ranging from 0.69 (MSWEPv2.8) to 0.76 (CR2MET and ERA5). These results are in agreement with (Beck et al., 2016), who reported that the inclusion of hydrological signatures in the objective function reduced regionalisation performance over more humid (tropical) catchments. The comparison of the two objective functions with the R_{reg} index (Figure S2 in the supplement material) shows that the uncertainty band was narrower in the northern region when using KGE', with median values ranging from 2.20 (ERA5) to 1.16 (RF-MEP). Conversely, the use of the AOF over south-central and southern Chile provided better results, with median values ranging from 0.89 (ERA5) to 0.81 (RF-MEP), suggesting that there is a trade off between P_{reg} and R_{reg}. Interestingly, all P products, regardless of their differences, provided very similar spatial patterns of P_{reg} and R_{reg}, with the exception of R_{reg} for RF-MEP, where the use of the AOF consistently resulted in better P_{reg} and R_{reg} values.

Figure 11 displays Q time series for two catchments where good performance was obtained with feature similarity as the regionalisation technique, over a snow-dominated (in the north, Figure 11a) and a rain-dominated (in the south, Figure 11b) catchment, using CR2MET as the P forcing. In the snow-dominated catchment, the AOF resulted in a higher P_{reg}, while in the rain-dominated catchment, the KGE' led to a higher P_{reg} value. In the snow-dominated catchment, the AOF was better at reproducing the recession curves of small events at the expense of returning a wider uncertainty band (i.e., higher R_{reg}). In the rain-dominated catchment, both objective functions produced good representations of the recession curves, and again the KGE' showed a slightly wider uncertainty band (i.e., higher R_{reg}) than AOF.

5.3 Impact of nested catchments

Nested catchments play an important role in the performance of regionalisation methods because nested catchments are more likely to have a strong climatological and physiological similarity to each other. As observed in 9, 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 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 are nested in a larger one). As expected, parameter regression was affected the least by the exclusion of nested catchments, as their removal had 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 compared to the results reported in a case study over Austria (Neri et al., 2020, ; their Figure 9a), which could be attributed to: i) the degree of nestedness of the selected catchments,

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021



Hydrology and Earth System Sciences
Discussions

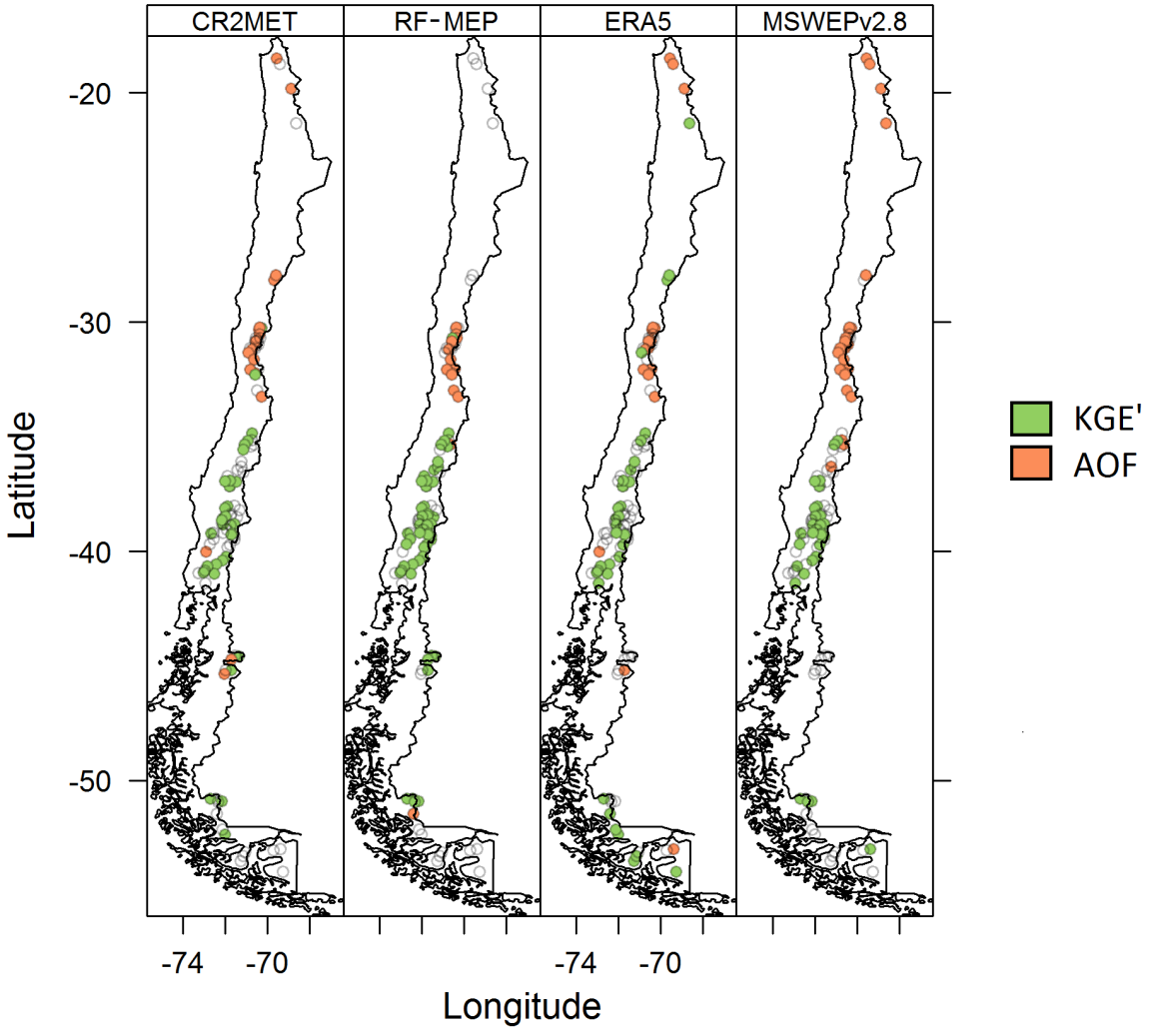


Figure 10. Comparison of whether P_{reg} is greater for KGE' and AOF according to catchment. Catchments with similar results (difference < 0.05) are plotted as empty circles.

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 3 nested catchments); and ii) the percentage of total catchments that are nested (42% in this study, compared to 65% in the Austrian case study).

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

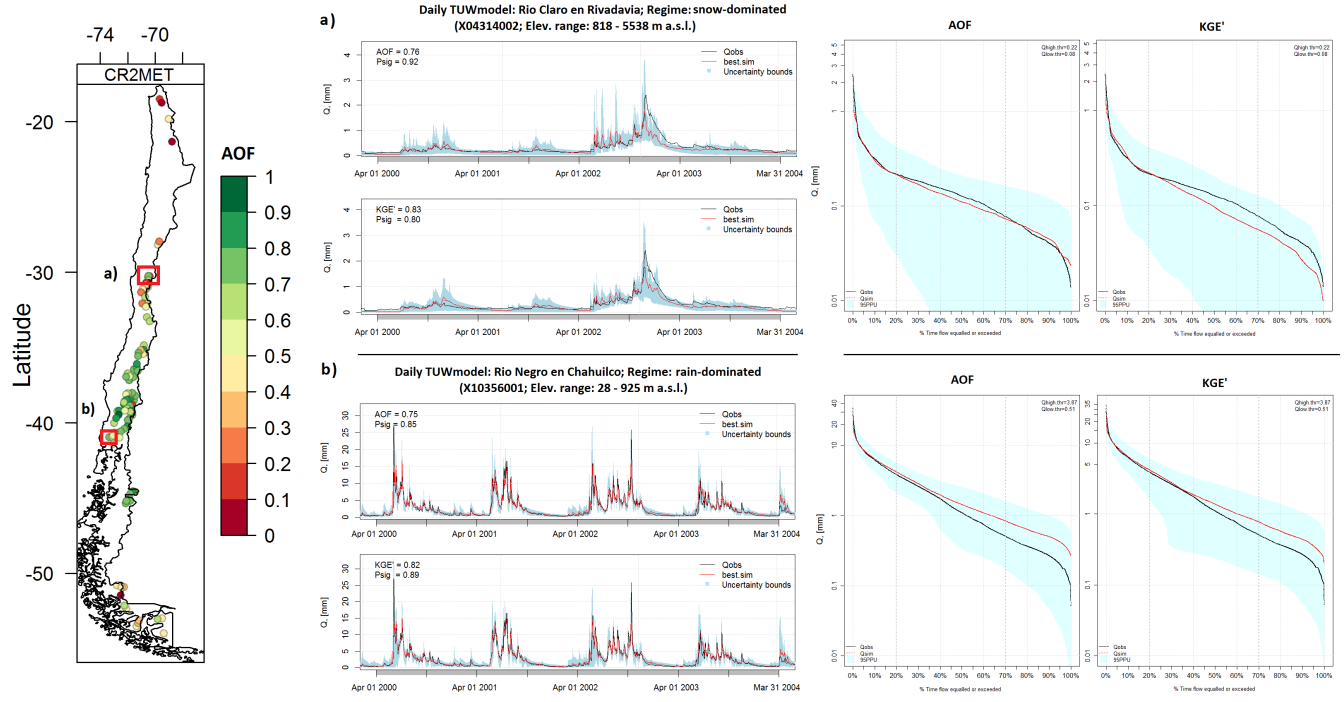


Figure 11. Visualisation of three hydrological years (2000–2005) of two well performing catchments: *a*) Rio Claro en Rivadavia (snow-dominated); and *b*) Rio Negro en Chahuilco (rain-dominated). The left panel shows the location of these catchments, the middle panel plots its hydrographs, and the right panel shows the flow duration curves for the KGE' and AOF calculated over 2000–2014, emphasising the low flows.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021


Hydrology and Earth System Sciences
Discussions

6 Conclusion

Streamflow prediction in ungauged catchments is an essential task for water resources management, and uncertainties arising from P products provide a challenge for the accurate prediction of daily Q. 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 model at the daily temporal scale, and two objective functions (KGE' and AOF) were used for calibration 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 with the best (worst) performance during independent calibration and verification did not necessarily provide the best (worst) results during regionalisation.

2. From the selected P products, the use of those that are corrected with daily gauge observations did not necessarily translate into improved hydrological model performance. However, we expect that P products with lower performances compared to 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 both the calibration and verification periods.

4. Feature similarity was the best performing regionalisation technique, regardless of the choice of gridded P product or the calibration criteria we explored (KGE' or AOF).

5. Spatial proximity was the second best performing regionalisation method. This can be explained by the fact that, in our study area, spatial proximity is a good proxy of climatic similarity for most neighbouring catchments.

6. Parameter regression provided the worst regionalisation performance, reinforcing the importance of transferring complete parameter sets to ungauged catchments.

7. 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.

8. When using feature similarity, smaller uncertainty bands were obtained using the ensemble of parameter sets obtained with AOF calibrations across the arid north, while the KGE' provided a more reliable ensemble in more humid central-southern Chile.

The results presented in this study are valid only for Chile and might not necessarily be valid for other regions. Despite this, they provide guidance for ongoing and future studies involving the application of gridded P products for regionalisation of

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021


Hydrology and Earth System Sciences
Discussions

hydrological model parameters in ungauged basins. The feature similarity procedure described here could be used to refine the parameter transfer approach that has been preliminary adopted in national scale hydrological characterisations for Chile (e.g., Bambach et al., 2018; Lagos et al., 2019). Additionally, further analyses could cover $i)$ 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 $v)$ assessments of the sensitivity of regionalisation results with respect to modified climate scenarios.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

Appendix A: 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 A1 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 high correlations between this variable and 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.

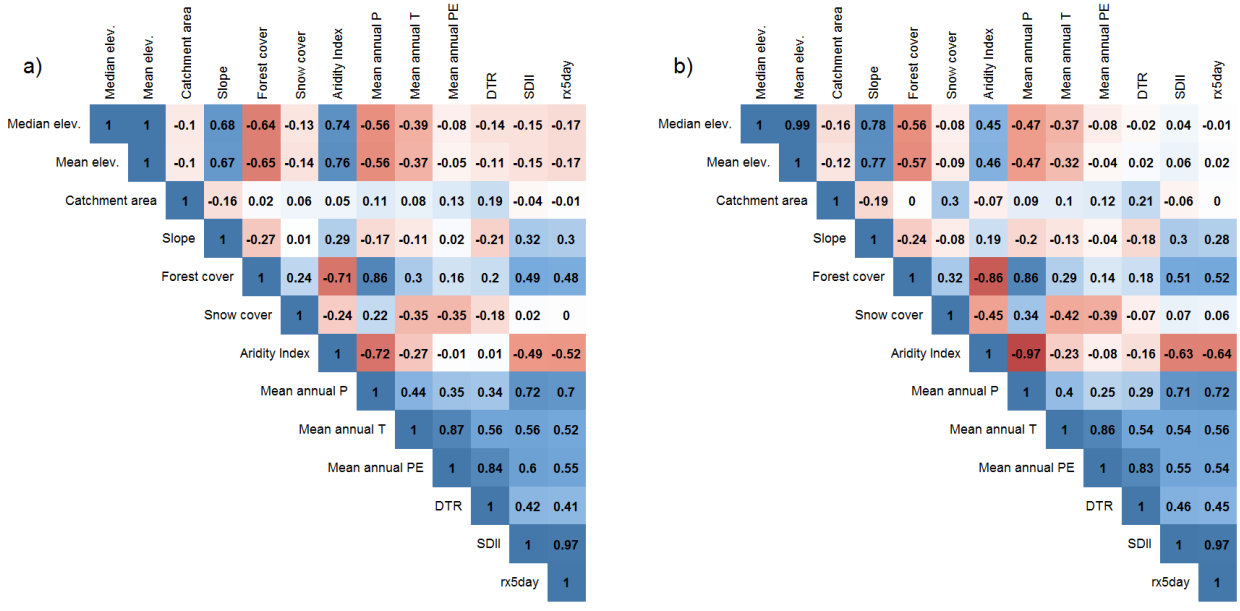


Figure A1. Correlation matrices of the catchment characteristics described in Table 4 using CR2METv2 as the P product for a) the Pearson correlation, to evaluate linear correlation, and b) the Spearman correlation to evaluate the monotonic correlation.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

Competing interests. no competing interests

Acknowledgements. The authors thank 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. 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 for developing and maintaining the `raster` and `TUWmodel` R packages, respectively.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

References

Abbaspour, K. C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J., and Srinivasan, R.: Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT, Journal of hydrology, 333, 413–430, 2007.

Abbaspour, K. C., Faramarzi, M., Ghasemi, S. S., and Yang, H.: Assessing the impact of climate change on water resources in Iran, Water resources research, 45, 2009.

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, https://www.hydrol-earth-syst-sci.net/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.

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, 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, 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.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

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.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021


Hydrology and Earth System Sciences
Discussions

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, 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.

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.

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.

Driessen, T., Hurkmans, R., Terink, W., Hazenberg, P., Torfs, P., and Uijlenhoet, R.: The hydrological response of the Ourthe catchment to climate change as modelled by the HBV model., Hydrology & Earth System Sciences, 14, 2010.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

Duan, Q., Sorooshian, S., and Gupta, V.: Effective And Efficient Global Optimization For Conceptual Rainfall-Runoff Models, Water Resources Research, 28, 1015–1031, https://doi.org/10.1029/91WR02985, 1992.

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, M., 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, https://doi.org/10.1016/j.jhydrol.2021.126047, *(In press)*, 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, 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.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

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.

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.

Jin, X., Xu, C.-y., Zhang, Q., and Chen, Y. D.: Regionalization study of a conceptual hydrological model in Dongjiang basin, south China, Quaternary International, 208, 129–137, 2009.

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, https://www.hydrol-earth-syst-sci.net/22/1775/2018/, 2018.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

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, https://hess.copernicus.org/articles/24/5149/2020/, 2020.

Nijzink, R., Almeida, S., Pechlivanidis, I., Capell, R., Gustafssons, D., Arheimer, B., Parajka, J., Freer, J., Han, D., Wagener, T., et al.: Constraining conceptual hydrological models with multiple information sources, Water Resources Research, 54, 8332–8362, 2018.

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.

Pushpalatha, R., Perrin, C., Le Moine, N., and Andréassian, V.: A review of efficiency criteria suitable for evaluating low-flow simulations, Journal of Hydrology, 420, 171–182, 2012.

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.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

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.

Samuel, J., Coulibaly, P., and Metcalfe, R. A.: Estimation of continuous streamflow in Ontario ungauged basins: comparison of regionalization methods, Journal of Hydrologic Engineering, 16, 447–459, 2011.

Santos, L., Thirel, G., and Perrin, C.: Pitfalls in using log-transformed flows within the KGE criterion, 2018.

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.

Schuol, J., Abbaspour, K. C., Srinivasan, R., and Yang, H.: Estimation of freshwater availability in the West African sub-continent using the SWAT hydrologic model, Journal of hydrology, 352, 30–49, 2008.

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., and Parajka, J.: The impact of the variability of precipitation and temperatures on the efficiency of a conceptual rainfall-runoff model, Slovak Journal of Civil Engineering, 24, 1–7, 2016.

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.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

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.

Villarini, G. and Krajewski, W. F.: Empirically-based modeling of spatial sampling uncertainties associated with rainfall measurements by rain gauges, Advances in Water Resources, 31, 1015–1023, 2008.

Vis, M., Knight, R., Pool, S., Wolfe, W., and Seibert, J.: Model calibration criteria for estimating ecological flow characteristics, Water, 7, 2358–2381, 2015.

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.

Wallner, M., Haberlandt, U., and Dietrich, J.: A one-step similarity approach for the regionalization of hydrological model parameters based on Self-Organizing Maps, Journal of hydrology, 494, 59–71, 2013.

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.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

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, http://www.sciencedirect.com/science/article/B6V6C-3SX81SS-6/2/4893345d9204707c4243f32a949063b0, 1998.

Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, Water Resources Research, 44, W09 417, https://doi.org/10.1029/2007WR006716, 2008.

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 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/https://doi.org/10.1016/j.jhydrol.2021.125969, http://www.sciencedirect.com/science/article/pii/S0022169421000160", 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.

https://doi.org/10.5194/hess-2021-156
Preprint. Discussion started: 23 April 2021

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

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.