The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies – Chile dataset

Camila Alvarez-Garreton[1,2], Pablo A. Mendoza[3], Juan Pablo Boisier[1,4], Nans Addor[5], Mauricio Galleguillos[1,6], Mauricio Zambrano-Bigiarini[1,7], Antonio Lara[1,2], Cristóbal Puelma[1,6], Gonzalo Cortes[8], Rene Garreaud[1,4], James McPhee[3,9], Alvaro Ayala[10,11]

[1]Center for Climate and Resilience Research (CR2), Santiago, Chile
[2]Instituto de Conservación, Biodiversidad y Territorio, Universidad Austral de Chile, Valdivia, Chile
[3]Advanced Mining Technology Center, Universidad de Chile, Santiago, Chile
[4]Department of Geophysics, Universidad de Chile, Santiago, Chile
[5]Climatic Research Unit, School of Environmental Sciences, University of East Anglia, UK.
[6]Faculty of Agronomic Sciences, Universidad de Chile, Santiago, Chile
[7]Department of Civil Engineering, Faculty of Engineering and Sciences, Universidad de La Frontera, Temuco, Chile
[8]Department of Civil and Environmental Engineering, University of California, Los Angeles, California, USA
[9]Department of Civil Engineering, Universidad de Chile, Santiago, Chile
[10]Laboratory of Hydraulics, Hydrology and Glaciology (VAW), ETH Zurich, Zurich, Switzerland
[11]Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland

Correspondence to: Camila Alvarez-Garreton (camila.alvarez@uach.cl)

Abstract. We introduce the first catchment dataset for large sample studies in Chile (South America). The dataset provides boundaries for 516 catchments along with daily streamflow records and basin-averaged daily time series of precipitation (from a national and three global datasets), maximum, minimum and mean temperature, potential evapotranspiration (PET; from two datasets), and snow water equivalent. We calculated hydro-climatological attributes using these time series, and processed diverse data sources to extract topographic, geological and land cover features. Relying on publicly available reservoirs and water rights data for the country, we computed the degree of anthropic intervention within the catchments. To facilitate the use of this dataset and promote common standards in large-sample studies, we computed most catchment attributes introduced by Addor *et al.*, (2017) in their Catchment Attributes and MEteorology for Large-sample Studies (CAMELS) dataset, and proposed several others.

The dataset presented here (named CAMELS-CL) is used to evaluate biases and uncertainties in basin-wide precipitation and PET. Looking at the watershed water balances, we found large discrepancies between precipitation products in arid regions, and a systematic precipitation underestimation in headwater mountain catchments (high elevations and steep slopes) over humid regions. The contrast of PET products with ground data indicated a fairly good performance of both products in humid regions ($r > 0.91$) and lower correlation ($r < 0.76$) in hyper-arid regions. Within the assessed products, a satellite-based PET showed a consistent overestimation of observation-based PET. Finally, we explore the effects of local perturbations on catchment response by analysing the relationship between hydrological signatures and a proposed human intervention attribute. We showed that larger anthropic intervention is correlated with lower than normal annual flows, runoff ratios, elasticity of runoff with respect to precipitation, and flashiness of runoff, especially in arid catchments.

CAMELS-CL provides unprecedented information in a continent largely underrepresented in large-sample studies. We envisage its use in a myriad of applications, including catchment classification and regionalisation studies, water supply modelling under different management scenarios, the characterisation of drought history and projections, and the exploration of climate change impacts on hydrological processes. This effort is part of an international initiative to create a multi-national large sample datasets freely available for the community. CAMELS-CL is freely available from http://www.cr2.cl/download/camels-cl/.

1 Introduction

Large-sample hydrology has been recognised as a fundamental tool to advance hydrological science (e.g., Andréassian *et al.*, 2006; Ehret *et al.*, 2014). The insights provided by studying a large set of catchments complement the findings from intensive place-based studies, where more detailed analyses are conducted over a small number of catchments. A common approach in large-sample studies is to explore interrelationships between catchment attributes describing landscape, climate and hydrologic behaviour, typically obtained from topography, soil types, geology, land cover and hydro-meteorological datasets (e.g., Oudin *et al.*, 2008; Sawicz *et al.*, 2011; Gupta *et al.*, 2014; Newman *et al.*, 2015; Addor *et al.*, 2017). Accounting for catchments attributes in a comprehensive dataset serves various purposes. For example, comparative hydrology and catchment classification studies use these attributes to explore catchment (dis)similarities (e.g., McDonnell and Woods, 2004; Wagener *et al.*, 2007; Sawicz *et al.*, 2011; Berghuijs *et al.*, 2014). Likewise, regionalisation studies incorporate catchment attributes to identify (hydrologically and physically) similar catchments that can be used to transfer model information from gauged to ungauged locations (Blöschl et al., 2013; Sawicz et al., 2011) – a fundamental motivation of the Predictions in Ungauged Basins (PUB) initiative (Sivapalan et al., 2003). In summary, the main goal of large-sample applications is to learn from diversity in order to define generalizable principles that can help to improve the predictability of the water cycle. This is addressed by disentangling the interplay between landscape, climate and hydrologic behaviour, which provides insights on hydrological systems and on suitable model structures to represent them.

As highlighted by Gupta *et al.*, (2014), a key challenge in large-sample hydrology is data accessibility, which is particularly critical in data-scarce regions such as South America (see Fig. 2 in Gupta *et al.*, 2014). Although there is a tendency for large-sample datasets to be shared worldwide (see examples in Gupta *et al.*, 2014), freely available hydro-meteorological records from individual countries typically use different formats and come from different providers. Moreover, they are rarely spatially aggregated to the catchment scale, which makes it difficult for researchers and practitioners to use them for basin-oriented applications.

In this paper, we introduce a unique dataset including 516 catchments in Chile and show how this dataset serves to improve our understanding of hydrological systems and their predictability by the analysis of (1) the uncertainties in two key meteorological variables (precipitation and PET), and (2) the impacts of anthropic intervention on catchment response. Chile extends over 4,300 km along the west of South America (17.8°S to 55.0°S), encompassing several climatic conditions, including hyper-arid, Mediterranean and hyper-humid regimes. Chile also exhibits a complex topography, dominated by the Andes Cordillera, the longest mountain chain in the world, with elevations up to 7000 m a.s.l. (178 catchments have a mean elevation greater than 2000 m a.s.l.).

The dataset built here consists on catchment boundaries in shapefile format, hydro-meteorological time series, and a suite of catchment attributes calculated from climate, hydrology, topography, geology, land cover, and water use characteristics. To facilitate and encourage the use of the dataset presented here, and in order to promote common standards and formats in large-sample studies, we compute five (out of six) classes of catchment attributes (location and topography, geology, land cover characteristics, climatic indices and hydrological signatures) used in Addor et al., (2017, referred as A17 hereafter). A17 introduced the Catchment Attributes and MEteorology for Large-sample Studies dataset (CAMELS dataset), which encompasses meteorological and streamflow datasets collated by Newman et al., (2015) and provides quantitative estimates of a wide range of attributes for 671 catchments in the contiguous United States (CONUS). The CAMELS dataset has already been used in a myriad of applications, including assessment of streamflow skill elasticity to initial conditions and climate prediction (Wood et al., 2016), snow data assimilation for seasonal streamflow prediction (Huang et al., 2017), continental-scale hydrologic parameter estimation (Mizukami et al., 2017), and climate change impacts on the hydrology of the CONUS, among others. Following this nomenclature, we name our dataset CAMELS-CL, which stands for CAMELS dataset in Chile. We add an attribute class not covered by A17: the degree of human intervention in each catchment. This novel information is valuable since anthropogenic activities may have major impacts on catchment behaviour, but human influence is often difficult to quantify, especially for hundreds of catchments.

CAMELS-CL is applied to analyse uncertainties in precipitation and potential evapotranspiration estimates, and to quantify anthropic impacts on catchment response. First, we compare the different precipitation products and evaluated them based on the observed water balance. This analysis includes one national dataset (CR2MET) and three widely used global datasets (CHIRPS, MSWEP and TMPA); thus, the results may have implications beyond the domain covered by CAMELS-CL. Secondly, we assess the PET products based on an independent set of PET point values calculated from meteorological records. Finally, we analyse human influence on catchment behaviour by relating hydrological signatures with a human intervention attribute calculated from water extraction information.

The paper is structured as follows: Sect. 2 describes the study area; Sect. 3 describes the collected datasets (Sect. 3.1) and provides a description of the derived catchment attributes with a discussion of their spatial distribution (Sect. 3.2); Sect. 4 presents the precipitation (Sect. 4.1) and potential evapotranspiration (Sect. 4.2) uncertainty analyses; Sect. 5 presents the analysis of human influence on catchment behaviour; and Sect. 6 presents the main conclusions of the paper.

2 Study area

The area covered by CAMELS-CL corresponds to continental Chile, a territory with a distinct geographical configuration that spans 4,300 km along a north-south axis. The country lies on the Nazca and Antarctic tectonic plates. The tectonic activity in the Quaternary (early Pleistocene) led to the formation of the three main physiographic characteristics of the territory (from west to east): the coastal range, the intermediate depression, and the Andes Cordillera. The Andes range defines the east border with Bolivia and Argentina for most of the country. Featuring altitudes well above 3,000 m a.s.l., with summits of up to 7,000 m a.s.l. (e.g. Aconcagua mountain or Ojos del Salado volcano), the Andes acts as an effective barrier for atmospheric flows, leading to particularly large precipitation amounts at high elevations (Garreaud, 2009).

Chile has 16 administrative regions (Fig. 1) split into four macro-zones defined by the Chilean Water Directorate (DGA), based on hydrological, climatic and topographic features (DGA, 2016a): North (from Arica and Parinacota to Coquimbo regions); Central (from Valparaiso to Maule regions); South (from Bio-Bio to Los Lagos regions); and Austral (from Aysén to Magallanes regions). To provide a more detailed discussion, we divided the North macro-zone into Far North (from Arica and Parinacota to Antofagasta regions) and Near North (from Atacama to Coquimbo regions), and the Austral macro-zone into Austral zone (Aysen region) and Southern Patagonia (Magallanes region). The resulting six macro-zones are presented in Fig. 1.

The country includes five primary climatic regimes according to Köppen's climate classification (Kottek et al., 2006). The Far North is dominated by a cold desert climate (BWk) and tundra (ET) along the Andes range. The Near North is characterised by cold desert climate in the Atacama region and a cold semi-arid climate (BSk) in the Coquimbo region. The Central zone is dominated by a sub-humid Mediterranean climate (Csb). The Southern zone includes a humid Mediterranean climate in Bio-Bio and Araucanía regions, and a temperate rain-oceanic climate (Cfb) in Los Rios and Los Lagos regions. The Austral and Southern Patagonia zones are dominated by rain-cool oceanic (Cfc) and cold steppe (BSk) climates.

3 CAMELS-CL dataset

3.1 Input data

3.1.1 Topography and catchment boundaries

The first step in the development of CAMELS-CL was the delimitation of catchment boundaries (Fig. 2). An official database for Chilean hydrographic network was developed by the Instituto Geográfico Militar in 1984 (IGM, 1984) and updated by the DGA in 2014 (DGA; CIREN, 2014). This network was made following Strahler hierarchy (Strahler, 1957), using the 30-m ASTER GDEM (Tachikawa et al., 2011) elevation data. The DGA network includes 102 catchments, 491 sub-catchments and 1481 sub-sub-catchments, and has been largely used by government agencies, the private sector and the general public. However, a key limitation of this hydrographic network is that – given the methodology used for its implementation – the existing streamflow gauges do not correspond with catchment outlets. Furthermore, DGA catchment boundaries are truncated at the administrative national border, even when there are some catchments contributing with runoff from Bolivian and Argentinian territories. Since any hydrologic application within a controlled basin requires the total contributing area associated to the measured streamflow, and there is no official catchment boundary database, different studies have applied their own basin delineations, making it difficult to compare results.

To overcome this limitation, we created our own catchment boundaries database for CAMELS-CL, defining the basin outlets at the location of 516 selected streamflow gauges (Sect. 3.1.5), and following only topographic-driven limits (not the administrative national border). A key challenge for this task is the mismatch between some station geographic coordinates reported by the DGA and the river network location, according to Google Earth imagery (Google, 2016). For some of those cases, expert advice was obtained from DGA technicians regarding gauge locations, while, for others, ancillary information (e.g., gauge name, road maps, Google Earth imagery) was used to determine the most probable location.

Basin delineation was performed in Quantum GIS (QGIS Development Team, 2015) by using watershed delineation packages from the Geographic Resources Analysis Support System (GRASS) (Neteler et al., 2012) and 30-m ASTER GDEM (Tachikawa et al., 2011) as input elevation data. Given the topographic characteristics of Chile, several catchments collected in this dataset are nested. To report this, we generated a logical catchment hierarchy matrix indicating those basins that are contained within another catchment of the dataset. The hierarchy matrix can be used to filter independent catchments, which is required for some applications such as hydrological modelling of large basins, catchment classification and parameter regionalisation.

The main topographic properties (area, median, mean, minimum and maximum elevation, mean slope) for each catchment were computed from ASTER GDEM 30-m raster data (Tachikawa et al., 2011), clipped by the catchment boundary polygons and processed with the R raster package (Hijmans, 2016). An important limitation of this dataset is that its spatial resolution is relatively coarse, which can lead to errors when delineating catchments over very flat regions (such as the Far North, see Sect. 3.2.1).

3.1.2 Geology

Catchment-scale geological characteristics were retrieved from the Global Lithological Map database (GLiM) produced by Hartmann and Moosdorf (2012). GLiM is a compilation of national datasets into a unified global map. In the case of Chile, Hartmann and Moosdorf (2012) relied on the map produced by the Servicio Nacional de Geología y Minería (Sernageomin, 2004), which has a resolution of 1:1,000,000 and is the most complete and commonly used map for the country. For each catchment, we reported the most and second-most frequent geological class, as well as the fraction of the catchment they cover. We also extracted the fraction of the catchment described as "carbonate sedimentary rocks", as it is a useful indicator of the presence of karstic systems.

3.1.3 Land cover

We used the 30-m resolution land cover map provided by Zhao *et al.* (2016), which integrates multi-seasonal Landsat 8 imagery acquired during 2013 and 2014. The classification scheme adopted by Zhao *et al.* (2016) was designed with Chilean geographers and biodiversity researches, mainly based on the FROM-GLC project (Gong et al., 2013), which is similar to the

Land Cover Classification System (Di Gregorio and Jansen, 2005). This classification scheme is compatible with other land cover classification systems such as FAO or IGBP, with minor ancillary data. It consists on 10 main (level-1) classes (Fig. 1): croplands; forests, grasslands; shrublands; wetlands; water bodies; impervious surfaces; barren lands; and snow and ice. Some of the classes were refined in level-2 (e.g., separate native forest and exotic forest plantation) and level-3 subclasses (a total of 30 and 35 subclasses, respectively). For CAMELS-CL, we used the R raster package (Hijmans, 2016) to clip the land cover map within each catchment boundary polygon, and compute the fractional area associated with each class or subclass (as described in Table 3).

3.1.4 Glaciers

Glaciers in Chile can be found at several locations, varying from small ice bodies at high-elevation sites of the Atacama region, to alpine glaciers in the Central Zone, and the large Patagonian ice fields in the Austral and Southern Patagonia regions (Pellicciotti et al., 2014). Even though the land cover map from Zhao et al. (2016) identify areas of snow and ice, we included a global glacier inventory for calculating the degree of glacierisation of the selected catchments. Glacier inventories have the advantages of using geomorphologic glacier-delineation techniques, and the recognition of debris-covered areas that cannot be identified by land cover classification schemes. In this study, we used the latest version of the Randolph Glacier Inventory (RGI 6.0; RGI Consortium, 2017). RGI 6.0 is a globally complete inventory of glacier outlines and it is widely used in regional and global studies on land surface fluxes, climatology and meteorology (e.g., Huss and Hock, 2015; Marzeion et al., 2012; Mernild et al., 2017). We preferred to use RGI 6.0 rather than the Chilean glacier inventory from DGA (DGA, 2014) because there are portions of some catchments lying on Argentinean territory (Fig. 2). The RGI 6.0 was clipped within each catchment and two attributes were computed: the total glacierised area (km^2) and the percentage of glacierised area in the catchment (%) (further details in Sect. 4).

3.1.5 Streamflow

We compiled daily streamflow records for gauges maintained by the DGA, available from the CR2 Climate Explorer (http://explorador.cr2.cl/). From the 809 gauges included there, we selected those currently operational (independently of their data period), or suspended after 31 December 1980 with a record period longer than 10 years. We also discarded gauges located in artificial channels, ending up with 516 selected gauges. Figure 3 illustrates the availability of daily streamflow records for different time periods (represented with different colours). Note that hydrological year is considered from April 1st to March 31st. As expected, the number of stations decreases with longer data availability. For example, if only stations with at most 5% of missing data were selected, this would lead to a subset of 90 to 115 stations (which corresponds to 18% and 22% of the total number of catchments within the database, respectively), depending on the time period considered. When considering all stations with at most 30% of missing data, then 249 to 258 stations (48% and 50% of catchments, respectively) would meet this criterion, depending on the period (Fig. 3). Figure 4a presents the mean annual discharge for each station (computed for the entire record period).

3.1.6 Precipitation

In most cases, precipitation is the main driver of hydrological systems. However, its estimation is highly uncertain, even in densely monitored regions (Tian and Peters-Lidard, 2010; Woldemeskel et al., 2013). This limitation is aggravated in regions with difficult accessibility, where only a sparse network of meteorological stations is available. In order to account for robust precipitation estimates and to characterise the uncertainty of this variable, we processed catchment-scale precipitation from four different products, whose main characteristics are summarised in Table 1. Each one of these products was clipped and averaged within the catchment boundaries, resulting in four daily time series for each catchment, named $precip_{cr2met}$, $precip_{chirps}$, $precip_{mswep}$ and $precip_{tmpa}$.

The $precip_{cr2met}$ times series was derived from CR2MET, a spatially-distributed daily precipitation product developed for Chile, which is currently being used by DGA to update the national water balance (DGA, 2017). The CR2MET product is partly based on a statistically downscaled ERA-Interim reanalysis data (Balsamo et al., 2015). The method builds on multiple linear regression models to transfer precipitation, moisture fluxes and other variables from ERA-Interim onto regional (0.05º) precipitation. The statistical models, which also consider a number of topographic parameters, were calibrated with a large network of quality-controlled rain-gauge records. Depending on the distance of a given grid-cell to neighbouring stations, the final product was obtained from merging downscaled precipitation and spatially interpolated in-situ observations. Further information about formulation, quality control and product assessments can be found in DGA, (2017).

The three satellite-based precipitation products used in CAMELS-CL were selected based on the exhaustive comparison and evaluation reported by Zambrano-Bigiarini et al. (2017) for the entire Chilean territory. The variable $precip_{chirps}$ was computed from the Climate Hazards Group InfraRed Precipitation with Station data version 2 (CHIRPS, Funk et al., 2015), a long term (1981 to near-present), quasi-global (50°N to 50°S) daily dataset available at a spatial resolution of 0.05°, designed to monitor agricultural drought and global environmental changes over land. CHIRPS uses the Tropical Rainfall Measuring Mission Multi-Satellite Precipitation Analysis version 7 (TRMM 3B42v7) in order to calibrate global Cold Cloud Duration rainfall estimates (Funk et al., 2015). CHIRPS also incorporates surface rain-gauge data in order to reduce biases of its estimates, based on public and private monthly data. Originally, this dataset spanned from 50ºN to 50ºS, but since November 2012 data is not being produced south of 46ºS. More information can be found in Funk et al. (2015).

The $precip_{mswep}$ was computed from the Multi-Source Weighted-Ensemble Precipitation (MSWEP, Beck et al., 2017) data version 1.1, a fully global precipitation dataset released in June 2016, with a 3-hourly temporal and 0.25º spatial resolutions, specifically produced for hydrological modelling applications. MSWEP was designed to improve the performance of satellite products in representing precipitation over mountainous, tropical, and snowmelt-driven regions. The algorithm used in MSWEP merges observed rain-gauge data, satellite observations and reanalysis data to provide reliable precipitation estimates over the entire globe. In this paper, we used daily data from MSWEP version 1.1, but newer versions (already available) will be included in CAMELS-CL after validation with ground measurements in Chile.

Finally, $precip_{tmpa}$ was computed from the Tropical Rainfall Measuring Mission (Huffman et al., 2007) Multi-satellite Precipitation Analysis (TMPA) dataset, which provides quasi-global (50°N to 50°S) precipitation estimates at a spatial resolution of 0.25°. TMPA integrates infra-red and passive microwave data from a wide variety of satellite-borne precipitation-related sensors. In this study, we used the TRMM research product 3B42v7, which makes use of Global Precipitation Climatology Project (GPCP; Adler et al., 2003) and Climate Assessment and Monitoring System (CAMS, Ropelewski et al., 1984) data to rescale its estimations on a monthly basis.

3.1.7 Temperature

Daily time series of minimum, maximum and mean temperatures for each catchment were also derived from the CR2MET dataset (DGA, 2017). Daily minimum and maximum temperatures in CR2MET (CR2MET/T_{max} and CR2MET/T_{min}, respectively) were mapped for the period 1979-2016 using a different approach than the one used for precipitation (Sect. 3.1.6). In this case, the method uses land-surface temperature (LST) estimates from Moderate Resolution Imaging Spectro-radiometer (MODIS) satellite retrievals, in addition to near surface temperature provided by ERA-Interim. Multivariate regression models for both CR2MET/T_{max} and CR2MET/T_{min} were developed using LST as part of the explanatory variables and local temperatures records in Chile as target data. Given the data gaps and relatively short period available for LST, the final product was derived for the whole period (1979-2016) by fitting the ERA-Interim data to the preliminary (incomplete) MODIS-based product. To get daily mean temperatures (CR2MET/T_{mean}), the long-term CR2MET/T_{max} and CR2MET/T_{min} were used to adjust the ERA-Interim 3-hourly near surface temperature. The adjusted 3-hourly data was then averaged to derive CR2MET/T_{mean}. Gridded daily mean, minimum and maximum temperatures from CR2MET (0.05° lat-lon resolution) were clipped to obtain basin-averaged daily time series, named T_{mean}, T_{min} and T_{max}, respectively.

3.1.8 Potential evapotranspiration

We processed catchment-scale PET from two different sources. The first PET product uses the formulae proposed by Hargreaves and Samani (1985), which is solely based on surface temperature data (see Hargreaves and Allen, 2003 for further details). We used CR2MET/T_{max} and CR2MET/T_{min} (described in Sect. 3.1.7) to generate a gridded PET estimate (PET_{har}). The second PET data used in CAMELS-CL (PET_{mod}) is that provided by the MODIS PET product (MOD16 collection 5; Mu et al., 2005), which is processed from different sources of information, including leaf area index and fractional photosynthetically active radiation, FPAR/LAI (MOD15A2; Myneni et al., 2002), land cover type 2 (MOD12Q1; Friedl et al., 2002), albedo (MCD43B2 and MCD43B3; Jin et al., 2003; Lucht et al., 2000), and daily meteorological reanalysis data from NASA's MERRA GMAO (GEOS-5). MOD16 is calculated following the Penman-Monteith approach (Howell and Evett, 2001), and the final product is available at an 8-day temporal resolution for the period 2000-2014, on a 1×1 km^2 grid. Such as for other gridded variables, the PET_{har} and PET_{mod} products were clipped and averaged within basin boundaries to generate daily (called pet_{har}) and 8-day (called pet_{mod}) catchment-scale time series, respectively.

3.1.9 Snow water equivalent

We processed snow water equivalent (SWE) daily data using the 180-m resolution SWE product generated by Cortés and Margulis (2017). Cortés and Margulis (2017) obtained SWE ensemble estimates from forward modelling "prior" values, which were then conditioned trough the assimilation of historical fractional snow-covered area (fSCA) data from Landsat TM, ETM+ and OLI sensors. The "posterior" SWE and fSCA estimates were probabilistically conditioned on the observed depletion record from Landsat, the uncertainty in fSCA observations, and the forward model state uncertainty. The fSCA retrieval was obtained with a spectral un-mixing algorithm (Cortés et al., 2014). The forward models for prior ensemble generation were the SSiB3 land surface model (Yang et al., 1997) and a Snow Depletion Curve model (SDC; Liston, 2004). Detailed assessments of this reanalysis framework were performed for the Sierra Nevada using in-situ sensor data (Margulis et al., 2016), and for the Andes (Cortés et al., 2016) using snow survey points, site-years of peak annual snow pillow and snow course SWE observations. Verification results showed unbiased posterior SWE estimates with a correlation coefficient of 0.73, RMSE of 0.29 m and mean error less than 0.01 m using snow pillow and snow course peak SWE. Results using snow survey data showed similar unbiased estimates, with a correlation coefficient of 0.50, RMSE of 0.29 m and mean error less than 0.01 m. The daily SWE gridded product generated by Cortés and Margulis (2017) was clipped and averaged within the catchment boundaries to obtain a daily time series for each catchment.

3.1.10 Water rights and reservoirs information

A public reservoir dataset (http://www.ide.cl/descarga/capas/item/embalses-2016.html) was processed to identify the presence of dams within catchments. We also compiled and processed granted water rights available from the Water Atlas (developed by the DGA; DGA, 2016a). This water allocation dataset includes information about the water source (surface or groundwater), the type of right (i.e., consumptive or non-consumptive), its use (i.e., industrial, irrigation, domestic and drinking water, hydroelectric power, pisciculture, mining, and classified as "other uses"), the annual allocated flow (expressed in units of volume per time or as "shares"), and temporal allocation (i.e., permanent and continuous, permanent and discontinuous, permanent and alternated, eventual and continuous, eventual and discontinuous, or eventual and alternated). A detailed explanation of this water right classification can be found in Carey (2014). A key limitation of this dataset is the lack of information on the actual use of granted rights (Larraín, 2006). Additionally, some water right records have incomplete information (e.g., missing coordinates, water volume assigned and temporal allocation).

Figure 5 illustrates water allocation in central-southern Chile (30-43ºS), showing surface and groundwater rights (all types). It is clear that groundwater rights dominate in the central Chile (31°S-36°S), especially in low elevation areas, compared to surface water rights. On the other hand, more surface water rights are granted in southern Chile, especially within high elevation areas towards the Andes.

3.2 Derived catchment attributes

We computed 67 catchment attributes grouped in six classes (Table 2). To motivate the use of common standards in the development of large sample catchment datasets, we included most of the attributes presented by A17 in their CAMELS database. A comparative summary between CAMELS and CAMELS-CL attributes is presented in Table 2, from which one can note that the attributes from classes climatic indices and hydrological signatures were fully adopted from A17. The attributes from the class soils characteristics were not computed at this stage since there is no publicly available national dataset. Given the differences in input datasets, some of the attributes in A17 from the classes location and topography, geologic characteristics, and land cover characteristics, were not computed here. On the other hand, new attributes were derived for the classes location and topography (Sect. 3.2.1), land cover characteristics (Sect. 3.2.3) and hydrological signatures (Sect. 3.2.5). Further, a new class was added to describe the degree of intervention within the catchments (Sect. 3.2.6).

A complete list of catchment attributes included in CAMELS-CL, their description and the corresponding data sources are presented in Table 3. To ensure the reproducibility of our results, the reference to the explicit formulation of climatic indices and hydrological signatures is provided in Table 3. Discussions on the spatial distribution of the catchment attributes, separated by class, are presented in the following sub-sections.

3.2.1 Location and topography

Figure 6 shows six (out of 14) location and topography attributes. Figure 66a presents the elevation of catchment outlets, illustrating two main elevation gradients: (i) a north-to-south (N-S) mean elevation decrease, starting with high elevation basins in the Far North macro-zone – which corresponds to the southern portion of the Altiplano plateau (18°S – 22°S) (Allmendinger et al., 1997) –, towards lower elevations in the southern macro-zones; and (ii) an east-to-west (E-W) gradient, dominated by high elevations in the Andes (located along the east border) decreasing towards sea level at the west border. This gauge elevation attribute can be used to classify catchments based on their location with respect to the coast or the Andes. We proposed the attribute location_type (see Table 3 and Fig. 6f) with three categories: coastal (or low elevation), foothills and altiplano catchments, defined by gauge elevations lower than 50 m a.s.l., between 1,000-1,200 m a.s.l., and above 3,500 m a.s.l., respectively.

Figure 66b reveals smoother N-S and E-W gradients of basin-averaged elevations, compared to gauge elevation gradients (Fig. 6a). This is because the mean elevation calculated for downstream catchments includes nested catchments (located at higher altitudes). The spatial distribution of mean catchment slopes follows different patterns depending on the macro-zone (Fig. 6c). The Far North – dominated by the flat Altiplano Plateau – exhibits relatively small variations in mean slopes, with relatively low values. From Near North to Austral Zone, the mean slope shows a spatial distribution similar to that from mean elevation, with a E-W gradient dominated by high slopes in the Andes and flatter areas towards the sea. In southern Patagonia, such E-W gradient is reversed given the relative position of the Andes.

The spatial distribution of basin areas shows a general increase from east to west (Fig. 6d), which is consistent with smaller headwater catchments at the Andes, and larger downstream catchments towards the sea. Some exceptions to this E-W distribution pattern are catchments located near the east border, featuring either a N-S drainage direction, or a portion of their total contributing area in Argentina (beyond the east national border). Additional exceptions to such E-W distribution are small inner sub-catchments near the west border, or small headwater catchments originated at the Chilean Coastal Range, which runs from north to south along the Pacific coast and reaches up to 3,000 m a.s.l. in the Antofagasta region (Figueroa and Moffat, 2000).

Because all catchments were delineated using available streamflow gauge locations as outlets (Sect. 3.1.1), the contributing area (Fig. 6d) is not necessarily correlated with the number of nested basins within each catchment (Fig. 6e). For example, some small catchments might be highly instrumented (i.e., with many controlled nested basins, because of – for example – water allocation priorities), while large but poorly instrumented catchments might not have inner basins defined.

3.2.2 Geological attributes

Overall, the most common dominant geological classes within CAMELS-CL catchments are acid plutonic rocks (24%), acid volcanic rocks (20%) and pyroclastic (14%). In the Far North zone, there is a strong presence of Pyroclastics, Siliclastic sedimentary rocks and Intermediate volcanic rocks (Fig. 7a and 7b), which can result in the connection of groundwater systems through fractured volcanic rocks (DGA, 1986). This means that there might be differences between surface catchment boundaries (based on topography) and the extension of groundwater systems, which should be considered when analysing the hydrologic response at the basin scale. Figure 7a also indicates that strong geological differences may exist between neighbouring catchments. Furthermore, one can see generally high geological variability within the catchments. Indeed, the dominant geological class covers less than half of the contributing area in most catchments, as indicated by the histogram in Fig. 7c. The presence of carbonate sedimentary rocks is particularly low (Figure 7e), with only 24 catchments having at least 10% of this type of rock. This suggests low formation of karst, a subsurface characteristic featuring large fissures and voids, which results in fast infiltration rates and preferential permeability channels (La Moreaux et al., 1984).

3.2.3 Land cover attributes

As summarised in Table 2, five land cover attributes in A17 were not computed since the land cover map used here (from Zhao *et al.*, 2016) does not provide information about leaf area index, green vegetation fraction, or depth. Instead, we included land cover attributes based on the catchment area encompassed by the main classes of the land cover dataset (Table 3). The first nine land cover attributes described in Table 3 were computed as the percentage of the catchment area covered by levels 1 and 2 land cover classes defined by Zhao *et al.*, (2016). We also computed a forest plantation index to quantify the ratio between forest exotic plantation (mainly *Pinus radiata* and *Eucalyptus spp*) and native forest within a catchment, which is critical information for forest hydrology and ecosystems studies (e.g., Lara *et al.*, 2009a).

Considering that several catchments (almost 50; Fig. 1) extend beyond the Chilean territory, a key limitation of the land cover attributes derived from the map developed by Zhao *et al.*, (2016), is the lack of information outside the national boundary. To account for this, we generated an attribute indicating the fractional catchment area contained within the land cover map, serving also as a quality flag for basin-averaged land cover characteristics.

Figure 8 illustrates a sub-set of the land cover attributes listed in Table 3. Fig. 8a shows the forested (native forests and forest plantation types) catchment area, which prevails in the Southern Zone, Austral Zone and Southern Patagonia. In forested catchments, exotic forest plantations dominate the coastal areas of the Central and Southern Zones, with forest plantation indices up to one (Fig. 8b). Such distribution can be attributed to the extensive land use change experienced in south-central Chile over the last 50 years, where native forests have been progressively converted into agricultural lands and forest plantations (Armesto et al., 2010; Miranda et al., 2015). This conversion has had significant impacts in forest ecosystem services such as water provision (Jones et al., 2017; Lara et al., 2009).

Figures 8c and 8d show that the Far and Near North Zones have more homogeneous land cover types, with shrublands and impervious lands occupying more than 60% of the catchment areas. In southern areas, the coverage of the dominat classes decreases substantially, transitioning towards a mosaic of different land cover types. Missing land cover data is presented in Fig. 8e, which should be accounted for if the land cover attributes of the affected catchments (i.e., the ones with portions in Argentina, as shown in Fig. 2) are used for applications such as catchment classification or parameter regionalisation.

Because of the glaciological contributions to the water balance within the domain (Mernild et al., 2017; Le Quesne et al., 2009), we added two attributes (Table 3) based on information from the glaciers inventory described in Sect. 3.1.4. We found that 255 catchments (48% of the total) have some degree of glacierisation, reaching up to 62% in the Geike River catchment in the Southern Patagonia. The glaciers included in CAMELS-CL span 7,321 km^2, corresponding to almost a quarter of the glacierised area in the Southern Andes (RGI Consortium, 2017). The catchments with the largest degree of glacierisation (more than 15%) are located in the Austral and Southern Patagonia regions, followed by the Olivares and Volcan river catchments (about 14%) in the Central Zone.

3.2.4 Climatic indices

To allow the comparison between CAMELS (A17) and CAMELS-CL, the climatic indices were computed for the same period as in A17, i.e., water years 1990 to 2009, corresponding to 1 April 1990 to 31 March 2010 for Chile. If the indices are required for different periods, the formulae provided in the references from Table 3 can be used with the raw hydro-meteorological time series (available from CAMELS-CL website). The complete spatial and temporal coverages of the meteorological variables allow the estimation of climatic indices for all 516 catchments – in contrast to hydrological signatures, computed for a sub-set of catchments (Sect. 3.2.5). Precipitation and PET-based attributes were calculated for all precipitation products (Sect. 3.1.6), using the daily PET product (Sect. 3.1.8).

The climatic attributes presented in Fig. 4 and 9 reveal basic features of the Chilean climatology, described in more detail by Miller (1976) and Garreaud et al. (2017), among others. Mean annual precipitation ranges from less than 10 mm in the Atacama

Desert (northern Chile) to more than 3,000 mm in western Patagonia (Fig. 4b). Such marked precipitation gradient reflects the relative influence of the subtropical, semi-permanent Southeast Pacific anticyclone, and the frequent incursion of the frontal systems at higher latitudes. The frequency of high precipitation events also increases southward, with a maximum in south-central Chile (Fig. 4d). The Andean domain in the Far North (Chilean Altiplano) is influenced by the monsoonal regime developing over the interior of the continent and receives about 300 mm/year above 4,000 m a.s.l. Superimposed on its N-S gradient, precipitation also varies from east to west due to orographic enhancement over the windward slope of the Andes cordillera (a factor of 2-3 between lowlands and windward slopes; Viale and Garreaud, 2014). PET has a more restricted range than precipitation (400-1,400 mm/year; Fig. 4d). Therefore, the aridity index (PET/P, Fig. 9c) is higher in northern Chile (> 1.0) compared to that of the southern regions (< 1.0). A positive precipitation seasonality (Fig. 9a) in northern Chile indicates precipitation peaks during summer (djf), which is consistent with the monsoonal precipitation events in the Far North during this season (Fig. 9f). In contrast, the negative seasonality values obtained for all macro-zones, except the Far North and Southern Patagonia, illustrate the increased storm frequency and high precipitation events in most of the country during the winter (jja) (Fig. 9a). Seasonality values close to zero indicate uniform precipitation throughout the year in Southern Patagonia (Fig. 9a and 9f). The zero-temperature isotherm during winter storms ranges between 1,500 and 4,000 m a.s.l., so most of the precipitation is liquid along the coast and interior valleys (Fig. 9b), with snow prevailing in high-elevation basins.

3.2.5 Hydrological signatures

Hydrological signatures were computed for the period 1 April 1990 to 31 March 2010, as in Sect. 3.2.4. To exclude the effects anthropic intervention on hydrologic response, we selected 109 catchments with valid daily streamflow records in at least 85% of the period, based on the following criteria: interv_degree lower than 0.1 (i.e., less than 10% of the annual streamflow allocated to surface rights), large_dam equal to zero (absence of large dams within the catchment), imp_frac lower than 5% (negligible urban areas), and copr_frac lower than 20% (negligible irrigation effects). Further, we excluded glacier dominated catchments by selecting glacier_frac lower than 5%. It should be noted that, despite of calculating hydrological indices for a subset of catchments, raw daily time series for all 516 catchments are included in CAMELS-CL database. These time series and the formulae provided in Table 3 may be used if the signatures are required for different time periods.

Figure 10 illustrates the spatial distribution of 12 (out of 14) hydrological signatures (Table 3), revealing the leading patterns of catchment responses. Both mean daily flow and runoff ratio increase from the Far North to the Southern Zone, showing strong correlations with mean annual precipitation (Fig. 4b) and the aridity index (Fig. 9c). Further, a positive west-east gradient (i.e., increase towards the Andes) is observed for runoff ratio and mean half-flow dates within that domain. Higher values of the latter signature can be found in steep (Fig. 6b) snow-dominated (Fig. 9b) basins in Central Chile – where the most frequent season for low precipitation days is Dec-Feb (Fig. 9i).

The mid-segment slope of the flow duration curve (FDC, Fig. 10d) – a signature that quantifies flashiness of runoff – shows that slow basin-averaged responses occur in the Far North and part of the Near North, in spatial correspondence with high baseflow index (Fig. 10e) and low discharge precipitation elasticity (Fig. 10f). Such behaviour is expected in this region due to substantial subsurface and groundwater contributions to total runoff. Although flashiness of runoff and discharge elasticity to precipitation (baseflow index) are relatively higher (lower) and show some correlation towards the south, no clear spatial gradients are observed within the domain spanning from Central Chile to Southern Patagonia.

The examination of signatures related to extreme (high or low) streamflow conditions exposes some interesting features. Although no clear spatial relationship is observed between high flow signatures (Fig. 10g-i), similar spatial distributions of low flow frequencies (Fig. 10j) and mean low flow durations (Fig. 10k) are obtained across the country. Q95 (Fig. 10i) and Q5 (Fig. 10l) provide generally similar patterns to those of mean daily discharge (Fig. 10a), with positive increases from the Far North to the Southern Zones, and a positive west-east gradient. The comparison between the signatures displayed in Fig. 10g-l and climatic indices in Fig. 9d-i highlight the complex relationship between climate and hydrologic catchment behaviour. For example, the spatial structure in the frequency of low/high precipitation days is not reflected in the spatial distribution of high/low flow frequencies. A similar disjunction is observed between the duration of low precipitation (Fig. 9h) and low flow (Fig. 10k) events, whereas those catchments with low duration of high precipitation events also provide low durations in high flow events.

Sharp variations in hydrological signatures (Fig. 10) – in contrast to generally smooth patterns in climate indices (Fig. 9) – are the result of complex, non-linear process interactions across a range of spatiotemporal scales, enhanced by heterogeneities in topography, soils, vegetation, geology and other landscape properties. Careful attention should be paid to such interactions and to the uncertainties involved in the calculation of hydrological signatures, in particular when attempting to extrapolate hydrological behaviour from gauged to ungauged basins based on climatic similarities alone (Westerberg et al., 2016; Westerberg and McMillan, 2015).

3.2.6 Human intervention

Figure 11 summarises water rights records used to characterise human intervention degree within the catchments. We can see that the number of surface rights (Fig. 11a) increases from north to south, while the number of groundwater rights (Fig. 11d) increases from east to west. Although these values do not provide information about allocated volumes, they show how many water rights holders interact to coordinate the water use within a particular catchment. The CAMELS-CL database provides information about each water right within a catchment (not only the attributes representing synthetized information), in case a more detailed analysis is required.

In terms of allocated surface and groundwater flows (Fig. 11b and Fig. 11e, respectively), we only considered consumptive permanent water rights. Further, we considered only water rights recorded as volume per time, since water rights expressed as "shares" (6% of the national water rights database) were not provided with their corresponding conversion into volume units (DGA, 2016b). I should be noted that "shares" rights are the oldest (allocated prior to the 2005 water code reform), thus probably representing a majority of the rights within the Central Zone (region that concentrates the oldest rights).

The above limitations may lead to an underestimation of the allocated flow, due to (at least) the following reasons: (i) non-consumptive rights may have their restitution points outside the catchment boundaries (however, they were not considered allocated flow calculation); (ii) shares rights are disregarded; (iii) there is missing information, and therefore some rights may be omitted (Sect. 3.1.10). On the other hand, allocation estimates may differ considerably from the actual extraction within a catchment. Possible reasons for this are the sub/over use of a granted allocated flow and unauthorised extractions of surface and groundwater.

Despite the limitations of the water use dataset and the attributes presented in Fig. 11, water rights information is still critical to quantify human intervention, and it has not been officially processed at the catchment scale in Chile. To quantify the intervention degree within a catchment, we calculated the interv_degree attribute (described in Table 3 and illustrated in Fig. 11c) as the ratio between the annual surface flow allocated within a catchment, and the catchment mean annual runoff. This attribute indicates how much of the annual runoff generated – in average – within a catchment, corresponds to the water volume allocated as consumptive surface rights. Further, we defined a binary attribute to characterise the presence of reservoirs within a catchment (large_dam in Table 3), using 0 if there are no dams, and 1 if there is at least one dam.

To quantify the urbanised fraction of a catchment – another important factor modulating catchment response –, we used the impervious fractional area attribute (imp_frac in Table 3), which usually contains urban areas. However, this land cover type is the worst classified class, since urban areas have mixed pixels of vegetation and paved surfaces (Zhao et al., 2016). The urban fraction of the catchments (assumed to be equal than imp_frac) varied between 0% and 7% for most catchments (only one catchment had imp_frac = 25%).

4 Uncertainty in precipitation and PET

4.1 Precipitation assessment

To assess precipitation uncertainty, we looked at the inter-product differences across the study domain. To this end, we defined a precipitation spread attribute (p_mean_spread, Table 3) as the standard deviation of basin-averaged mean annual precipitation from the four different products, normalised by multi-product mean. To allow such inter-comparison, we used data from the concurrent period 1998-2014 (Table 2), and excluded catchments located southern than 50°S (since CHIRPS and TMPA cover up to 50°S). Given the different nature of the assessed precipitation products, the spread attribute can be interpreted as a measure of precipitation uncertainty. The underlying assumption is that similar values from different data sources indicate regions with higher confidence in precipitation estimates.

Figure 12 displays catchment-scale mean precipitation and the precipitation spread index for three macro-regions: North (northern than 34°S), which includes the Far North and Near North macro-zones; Central-South (between 34°S and 43°S); and Austral-Patagonia (southern than 43°S). Mean precipitation estimates (p_mean) have a larger spread in the North (Fig. 12a), which indicates larger uncertainties in this domain. We attribute these higher relative errors to the methodological challenges in detecting events and estimating their intensities in this arid sub-domain, where the occurrence of precipitation events is relatively rare (note the different scale used for p_mean in Fig. 12d). By contrast, considerably larger precipitation amounts (Fig. 12e-f) and lower spread values (Fig. 12b-c) are obtained in Central-South and Austral-Patagonia.

Although the effects of large precipitation uncertainty on streamflow modelling in the North are not straightforward to determine, some insights can be gained from our analyses. First, surface runoff is not very sensitive to variations in precipitation (i.e., small runoff elasticity values in Fig. 10f), suggesting a weak propagation of precipitation errors within a model. Second, groundwater has the largest contribution to streamflow in this domain (largest baseflow indices in Fig. 10e and sedimentary rocks as the most common geologic class illustrated in Fig. 7a), indicating how critical is to pursue a realistic representation of groundwater mechanisms in numerical models. Additionally, aquifer boundaries may be quite different from surface catchment boundaries, and therefore accurate delineations are needed to ensure a good representation of surface-groundwater interactions (e.g., Sar et al. 2015; Arkoprovo et al. 2012; Ivkovic et al. 2009).

We note that the ensemble spread of precipitation estimates is a measure of disagreement among the various products rather than a measure of accuracy, which should be quantified using ground observations (e.g., Zambrano-Bigiarini et al. 2017). Such analysis is beyond the scope of this paper, since the assessment of different precipitation products at the basin scale is typically conducted by forcing one or more hydrological models with the different precipitation datasets over the selected study area (e.g., Bisselink et al. 2016; Thiemig et al. 2013; Su et al. 2008).

As an alternative to the model-based approach, we examined the consistency of catchment precipitation estimates based on the long-term runoff ratios in 119 near-natural catchments (Fig. 12g-i), selected following the criteria presented in Sect. 3.2.5. Although there are large inter-product differences in runoff ratios in the North (consistent with large p_mean_spread values in Fig. 12a), relatively low runoff ratio values (<0.4) are obtained, as expected given the arid and semi-arid conditions in this region. By contrast, there are catchments with runoff ratios larger than 1 in Central-South and Austral-Patagonia, indicating that there is more water leaving the catchment than the total amount entering as precipitation. Assuming that streamflow data and catchment area are reliable, and that changes in storage and groundwater contributions are negligible, such cases indicate precipitation underestimation by the various products. In the Central-South (Austral-Patagonia) region, the MSWEP (CR2MET) dataset provides 8% (20%) of catchments runoff ratio > 1 – i.e., the smallest among all products. In both domains, the TMPA dataset provides the largest fraction of catchments with runoff ratio > 1 (54% in Central-South and 70% in Austral-Patagonia). Such underestimation of TMPA, as well as other satellite precipitation estimates, was also reported by Hobouchian et al., (2017) and Zambrano-Bigiarini et al. (2017).

To further explore differences and systematic biases in within the assessed products, we used the Budyko framework (Budyko, 1971) to diagnose the factors affecting the quality of the precipitation datasets. This framework links climate to catchment runoff and evapotranspiration in a simple and easy-to-interpret visualisation. Figure 13 shows the evaporative index (EI, the ratio of mean annual evapotranspiration to the mean annual precipitation), estimated as one minus the runoff ratio (i.e., assuming that changes in storage and groundwater contributions are negligible; Sposito, 2017), as a function of the aridity index for the 119 near-natural catchments over the period 1998-2014. This figure illustrates how the evapotranspiration and runoff rates within this highly diverse set of catchments are governed by the available energy and precipitation, e.g., – for a given amount of precipitation – runoff exceeds evapotranspiration when the available energy and PET are relatively low (points below the energy limited line in humid regions).

Negative EI values in Fig. 13 represent non-behavioural combinations of precipitation and runoff (Berghuijs et al., 2017). Under the assumption that precipitation estimates represent a relatively larger source of uncertainty compared to runoff, all the points with EI < 0 indicate those catchments where mean annual precipitation is underestimated (i.e., runoff ratios > 1). Therefore, Fig. 13a-d indicate that all precipitation products systematically fail in humid catchments with steep topography (slopes greater than 150 m km^{-1}), in agreement with the limitations reported for different satellite precipitation products over the same domain (Hobouchian et al., 2017). The systematic precipitation underestimation can be attributed to the complex topography of headwater catchments and the scarcity of ground stations at high elevations. In fact, 90% of the 500 rain gauges located south of 34°S are placed below 1,000 m a.s.l.

4.2 PET assessment

To assess the quality of the PET products described in Sect. 3.1.8, we used a different approach than in Sect. 4.1, since a basin-scale PET estimation cannot be evaluated based on the observed streamflow. In this case, the evaluation was made with an independent set of PET, which was calculated with daily observations for the period 2010-2014 from 75 meteorological stations maintained by the Chilean National Institute of Agricultural Research (INIA, 2017). For each site, we calculated two PET time series: (i) a daily time series calculated by using the Hargreaves formulae (Hargreaves and Allen, 2003) fed with INIA temperature observations, called $INIA_{har}$ hereafter, and (ii) an 8-day accumulated time series based on the FAO Penman-Monteith reference crop evapotranspiration (Allen et al., 1998), called $INIA_{ET0}$. $INIA_{har}$ and $INIA_{ET0}$ were used to evaluate the corresponding pixels of PET_{har} and PET_{mod}, respectively.

The evaluation metrics used in these comparisons, spatially averaged within the macro-zones, are summarised in Table 4. These results indicate good agreement between PET_{har} and $INIA_{har}$, with correlation coefficients (r) greater than 0.92 throughout the national territory, except in the Far-North macro-zone, where we found a weaker correlation (r = 0.76). The ratios between mean PET_{har} and mean $INIA_{har}$ indicate that PET_{har} underestimates (overestimates) up to an 8% the observation-based PET in the Far and Near North arid regions (the Southern and Austral Zones humid regions).

The comparison between PET_{mod} and the $INIA_{ET0}$ led to r values greater than 0.80 within the domain, except for the Far North, where the correlation was below 0.20 in all available stations. The ratios between the means of PET_{mod} and $INIA_{ET0}$ indicate that the first one systematically overestimates stations estimates, which was also found by Westerhoff, (2015). These systematic biases may be explained by the theoretical differences between $INIA_{ET0}$ and PET_{mod} calculated in MOD16. $INIA_{ET0}$ represents a potential condition for a regular crop height of 0.12 m and a fixed surface resistance and albedo, which is not the case for the PET_{mod}, which includes a more complete parameterisation of those variables according to vegetation characteristics. Further, $INIA_{ET0}$ uses local meteorological observations, while PET_{mod} uses global sources that may not capture meteorological variations at the local scale. If an application requires it (e.g., irrigation or hydrological modelling applications), the biases reported for PET_{mod} can be corrected with conventional statistical methods (e.g., Maraun and Widmann, 2018). Other spatio-temporal analyses (e.g., drought monitoring) may directly apply PET_{mod} due to its high correlation with ground ET_0 estimates.

Since INIA records were used differently for evaluating PET_{har} and PET_{mod}, a direct comparison between both assessments is not possible, although they provide valuable information about the quality of each PET product across the territory. Furthermore, the formulation behind the two gridded products have different trade-offs. PET_{mod} is based on the Penman-Monteith equation that solves the surface energy balance, including parameters such as albedo and FPAR/LAI, whereas PET_{har} is calculated from an empirical approach based only on air temperature. PET_{har} has a coarser spatial resolution (5×5 km^2) compared to PET_{mod} (1×1 km^2), which may induce to larger errors over complex topography (e.g., mountain catchments) due to the local variations in potential evapotranspiration with changes in slope and aspect. On the other hand, PET_{har} covers a longer period (1979-2016, the same as T_{min} and T_{max} from Sect. 3.1.7) compared to PET_{mod} (2000-2014), which is more suitable to establish climatic trends.

5 Impacts of human intervention on catchment behaviour

Large sample hydrology is a suitable framework to explore anthropic impacts on catchment behaviour through comparative analysis over a broad range of hydroclimatic conditions and catchment characteristics. Such assessment constitute a fundamental element to be considered when addressing the question of how climate change will affect global water supply (Vörösmarty et al., 2007). However, it remains unsolved how to generalise the results from different studies. For example, Poff et al. (2006) examined the effects of land use on hydrological regimes (e.g., peak and low flows, runoff variability) in 158 basins within the CONUS, finding region-dependent changes in specific metrics. Ochoa-Tocachi et al. (2016) analysed the impacts of land use on the hydrology of 25 Andean catchments, finding that anthropogenic influences propagate towards increased streamflow variability and decreased catchment regulation capacity and water yield. More recently, Tijdeman et al. (2018) examined the effects of human intervention on streamflow drought characteristics across 187 catchments in England and Wales, concluding that most human-influenced catchments did not have drought characteristics different from those expected for near-natural conditions.

In this work, we used hydrological signatures to describe catchment behaviour and the interv_degree attribute (Sect. 3.2.6) to characterise the level of human intervention. Figure 14 (panels a-d) display scatter plots between four hydrological signatures and the logarithm of the intervention degree index – which accounts for consumptive, continuous surface water rights. Different colours indicate the aridity of each catchment, which is a major driver of hydrological behaviour (as showed in Fig. 13). These plots show that larger human intervention is associated to decreased annual flows and runoff ratios, especially in drier catchments. Interestingly, a larger number of consumptive surface rights (larger interv_degree values) is reflected on decreased elasticity of runoff with respect to precipitation, supported by low p-values. Since these scatter plots do not allow to separate the effects of aridity and human intervention, we binned the data and used boxplots to disentangle such effects. Figures 14e-h show the boxplots for these four hydrological signatures for the classified catchments, binned according to their aridity (humid: aridity below 0.8, medium: aridity between 0.8 and 1.5, and arid: aridity above 1.5) and their degree of human intervention. Catchments with low (high) intervention were defined by the interv_degree attribute lower (greater) than 5% and the large_dam attribute equal to zero (one).

The dispersion in hydrological signatures among wet and medium catchments is large, and no significant difference is found between catchments with high and low human intervention. In contrast, dry catchments (zoomed view in Fig. 14i-l) reveal differences in hydrological signatures for high and low human intervention (i.e., median values in highly disturbed catchments are below the first quantile of low intervention catchments). In agreement with the scatter plots, the annual flows (Fig. 14i) and runoff ratios (Fig. 14j) decrease in catchments with larger number of consumptive surface rights, which is expected due to withdrawals of water within water-scarce regions. Further, highly disturbed arid catchments feature lower runoff sensitivities to precipitation compared to less disturbed ones (Fig. 14k), which could be attributed to altered runoff generation mechanisms associated to water withdrawals and reservoirs. Figure 14l shows that there is likely less variation in daily runoff – represented by the mid-segment slope of the flow duration curves – within highly disturbed arid catchments. The results found in arid catchments (i.e., water-scarce regions) provide new evidence of the potential impacts of human intervention on water supply, however, further research is needed to assess the causality of the correlations found here.

6 Concluding remarks

The CAMELS-CL dataset presented here provides novel information at the catchment scale in Chile, a region that is largely underrepresented in large-sample studies. CAMELS-CL includes daily streamflow data and a suite of basin-averaged hydro-meteorological variables, including precipitation, temperature, potential evapotranspiration, and snow water equivalent for 516 catchments in the country. The dataset also includes shapefiles of the drainage area boundaries related to streamflow gauge locations, overcoming the main limitations of the official national hydrographic network (DGA; CIREN, 2014). Further, we synthesised diverse and complementary datasets to compute 67 catchment attributes describing topography, geology, land cover, climate, hydrology, and anthropic intervention.

In this paper, we described the advantages and main limitations of the datasets used to derive the various catchment attributes, which should be considered when using CAMELS-CL for selecting catchments and interpreting results. The main spatial patterns of catchment attributes and their inter-relationships were analysed across the entire domain (4,300 km), which includes high altitude catchments and five different primary climatic regimes.

The CAMELS-CL dataset is used, further, to assess hydro-meteorological biases and uncertainties in a large ensemble of watersheds in Chile, based on the contrast of various precipitation products – one national (CR2MET) and three widely used global products (CHIRPS, MSWEP and TMPA). Large discrepancies between products were detected in arid regions, which were explained by the methodological challenges associated with the very rare occurrence of events in this region. Based on a water balance analysis using Budyko curves, we found systematic precipitation underestimation in headwater mountain catchments (high elevations and steep slopes) over humid regions. For these topographic characteristics and climatic conditions, all products failed at estimating precipitation amounts that closed the water balance, with the TMPA product featuring the largest errors – in agreement with previous studies over the same domain (Hobouchian et al., 2017; Zambrano-

Bigiarini et al., 2017). Such errors were attributed to the complex topography of headwater catchments and the scarcity of ground stations at high elevations (90% of rain gauges located south of 34°S are placed below 1,000 m a.s.l.). These limitations restrict our understanding of hydrological processes, posing challenges for streamflow modelling, water management and allocation. To alleviate these constraints, efforts should be put in improving the surface monitoring network at high elevations (> 1,000 m a.s.l.). This would help to improve remotely-sensed and model-based precipitation estimates in complex terrains.

Further, we assessed the PET_{har} and PET_{mod} products with an independent observation-based PET. In general, both products showed good correlations with the observation-based PET for the complete domain, except in the Far North arid region. Regarding mean biases, PET_{har} showed slight underestimation (overestimation) in the Far and Near North arid regions (the Southern and Austral Zones humid regions), compared with the observed-based PET. PET_{mod} on the other hand, showed a systematic and more significant overestimation of the observation-based PET within the complete domain, which was attributed to the theoretical differences between their formulations.

The assessment of precipitation and PET products showed different performances within the domain. Therefore, the choice of these products or similar datasets must be carefully made based on the application and requirement of the study.

Finally, we used CAMELS-CL to explore the impact of human activities on catchment behaviour, a key element to be considered when evaluating climate change impacts on water supply. We showed that larger human intervention is correlated with lower than normal annual flows, runoff ratios, elasticity of runoff with respect to precipitation, and flashiness of runoff, especially in drier catchments. These results not only illustrate how catchment behaviour can change with human intervention, but also reveal the potential of anthropic indices to predict shifts in hydrological systems.

In summary, this paper contributes to the scientific community by: (i) providing a unique dataset that can be used to advance our understanding of hydrological systems by learning from diversity, (ii) analysing the dominant spatial patterns of physical, climatic and hydrological attributes within the domain, (iii) assessing the quality of one national and three global precipitation datasets based on the observed water balance, (iv) assessing two PET products based on an independent set of PET point values calculated from meteorological records, and (v) examining the interplay between human intervention and changes in observed catchment response .

Other research questions that can addressed with CAMELS-CL may be related with – but not limited to – catchment classification, similarity and regionalisation, model parameter estimation, dominant controls on runoff generation, the impacts of different land cover types on catchment response, characterisation of drought history and projections, and climate change impacts on hydrological processes. CAMELS-CL will be continuously updated to incorporate new records and new datasets, which may include soils characteristics, water quality, seismology records, socio-economic indices and energy generation data. Additionally, new and more detailed information about the Chilean cryosphere will be included, complementing the global inventory processed here with national inventories of Chile and Argentina. Time series of streamflow, meteorological variables, and all catchment attributes described in this paper are available from the Center for Climate and Resilience Research website (http://www.cr2.cl/download/camels-cl).

Acknowledgements

This research emerged from the collaboration with many colleagues at the Center for Climate and Resilience Research (CR2, CONICYT/FONDAP/15110009). Camila Alvarez-Garreton is funded by FONDECYT Postdoctoral Grant N°3170428. Pablo Mendoza received additional support from FONDECYT Postdoctoral Grant N° 3170079. Mauricio Zambrano-Bigiarini thanks FONDECYT 11150861 for financial support. The development of CR2MET was supported by the Chilean Water Directorate (DGA), through the National Water Balance Updating Project DGA-2319.

References

Addor, N., Newman, A. J., Mizukami, N. and Clark, M. P.: The CAMELS data set: Catchment attributes and meteorology for large-sample studies, Hydrol. Earth Syst. Sci., 21(10), 5293–5313, doi:10.5194/hess-21-5293-2017, 2017.

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P. and Nelkin, E.: The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present), J. Hydrometeorol., 4(6), 1147–1167, doi:10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2, 2003.

Allen, R. G., Pereira, L. S., Raes, D. and Smith, M.: FAO Penman-Monteith Equation, in Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements, pp. 17–28., 1998.

Allmendinger, R. W., Jordan, T. E., Kay, S. M. and Isacks, B. L.: The evolution of the Altiplano-Puna Plateau of the Central Andes, Annu. Rev. Earth Planet. Sci., 25(1), 139–174, doi:10.1146/annurev.earth.25.1.139, 1997.

Andréassian, V., Hall, A., Chahinian, N. and Schaake, J.: Introduction and Synthesis : Why should hydrologists work on a large number of basin data sets ?, IAHS-AISH Publ. 307, 1–6, 2006.

Arkoprovo, B., Adarsa, J. and Prakash, S. S.: Delineation of Groundwater Potential Zones using Satellite Remote Sensing and Geographic Information System Techniques : A Case study from Ganjam district , Orissa , India, Res. J. Recent Sci., 1(9), 59–66, 2012.

Armesto, J. J., Manuschevich, D., Mora, A., Smith-Ramirez, C., Rozzi, R., Abarzúa, A. M. and Marquet, P. A.: From the Holocene to the Anthropocene: A historical framework for land cover change in southwestern South America in the past 15,000 years, Land use policy, 27(2), 148–160, doi:10.1016/j.landusepol.2009.07.006, 2010.

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Munõz-Sabater, J., Pappenberger, F., De Rosnay, P., Stockdale, T. and Vitart, F.: ERA-Interim/Land: A global land surface reanalysis data set, Hydrol. Earth Syst. Sci., 19(1), 389–407, doi:10.5194/hess-19-389-2015, 2015.

Beck, H. E., Van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B. and De Roo, A.: MSWEP: 3-hourly 0.25° global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data, Hydrol. Earth Syst. Sci., 21(1), 589–615, doi:10.5194/hess-21-589-2017, 2017.

Berghuijs, W. R., Sivapalan, M., Woods, R. A. and Savenije, H. H. G.: Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales, Water Resour. Res., 50(7), 5638–5661, doi:10.1002/2014WR015692, 2014.

Berghuijs, W. R., Larsen, J. R., van Emmerik, T. H. M. and Woods, R. A.: A Global Assessment of Runoff Sensitivity to Changes in Precipitation, Potential Evaporation, and Other Factors, Water Resour. Res., 53(10), 8475–8486, doi:10.1002/2017WR021593, 2017.

Bisselink, B., Zambrano-Bigiarini, M., Burek, P. and de Roo, A.: Assessing the role of uncertain precipitation estimates on the robustness of hydrological model parameters under highly variable climate conditions, J. Hydrol. Reg. Stud., 8, 112–129, doi:10.1016/j.ejrh.2016.09.003, 2016.

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A. and Savenije, H.: Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales., 2013.

Budyko, M. I.: Climate and Life., 1971.

Carey: General Overview of Water Rights in Chile., 2014.

Cortés, G. and Margulis, S.: Impacts of El Niño and La Niña on interannual snow accumulation in the Andes: Results from a high-resolution 31 year reanalysis, Geophys. Res. Lett., 1–9, doi:10.1002/2017GL073826, 2017.

Cortés, G., Girotto, M. and Margulis, S. A.: Analysis of sub-pixel snow and ice extent over the extratropical Andes using spectral unmixing of historical Landsat imagery, Remote Sens. Environ., 141, 64–78, doi:10.1016/j.rse.2013.10.023, 2014.

Cortés, G., Girotto, M. and Margulis, S.: Snow process estimation over the extratropical Andes using a data assimilation framework integrating MERRA data and Landsat imagery, Water Resour. Res., 52(4), 2582–2600, doi:10.1002/2015WR018376, 2016.

DGA; CIREN: Redefinición de la clasificación red hidrográfica a nivel Nacional, 2014.

DGA: Mapa Hidrogeológico de Chile., 1986.

DGA: Glaciares de chile, 2014.

DGA: Atlas del Agua - Chile 2016. Capítulo 1: Chile en el mundo, Atlas del Agua Chile 2016, 24, 2016a.

DGA: Atlas del Agua - Chile 2016. Capítulo 4: Gestion del agua, Atlas del Agua Chile 2016, 30, 2016b.

DGA: Actualización del Balance Hídrico Nacional, SIT N° 417, Ministerio de Obras Públicas, Dirección General de Aguas, División de Estudios y Planificación, Santiago, Chile. Realizado por: Universidad de Chile & Pontificia Universidad Católica de Chile., 2017.

Ehret, U., Gupta, H. V., Sivapalan, M., Weijs, S. V., Schymanski, S. J., Bl??schl, G., Gelfan, A. N., Harman, C., Kleidon, A., Bogaard, T. A., Wang, D., Wagener, T., Scherer, U., Zehe, E., Bierkens, M. F. P., Di Baldassarre, G., Parajka, J., Van Beek, L. P. H., Van Griensven, A., Westhoff, M. C. and Winsemius, H. C.: Advancing catchment hydrology to deal with predictions under change, Hydrol. Earth Syst. Sci., 18(2), 649–671, doi:10.5194/hess-18-649-2014, 2014.

Figueroa, D. and Moffat, C.: On the influence of topography in the induction of coastal upwelling along the Chilean coast, Geophys. Res. Lett., 27(23), 3905–3908, doi:10.1029/1999GL011302, 2000.

Friedl, M. ., McIver, D. ., Hodges, J. C. ., Zhang, X. ., Muchoney, D., Strahler, A. ., Woodcock, C. ., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F. and Schaaf, C.: Global land cover mapping from MODIS: algorithms and early results, Remote Sens. Environ., 83(1–2), 287–302, doi:10.1016/S0034-4257(02)00078-0, 2002.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. and Michaelsen, J.: The climate hazards infrared precipitation with stations - A new environmental record for monitoring extremes, Sci. Data, 2, doi:10.1038/sdata.2015.66, 2015.

Garreaud, R., Alvarez-Garreton, C., Barichivich, J., Boisier, J. P., Christie, D., Galleguillos, M., LeQuesne, C., McPhee, J. and Zambrano-Bigiarini, M.: The 2010-2015 mega drought in Central Chile: Impacts on regional hydroclimate and vegetation, Hydrol. Earth Syst. Sci. Discuss., 1–37, doi:10.5194/hess-2017-191, 2017.

Garreaud, R. D.: The Andes climate and weather, Adv. Geosci., 22, 3–11, doi:10.5194/adgeo-22-3-2009, 2009.

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., Li, C., Li, X., Fu, W., Liu, C., Xu, Y., Wang, X., Cheng, Q., Hu, L., Yao, W., Zhang, H., Zhu, P., Zhao, Z., Zhang, H., Zheng, Y., Ji, L., Zhang, Y., Chen, H., Yan, A., Guo, J., Yu, L., Wang, L., Liu, X., Shi, T., Zhu, M., Chen, Y., Yang, G., Tang, P., Xu, B., Giri, C., Clinton, N., Zhu, Z., Chen, J. and Chen, J.: Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data, Int. J. Remote Sens., 34(7), 2607–2654, doi:10.1080/01431161.2012.748992, 2013.

Google: Google Earth Pro, Google [online] Available from: https://www.google.com/earth/, 2016.

Di Gregorio, A. and Jansen, L. J. .: Land Cover Classification System. Classification concepts and user manual., 2005.

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M. and Andréassian, V.: Large-sample hydrology: A need to balance depth with breadth, Hydrol. Earth Syst. Sci., 18(2), 463–477, doi:10.5194/hess-18-463-2014, 2014.

Hargreaves, G. H. and Allen, R. G.: History and evaluation of Hargreaves evapotranspiration equation, J. Irrig. Drain. Eng., 129(1), 53–63, doi:10.1061/(ASCE)0733-9437(2003)129:1(53), 2003.

Hargreaves, G. H. and Samani, Z. a.: Reference crop evapotranspiration from temperature, Appl. Eng. Agric., 1(2), 96–99, doi:10.13031/2013.26773, 1985.

Hartmann, J. and Moosdorf, N.: The new global lithological map database GLiM: A representation of rock properties at the Earth surface, Geochemistry, Geophys. Geosystems, 13(12), 1–37, doi:10.1029/2012GC004370, 2012.

Hijmans, R. J.: Raster: Geographic Data Analysis and modeling, R Packag. version 2.5-8. https//CRAN.R-project.org/package=raster, 1, r948, 2016.

Hobouchian, M. P., Salio, P., García Skabar, Y., Vila, D. and Garreaud, R.: Assessment of satellite precipitation estimates over the slopes of the subtropical Andes, Atmos. Res., 190, 43–54, doi:10.1016/j.atmosres.2017.02.006, 2017.

Howell, T. . and Evett, S. R.: The Penman-Monteith Method, Bushland, Texas USDA Agric. Res. Serv., 5646(806), 2001.

Huang, C., Newman, A. J., Clark, M. P., Wood, A. W. and Zheng, X.: Evaluation of snow data assimilation using the ensemble Kalman filter for seasonal streamflow prediction in the western United States, Hydrol. Earth Syst. Sci., 21(1), 635–650, doi:10.5194/hess-21-635-2017, 2017.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P. and Stocker, E. F.: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation

Estimates at Fine Scales, J. Hydrometeorol., 8(1), 38–55, doi:10.1175/JHM560.1, 2007.

Huffman, G. J., Adler, R. F., Bolvin, D. T. and Nelkin, E. J.: The TRMM Multi-satellite Precipitation Analysis (TMPA), in Satellite Rainfall Applications for Surface Hydrology, pp. 3–22., 2010.

Huss, M. and Hock, R.: A new model for global glacier change and sea-level rise, Front. Earth Sci., 3, doi:10.3389/feart.2015.00054, 2015.

IGM, I. G. M.: Hidrografía, in Geografía de Chile, p. 19., 1984.

INIA: AGROMET. Instituto Nacional de Investigacion Agriola. Ministerio de Agricultura., [online] Available from: https://agromet.cl, 2017.

Ivkovic, K. M., Letcher, R. A. and Croke, B. F. W.: Use of a simple surface-groundwater interaction model to inform water management, in Australian Journal of Earth Sciences, vol. 56, pp. 71–80., 2009.

Jin, Y., Schaaf, C. B., Woodcock, C. ., Gao, F., Li, X. and Strahler, A. H.: Consistency of MODIS surface bidirectional reflectance distribution function and albedo retrievals: 2. Validation, J. Geophys. Res., 108(D5), 4159, doi:10.1029/2002JD002804, 2003.

Jones, J., Almeida, A., Cisneros, F., Iroumé, A., Jobbágy, E., Lara, A., Lima, W. de P., Little, C., Llerena, C., Silveira, L. and Villegas, J. C.: Forests and water in South America, Hydrol. Process., 31(5), 972–980, doi:10.1002/hyp.11035, 2017.

Kottek, M., Grieser, J., Beck, C., Rudolf, B. and Rubel, F.: World map of the Köppen-Geiger climate classification updated, Meteorol. Zeitschrift, 15(3), 259–263, doi:10.1127/0941-2948/2006/0130, 2006.

Ladson, A.; Bronw, R.; Neal, B.; Nathan, R.: A standard approach to baseflow separation using the Lyne and Hollick filter, Tech. Pap., 17(1), 25–34, doi:http://dx.doi.org/10.7158/W12-028.2013.17.1., 2013.

Lara, A., Little, C., Urrutia, R., McPhee, J., Álvarez-Garretón, C., Oyarzún, C., Soto, D., Donoso, P., Nahuelhual, L., Pino, M. and Arismendi, I.: Assessment of ecosystem services as an opportunity for the conservation and management of native forests in Chile, For. Ecol. Manage., 258(4), 415–424, doi:10.1016/j.foreco.2009.01.004, 2009.

Larraín, S.: El agua en Chile: entre los derechos humanos y las reglas del mercado, Http://Polis.Revues.Org, (14), 2006.

Liston, G. E.: Representing subgrid snow cover heterogeneities in regional and global models, J. Clim., 17(6), 1381–1397, doi:10.1175/1520-0442(2004)017<1381:RSSCHI>2.0.CO;2, 2004.

Lucht, W., Schaaf, C. B. and Strahler, A. H.: An algorithm for the retrieval of albedo from space using semiempirical BRDF models, IEEE Trans. Geosci. Remote Sens., 38(2 II), 977–998, doi:10.1109/36.841980, 2000.

Maraun, D. and Widmann, M.: Statistical Downscaling and Bias Correction for Climate Research, , 170–200, 2018.

Margulis, S. A., Cortés, G., Girotto, M. and Durand, M.: A Landsat-Era Sierra Nevada Snow Reanalysis (1985–2015), J. Hydrometeorol., 17(4), 1203–1221, doi:10.1175/JHM-D-15-0177.1, 2016.

Marzeion, B., Jarosch, a. H. and Hofer, M.: Past and future sea-level change from the surface mass balance of glaciers, Cryosph., 6(4), 1295–1322, doi:10.5194/tc-6-1295-2012, 2012.

McDonnell, J. J. and Woods, R.: On the need for catchment classification, J. Hydrol., 299(1–2), 2–3, doi:10.1016/j.jhydrol.2004.09.003, 2004.

Mernild, S. H., Liston, G. E., Hiemstra, C. and Wilson, R.: The Andes Cordillera. Part III: glacier surface mass balance and contribution to sea level rise (1979–2014), Int. J. Climatol., 37(7), 3154–3174, doi:10.1002/joc.4907, 2017.

Miller, A.: The climate of Chile, in World survey of climatology, pp. 113–145., 1976.

Miranda, A., Altamirano, A., Cayuela, L., Pincheira, F. and Lara, A.: Different times, same story: Native forest loss and landscape homogenization in three physiographical areas of south-central of Chile, Appl. Geogr., 60, 20–28, doi:10.1016/j.apgeog.2015.02.016, 2015.

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

La Moreaux, P. E., Wilson, B. M. and Memon, B. A.: Guide to the hydrology of carbonate rocks., 1984.

Mu, Q., Zhao, M. and Running, S.: Brief Introduction to MODIS Evapotranspiration Data Set (MOD16), Water Resour. Res., (2005), 1–4, 2005.

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M., Morisette, J. T., Votava, P., Nemani, R. R. and Running, S. W.: Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data, Remote Sens. Environ., 83(1–2), 214–231, doi:10.1016/S0034-4257(02)00074-3, 2002.

Neteler, M., Bowman, M. H., Landa, M. and Metz, M.: GRASS GIS: A multi-purpose open source GIS, Environ. Model. Softw., 31, 124–130, doi:10.1016/j.envsoft.2011.11.014, 2012.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T. and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance, Hydrol. Earth Syst. Sci., 19(1), 209–223, doi:10.5194/hess-19-209-2015, 2015.

Ochoa-Tocachi, B. F., Buytaert, W., De Bièvre, B., Célleri, R., Crespo, P., Villacís, M., Llerena, C. A., Acosta, L., Villazón, M., Guallpa, M., Gil-Ríos, J., Fuentes, P., Olaya, D., Viñas, P., Rojas, G. and Arias, S.: Impacts of land use on the hydrological response of tropical Andean catchments, Hydrol. Process., 30(22), 4074–4089, doi:10.1002/hyp.10980, 2016.

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 Resour. Res., 44(3), 1–15, doi:10.1029/2007WR006240, 2008.

Pellicciotti, F., Ragettli, S., Carenzo, M. and McPhee, J.: Changes of glaciers in the Andes of Chile and priorities for future work, Sci. Total Environ., 493, 1197–1210, doi:10.1016/j.scitotenv.2013.10.055, 2014.

Poff, N. L. R., Bledsoe, B. P. and Cuhaciyan, C. O.: Hydrologic variation with land use across the contiguous United States: Geomorphic and ecological consequences for stream ecosystems, Geomorphology, 79(3–4), 264–285, doi:10.1016/j.geomorph.2006.06.032, 2006.

QGIS Development Team: QGIS Geographic Information System, Open Source Geospatial Found. Proj., doi:http://www.qgis.org/, 2015.

Le Quesne, C., Acuña, C., Boninsegna, J. A., Rivera, A. and Barichivich, J.: Long-term glacier variations in the Central Andes of Argentina and Chile, inferred from historical records and tree-ring reconstructed precipitation, Palaeogeogr. Palaeoclimatol. Palaeoecol., 281(3–4), 334–344, doi:10.1016/j.palaeo.2008.01.039, 2009.

RGI Consortium: Randolph Glacier Inventory -- A Dataset of Global Glacier Outlines: Version 6.0., 2017.

Ropelewski, C. F., Janowiak, J. E. and Halpert, M. S.: The Climate Anomaly Monitoring System (CAMS), in Climate Analysis Center, NSW, NOAA, Washigton, DC. [Available from the Climate Prediction Center, Camp Springs, MD 20746], p. 39., 1984.

Sar, N., Khan, A., Chatterjee, S. and Das, A.: Hydrologic delineation of ground water potential zones using geospatial technique for Keleghai river basin, India, Model. Earth Syst. Environ., 1(3), 25, doi:10.1007/s40808-015-0024-3, 2015.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A. and Carrillo, G.: Catchment classification: Empirical analysis of hydrologic similarity based on catchment function in the eastern USA, Hydrol. Earth Syst. Sci., 15(9), 2895–2911, doi:10.5194/hess-15-2895-2011, 2011.

Sernageomin: Mapa geologico de chile: version digital, Publ. Geol. Digit., 4, 25, 2004.

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J. J., Mendiondo, E. M., O'Conell, P. E., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S. and Zehe, E.: IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences, Hydrol. Sci. J., 48(6), 857–880, doi:10.1623/hysj.48.6.857.51421, 2003.

Sposito, G.: Understanding the budyko equation, Water (Switzerland), 9(4), 1–14, doi:10.3390/w9040236, 2017.

Strahler, A. N.: Quantitative analysis of watershed geomorphology, Eos, Trans. Am. Geophys. Union, 38(6), 913–920, doi:10.1029/TR038i006p00913, 1957.

Su, F., Hong, Y. and Lettenmaier, D. P.: Evaluation of TRMM Multisatellite Precipitation Analysis (TMPA) and Its Utility in Hydrologic Prediction in the La Plata Basin, J. Hydrometeorol., 9(4), 622–640, doi:10.1175/2007JHM944.1, 2008.

Tachikawa, T., Hato, M., Kaku, M. and Iwasaki, A.: Characteristics of ASTER GDEM version 2, in International Geoscience and Remote Sensing Symposium (IGARSS), pp. 3657–3660., 2011.

Thiemig, V., Rojas, R., Zambrano-Bigiarini, M. and De Roo, A.: Hydrological evaluation of satellite-based rainfall estimates over the Volta and Baro-Akobo Basin, J. Hydrol., 499, 324–338, doi:10.1016/j.jhydrol.2013.07.012, 2013.

Tian, Y. and Peters-Lidard, C. D.: A global map of uncertainties in satellite-based precipitation measurements, Geophys. Res. Lett., 37(24), doi:10.1029/2010GL046008, 2010.

Tijdeman, E., Hannaford, J. and Stahl, K.: Human influences on streamflow drought characteristics in England and Wales, Hydrol. Earth Syst. Sci., 22(2), 1051–1064, doi:10.5194/hess-22-1051-2018, 2018.

Viale, M. and Garreaud, R.: Summer Precipitation Events over the Western Slope of the Subtropical Andes, Mon. Weather Rev., 142(3), 1074–1092, doi:10.1175/MWR-D-13-00259.1, 2014.

Vörösmarty, C. J., Vo, C. J. and Green, P.: Global Water Resources : Vulnerability from Climate Change and Population

Growth, Science (80-.)., 284(2000), doi:10.1126/science.289.5477.284, 2007.

Wagener, T., Sivapalan, M., Troch, P. and Woods, R.: Catchment Classification and Hydrologic Similarity, Geogr. Compass, 1, 1–31, doi:10.1111/j.1749-8198.2007.00039.x, 2007.

Westerberg, I. K. and McMillan, H. K.: Uncertainty in hydrological signatures, Hydrol. Earth Syst. Sci., 19(9), doi:10.5194/hess-19-3951-2015, 2015.

Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A. and Freer, J.: Uncertainty in hydrological signatures for gauged and ungauged catchments, Water Resour. Res., 52(3), doi:10.1002/2015WR017635, 2016.

Westerhoff, R. S.: Using uncertainty of Penman and Penman-Monteith methods in combined satellite and ground-based evapotranspiration estimates, Remote Sens. Environ., 169, 102–112, doi:10.1016/j.rse.2015.07.021, 2015.

Woldemeskel, F. M., Sivakumar, B. and Sharma, A.: Merging gauge and satellite rainfall with specification of associated uncertainty across Australia, J. Hydrol., 499, 167–176, doi:10.1016/j.jhydrol.2013.06.039, 2013.

Wood, A. W., Hopson, T., Newman, A., Brekke, L., Arnold, J. and Clark, M.: Quantifying Streamflow Forecast Skill Elasticity to Initial Condition and Climate Prediction Skill, J. Hydrometeorol., 17(2), 651–668, doi:10.1175/JHM-D-14-0213.1, 2016.

Yang, Z. L., Dickinson, R. E., Robock, A. and Vinnikov, K. Y.: Validation of the snow submodel of the biosphere-atmosphere transfer scheme with Russian snow cover and meteorological observational data, J. Clim., 10(2), 353–373, doi:10.1175/1520-0442(1997)010<0353:VOTSSO>2.0.CO;2, 1997.

Zambrano-Bigiarini, M., Nauditt, A., Birkel, C., Verbist, K. and Ribbe, L.: Temporal and spatial evaluation of satellite-based rainfall estimates across the complex topographical and climatic gradients of Chile, Hydrol. Earth Syst. Sci., 21(2), 1295–1320, doi:10.5194/hess-21-1295-2017, 2017.

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

Figures

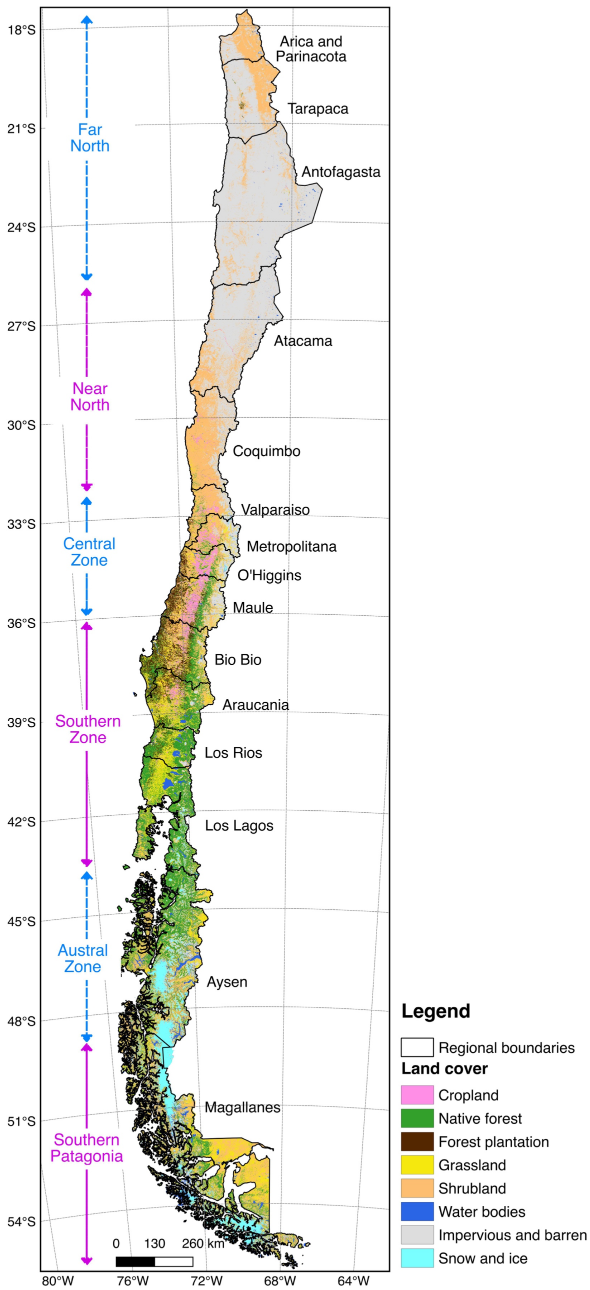

Figure 1: Chilean regional boundaries and names, and the six defined macro-zones (blue and magenta arrows). The background colour correspond to the main land cover classes, obtained from Zhao et al., (2016).

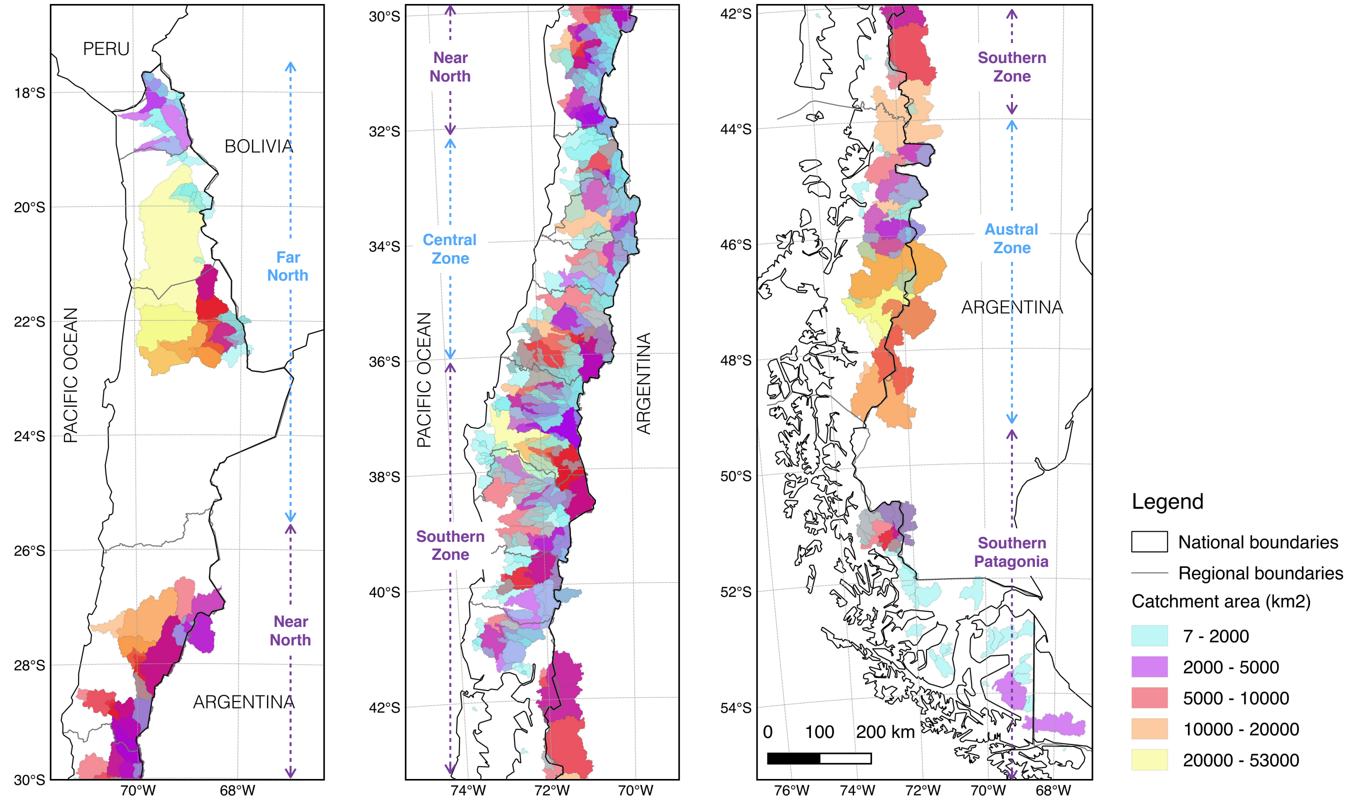

Figure 2: Catchment boundaries and contributing areas (km^2) of the 516 watersheds included in this study. The six defined macro-zones are indicated in blue and purple arrows.

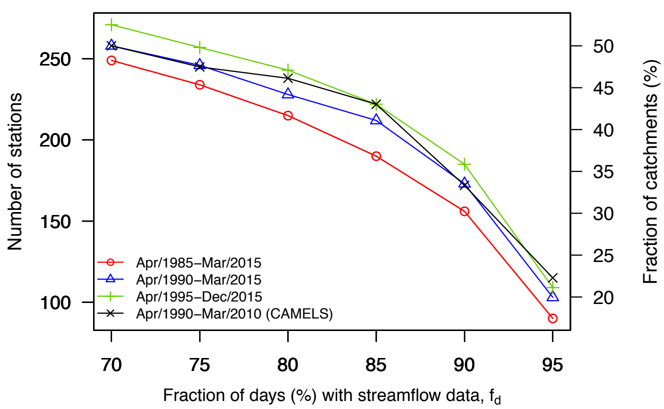

Figure 3: Number of stations (left y-axis) having at least f_d % of days with daily streamflow records, for different periods. The right y-axis shows the percentage of catchments (out of 516) that meets the criterion. The period used in the CAMELS dataset (Addor et al., 2017) was included as a reference.

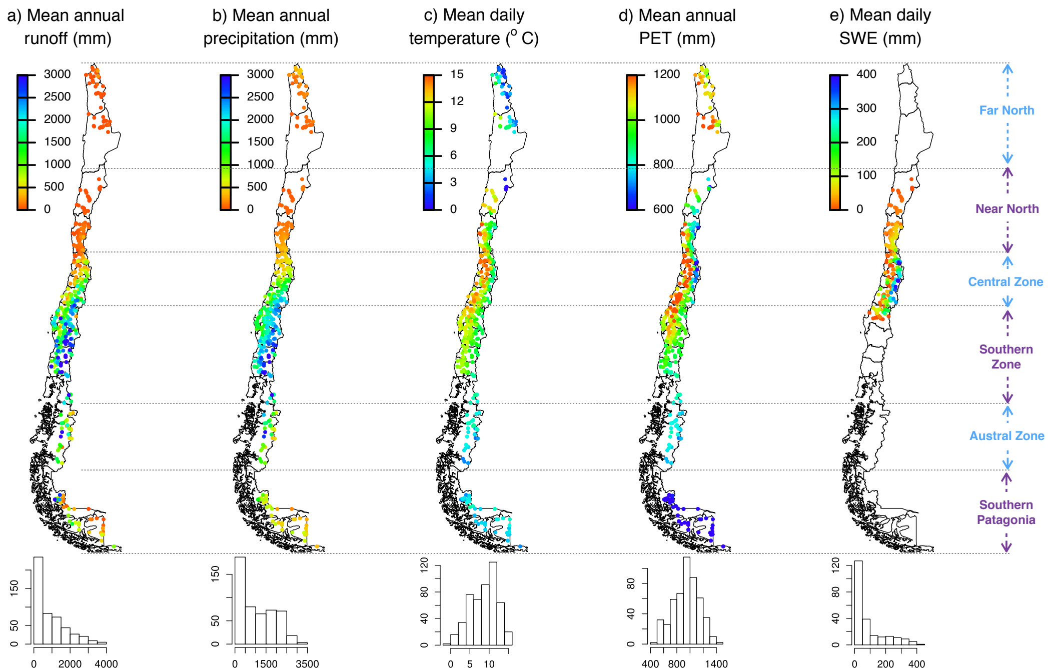

Figure 4: Mean annual hydro-meteorological variables, calculated for the complete recording period of each variable. Panels b and d were generated with, $\text{precip}_{\text{cr2met}}$ and pet_{har} products, respectively. The histograms indicate the number of catchments (out of 516) in each bin. The points represent the location of catchment outlets.

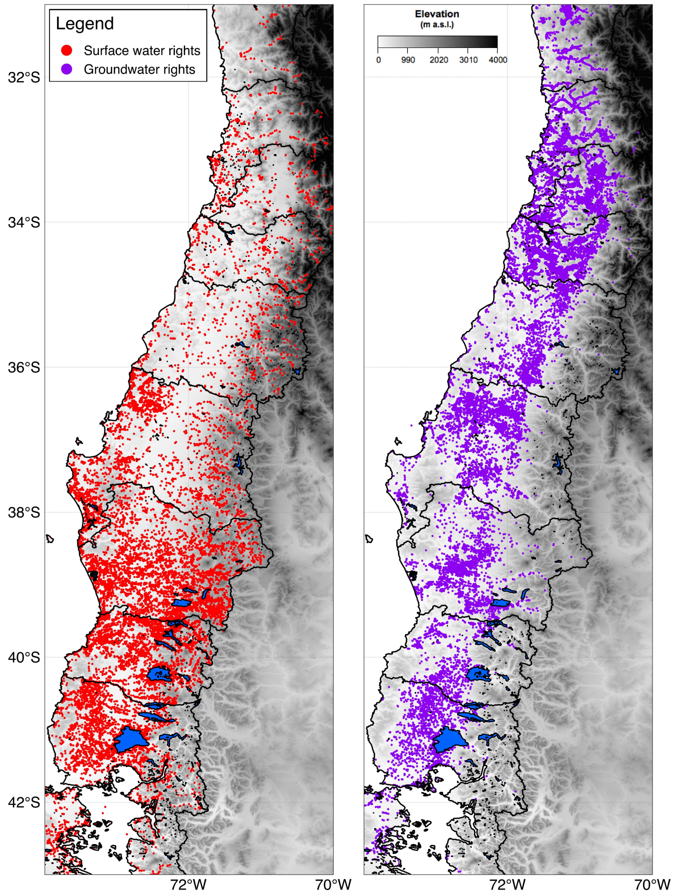

Figure 5: Surface (left panel) and ground (right panel) water rights (consumptive, non-consumptive, permanent, eventual and alternated) granted by the Chilean Water Directorate (DGA) for a portion of the country. Background colours represent topography (greyscale) and the main water bodies (highlighted in blue).

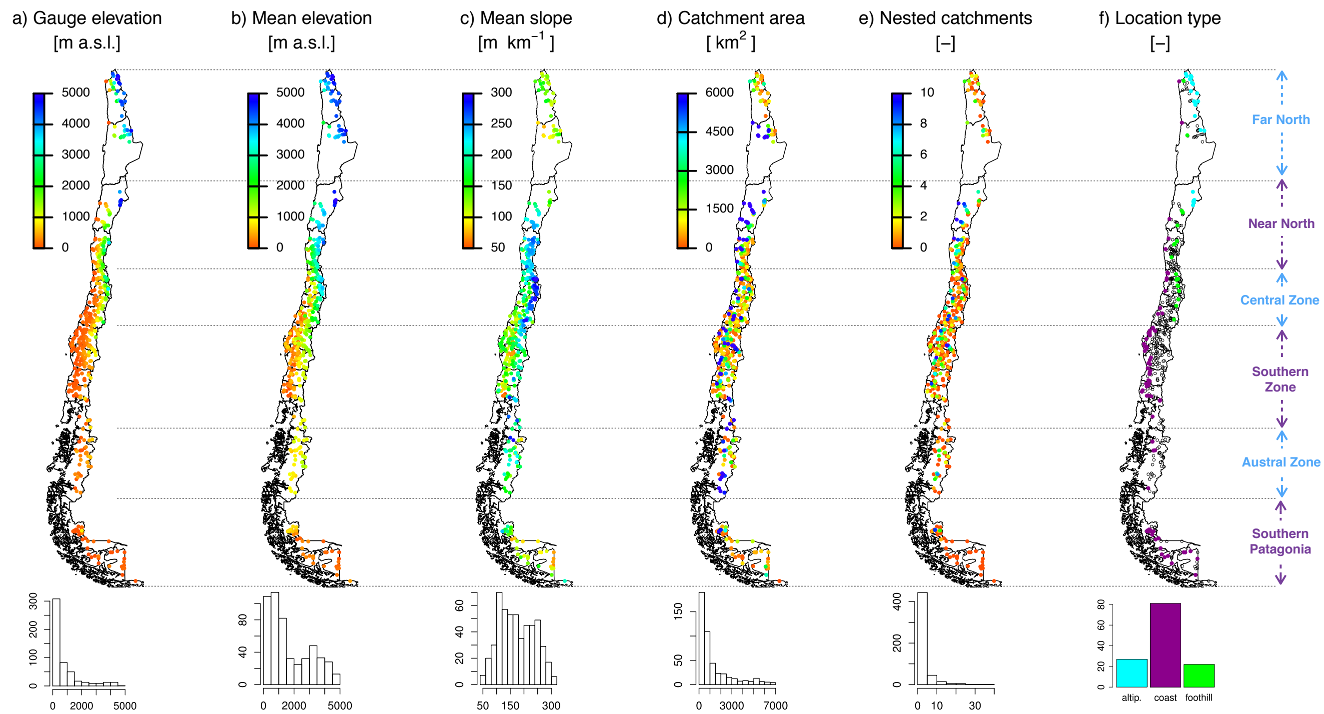

Figure 6: Location and topography. For visualization purposes, catchment areas (panel d) are shown up to their 90th percentile. The histograms indicate the number of catchments (out of 516) in each bin.

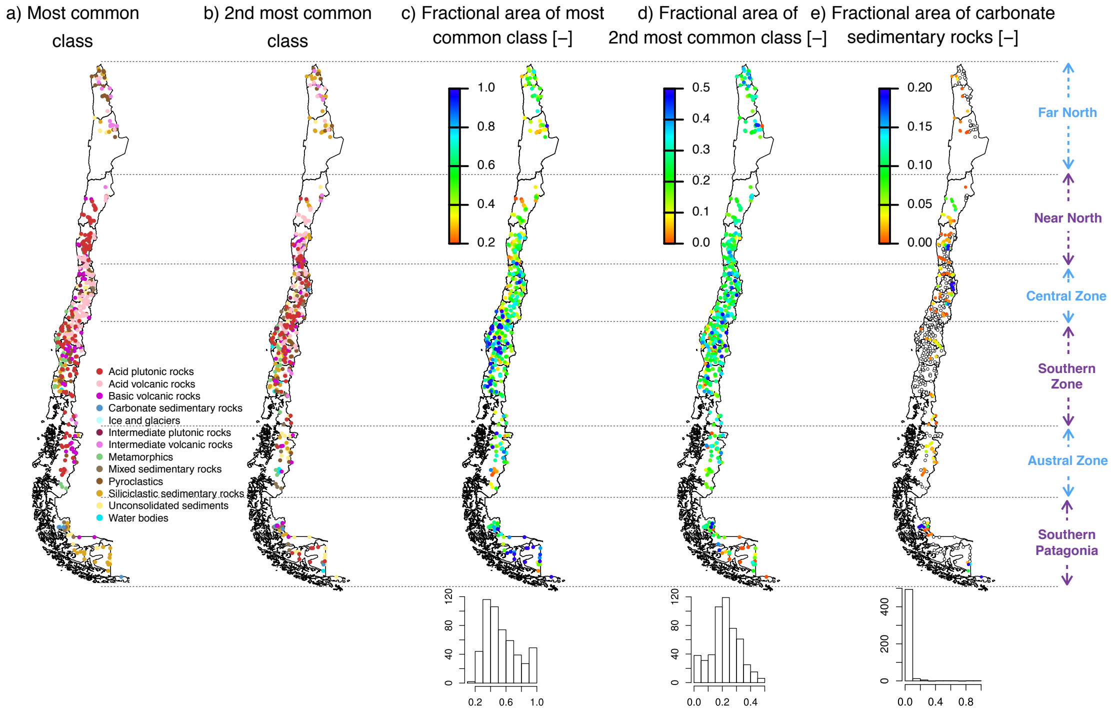

Figure 7: Geology attributes. The histograms indicate the number of catchments (out of 516) in each bin.

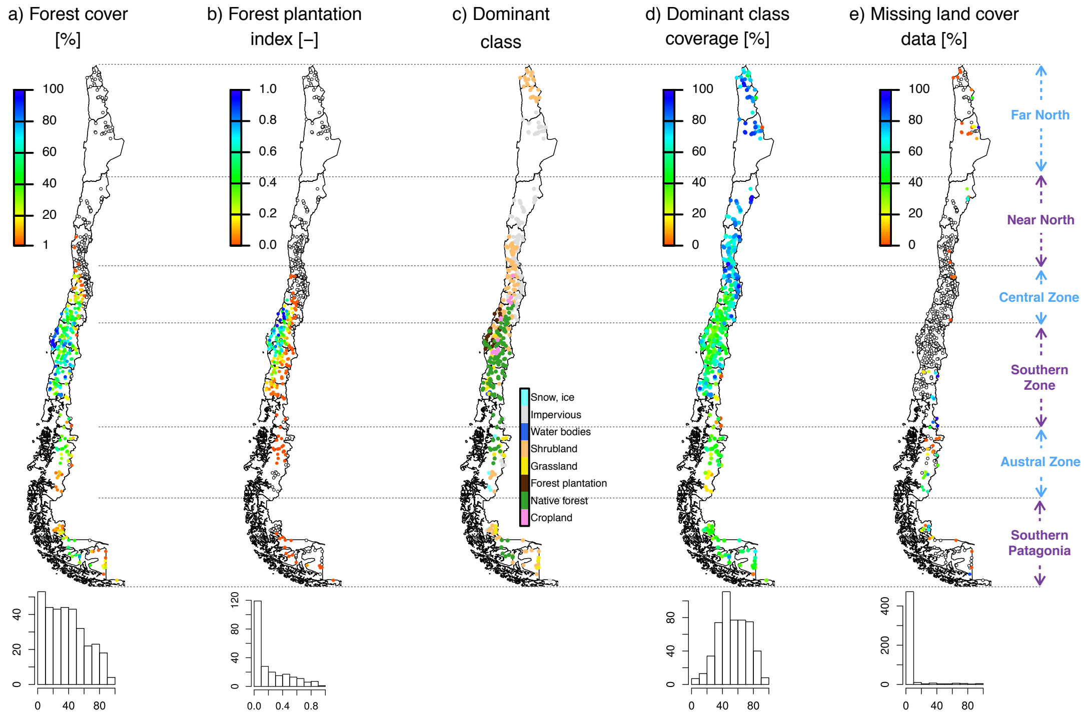

Figure 8: Land cover characteristics. Values below colour bar lower limits are shown blank (i.e., for panel e, this means that there is no missing land cover data within those catchments). The histograms indicate the number of catchments (out of 516) in each bin.

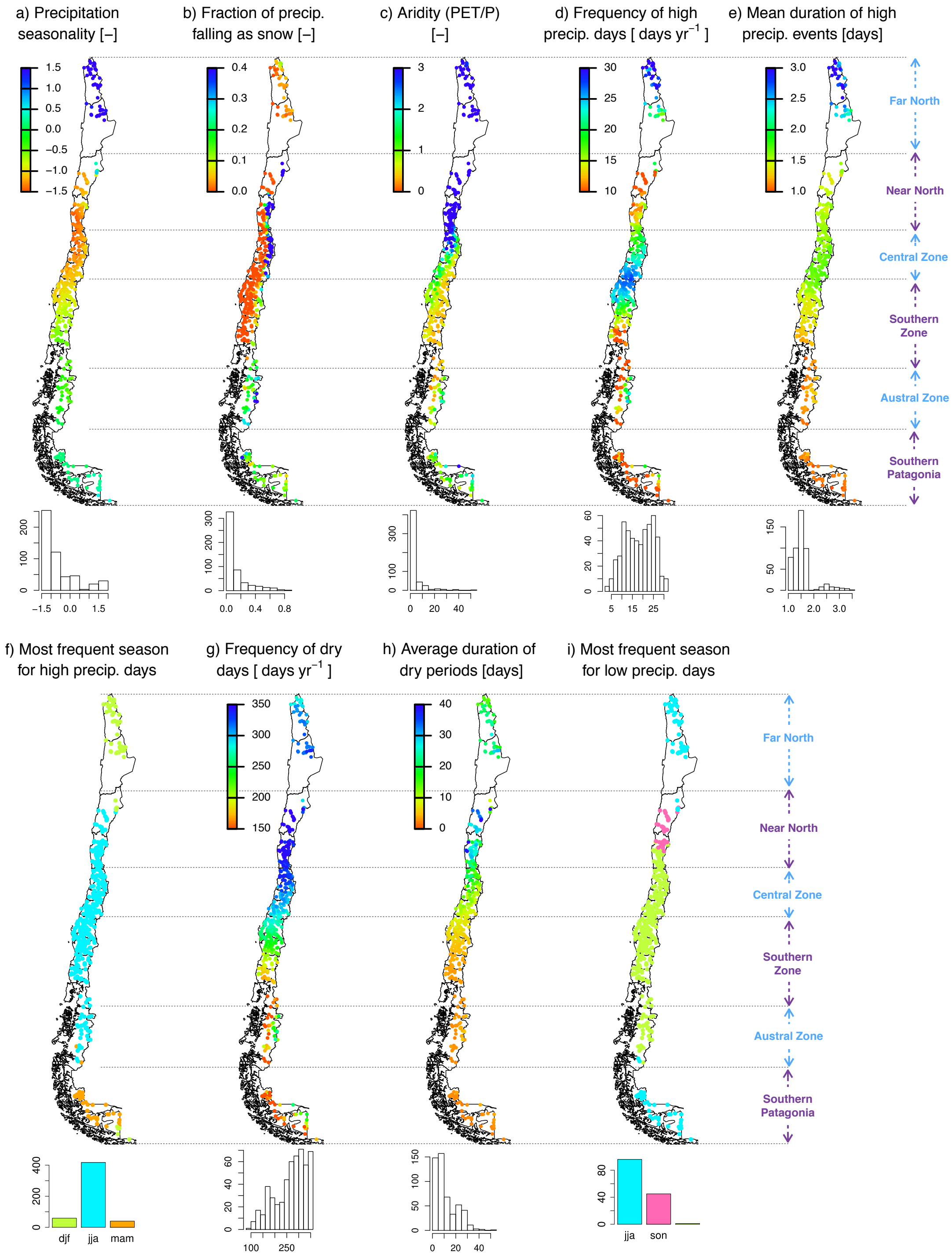

Figure 9: Climatic indices (calculated from precip$_{cr2met}$ product). The histograms indicate the number of catchments (out of 516) in each bin.

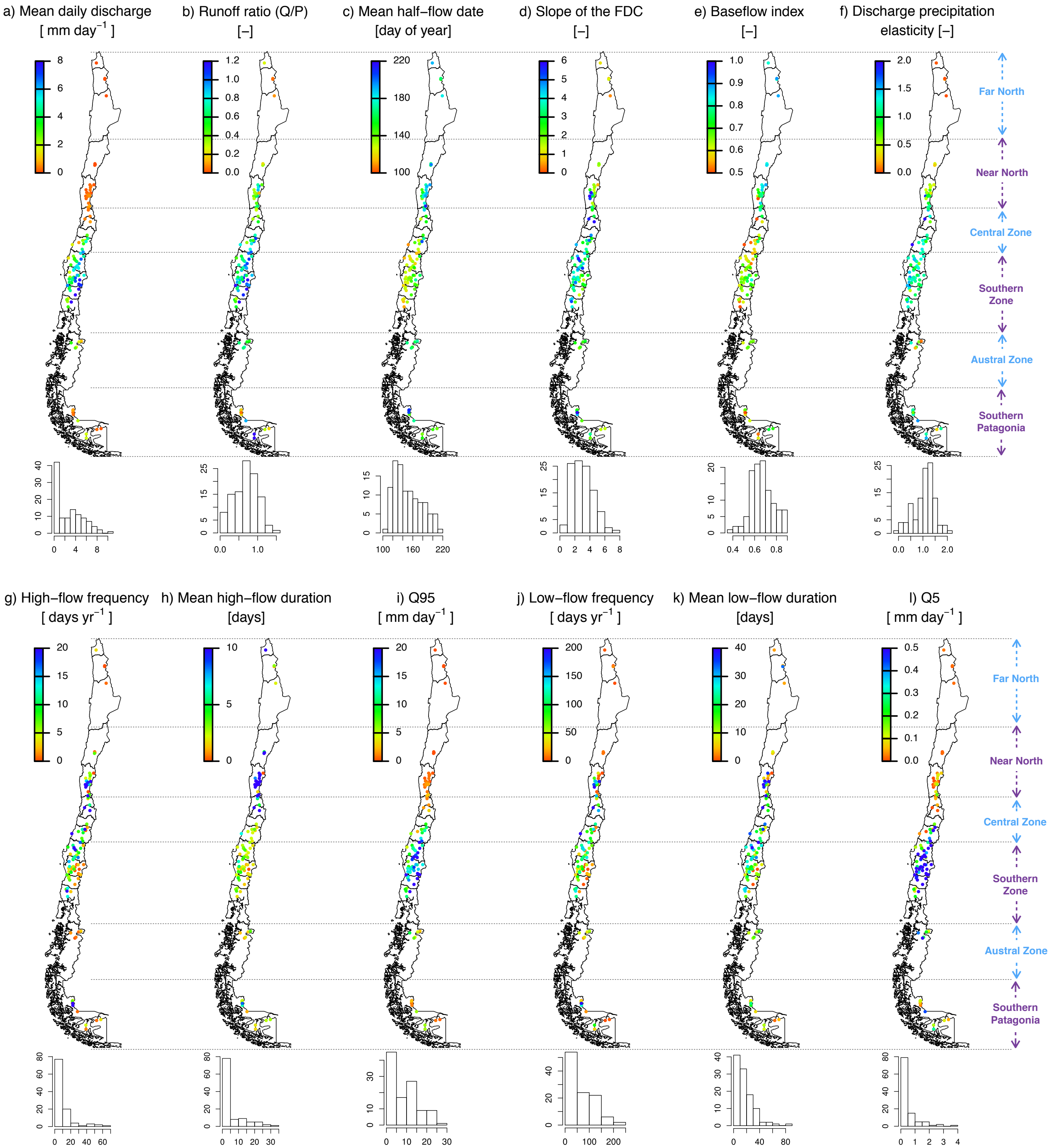

Figure 10: Hydrologic signatures for 109 near-natural catchments. The histograms indicate the number of catchments (out of 109) in each bin.

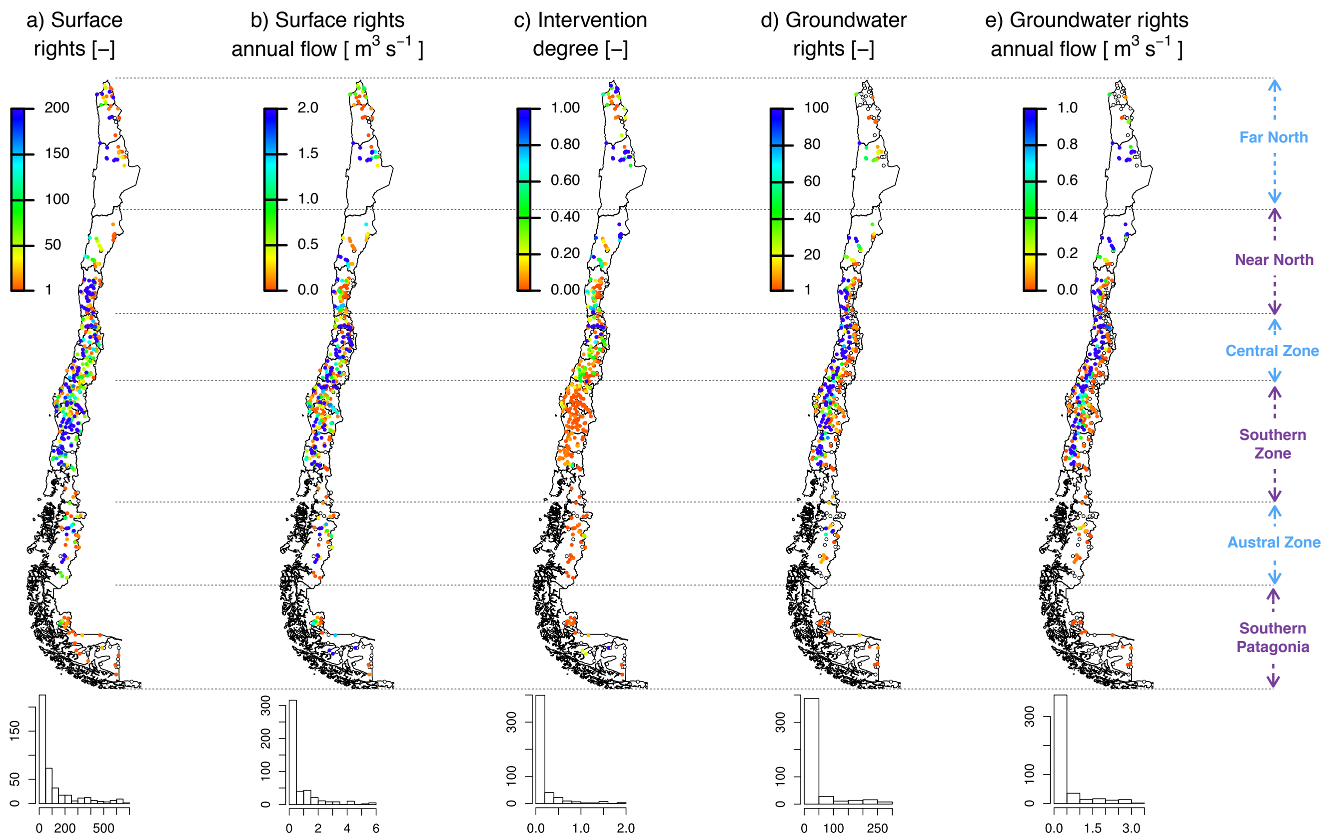

Figure 11: Water rights attributes. For visualization purposes, the attributes in panels a, b, d and e are shown up to their 90th percentile. Attributes below the lower colour bar value are blank and above the upper colour bar value are blue. The histograms indicate the number of catchments (out of 516) in each bin.

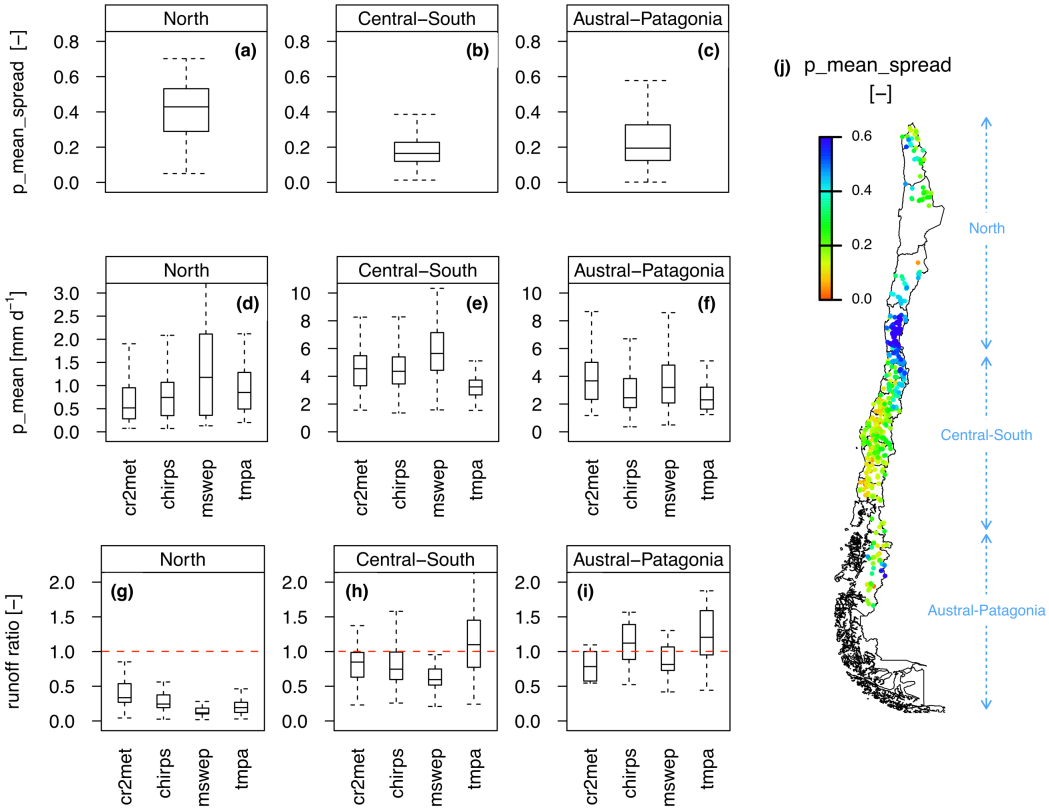

Figure 12: Precipitation spread (p_mean_spread in panels a-c), mean annual precipitation (panels d-f), and runoff ratio (panels g-i), for the different precipitation products. The domain was divided in three main regions: North (northern than 34°S); Central-South; and Austral-Patagonia. Panel j shows the spatial distribution of p_mean_spread in these sub-regions.

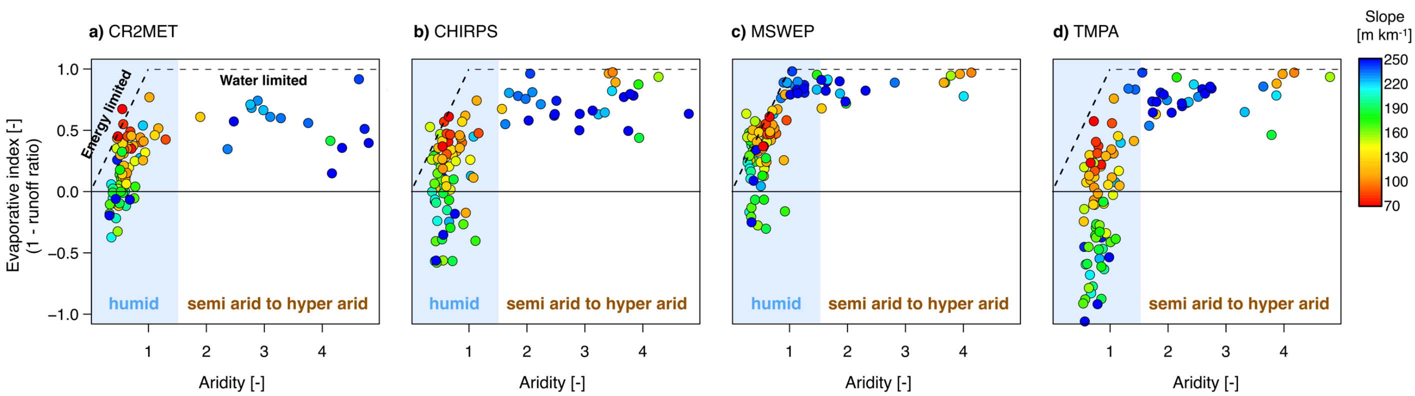

Figure 13: Water balance for 109 near-natural catchments, illustrated in a Budyko scheme for CR2MET (panel a), CHIRPS (panel b), MSWEP (panel c) and TMPA (panel d). Markers are coloured by the catchment mean slope.

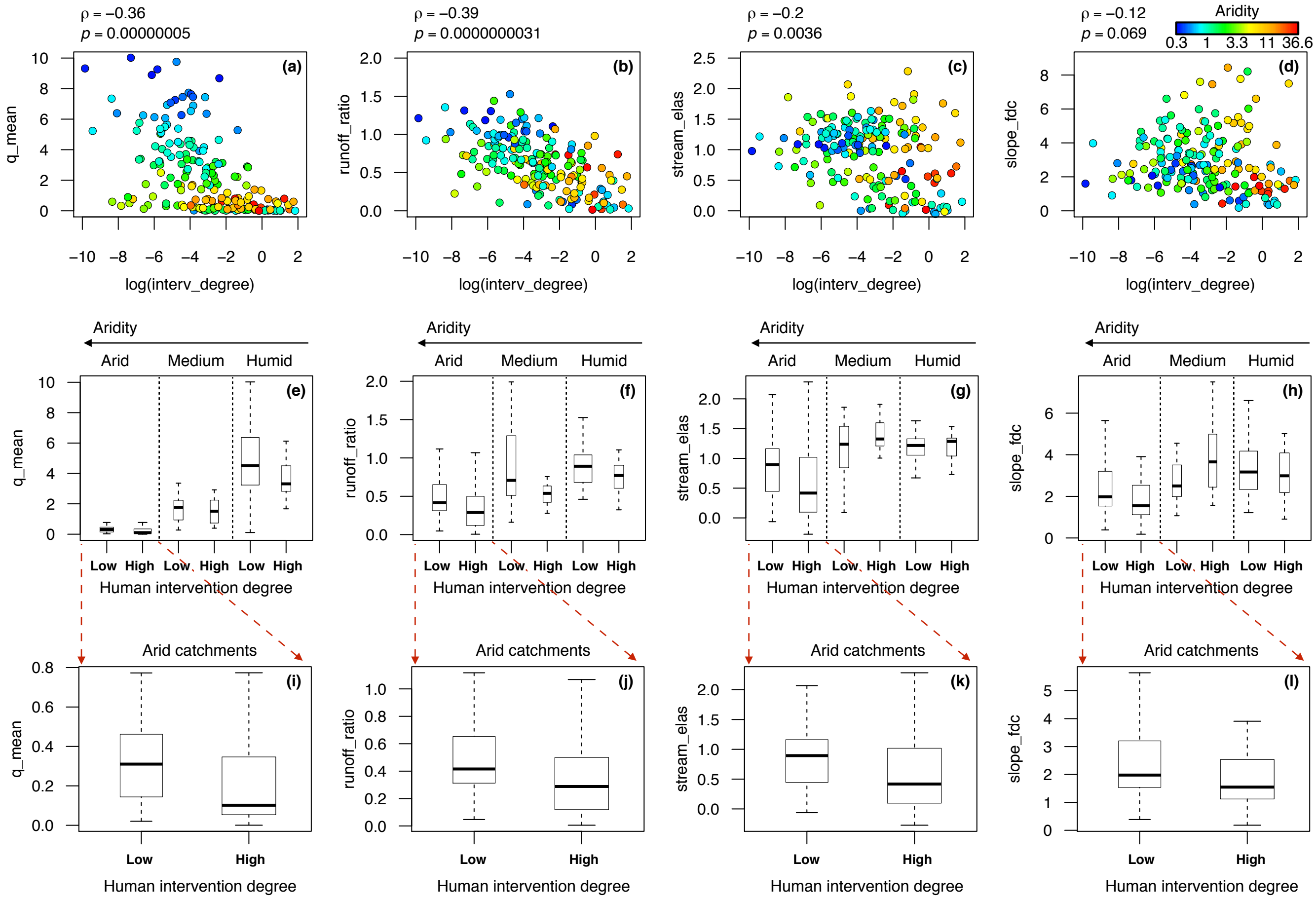

Figure 14: Panels a-d present the relation between four hydrological signatures and the log-transformed human intervention degree ("interv_degree" from Table 3). The spearman rank correlation coefficients and their p-values at 95% confidence are shown in each plot. The colour corresponds to the aridity index. Panels e-h show the boxplots (box widths are proportional to the number of catchments in each box) of the hydrological signatures for the catchments classified by their aridity (humid: aridity below 0.8, medium: aridity between 0.8 and 1.5, and arid: aridity above 1.5) and by their human intervention degree (low: interv_degree below 5%, and high: interv_degree greater than 5%). Panels i-l present a zoomed view of the arid catchments.

Tables

Table 1: Precipitation products

Name	*Description*	*Spatial resolution*	*Temporal resolution*	*Period of record*
precip$_{cr2met}$	Obtained from de CR2METv1.3 dataset (DGA, 2017)	0.05° lat-lon	daily	1979-2016
precip$_{tmpa}$	Obtained from TMPA 3B42v7 dataset (Huffman et al., 2007, 2010)	0.25° lat-lon	daily	1998-2016
precip$_{chirps}$	Obtained from Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) version 2 dataset (Funk et al., 2015)	0.05° lat-lon	daily	1981-2016
precip$_{mswep}$	Obtained from the Multi-Source Weighted-Ensemble Precipitation (MSWEP) v1.1 dataset (Beck et al., 2017)	0.25° lat-lon	daily	1979-2016

Table 2: Summary of attributes computed in CAMELS and CAMELS-CL.

Attribute class	*CAMELS (A17)*	*CAMELS-CL*
Location and topography	9 attributes	6 attributes adopted from A17 8 additional attributes
Geology	7 attributes	5 attributes adopted from A17
Soils characteristics	11 attributes	not computed
Land cover characteristics	8 attributes	3 attributes adopted from A17 13 additional attributes
Climatic indices	11 attributes	11 attributes adopted from A17 1 additional attribute
Hydrological signatures	13 attributes	13 attributes adopted from A17 1 additional attribute
Intervention degree	not computed	6 attributes
Total number of attributes	**59**	**38 adopted from A17** **29 introduced**

Table 3: Summary of catchment attributes. Climate indices and hydrological signatures were computed for the period 01/01/1990-31/12/2010. Index i refers to the precipitation product (i = 1, 2, 3, 4 for $precip_{cr2met}$, $precip_{tmpa}$, $precip_{chirps}$ and $precip_{mswep}$, respectively).

Attribute class	*Attribute name*	*Description*	*Unit*	*Data source*	*Reference*
Location and topography	gauge_id	catchment identifier (corresponds to the station code provided by DGA)	-	Gauges information collected from http://explorador.cr2.cl	Section 3.1
	gauge_name	gauge name (based on DGA records)	-		
	gauge_lat	gauge latitude (based on DGA records)	° South		
	gauge_lon	gauge longitude (based on DGA records)	° West		
	area	catchment area	km^2	ASTER GDEM 30-m raster data (Tachikawa et al., 2011)	
	elev_gauge	gauge elevation (catchment outlet) obtained from the 30-m ASTER GDEM elevation data and the location provided by DGA	m a.s.l.		
	elev_mean	catchment mean elevation	m a.s.l.		
	elev_med	catchment median elevation	m a.s.l.		
	elev_max	catchment maximum elevation	m a.s.l.		
	elev_min	catchment minimum elevation	m a.s.l.		
	slope_mean	catchment mean slope	$m\ km^{-1}$		
	nested_inner	number of inner catchments contained within gauge_id catchment (the gauge_id for the inner catchments can be obtained from the hierarchy matrix described in Sect. 3.1.1)	-	-	
	nested_outer	number of catchments containing gauge_id catchment (the gauge_id for the outer catchments can be obtained from the hierarchy matrix described in Sect. 3.1.1)	-		
	location_type	classification in "coastal (or low elevation)", "foothill" and "altiplano" catchments, based on gauge elevations (gauge_elev) below 50 m a.s.l., between 1000 and 1200 m a.s.l., and above 3,500 m a.s.l., respectively.	-	-	Section 3.2
Geological characteristics	geol_class_1st	most common geologic class in the catchment	-	Global Lithological Map database (GLiM) (Hartmann and Moosdorf, 2012)	Table 6 in A17
	geol_class_1st_frac	fraction of the catchment area associated with its most common geologic class	-		
	geol_class_2nd	second most common geologic class in the catchment	-		
	geol_class_2nd_frac	fraction of the catchment area associated with its second most common geologic class	-		
	carb_rocks_frac	fraction of the catchment area characterised as "carbonate sedimentary rocks"	-		
Land cover characteristics	crop_frac	percentage of the catchment covered by croplands, level 1. Includes five types of level 2 classes: rice fields; greenhouse farming; other croplands; orchards; and bare croplands	%	30-m resolution land cover map provided by Zhao *et al.* (2016)	Sections 3.1.3 and 3.2.3
	nf_frac	percentage of the catchment covered by forest (level 1) classified as natural broadleaf (level 2) or natural conifer (level 2).	%		
	fp_frac	percentage of the catchment covered by forest (level 1) classified as broadleaf plantations (level 2) or conifer plantations (level 2).	%		
	grass_frac	percentage of the catchment covered by grasslands, level 1. Includes three types of level 2 classes: pastures; other grasslands; and withered grasslands.	%		
	shrub_frac	percentage of the catchment covered by shrublands, level 1. Includes five types of level 2 classes: shrublands; shrubs and sparse trees mosaic; succulents; shrub plantations; and withered shrublands.	%		
	wet_frac	percentage of the catchment covered by wetlands and water bodies (level 1). Includes six types of level 2 classes: marshlands; mudflats; other wetlands; lakes; reservoirs/ponds; rivers; and ocean.	%		
	imp_frac	percentage of the catchment covered by impervious surfaces (level 1). Urbanised areas are usually contained in this class.	%		
	barren_frac	percentage of the catchment covered by barren lands (level 1). Includes three types of level 2 classes: dry salt flats; sandy areas; and bare exposed rocks			
	snow_frac	percentage of the catchment covered by snow and ice, level 1. Includes two types of level 2 classes: snow and ice.	%		
	fp_nf_index	forest plantation index: calculated as the ratio between fp_frac and the total forested area (fp_frac+nf_frac).	-		

	forest_frac	fraction of the catchment covered by forests, including native forest and forest plantation (fp_frac+nf_frac).	%		
	dom_land_cover	dominant land cover class	-		
	dom_land_cover_frac	fraction of the basin associated with dominant land cover class	%		
	land_cover_missing	percentage of the basin not covered by the land cover map	%		
	glaciers_area	glacierized area within the catchment	km^2	Randolph Glacier Inventory v. 6.0 (RGI Consortium, 2017)	Sections 3.1.4 and 3.2.3
	glaciers_frac	percentage of the catchment covered by glaciers.	%		
Climatic indices (computed for 1 April 1990 to 31 March 2010)	p_mean_i	mean daily precipitation of product *i*	mm day^{-1}	Precipitation, temperature and potential evapotranspiration products introduced in Sect. 3.1.6, 3.1.7 and 3.1.8, respectively.	Table 2 in A17
	p_mean_spread	coefficient of variation of basin-averaged mean annual precipitation from different products (standard deviation of p_mean_i from the four precipitation products, normalised by multi-product mean)	-		
	pet_mean	mean daily PET of pet$_{har}$ product	mm day^{-1}		
	aridity_i	aridity, calculated as the ratio of mean daily PET (pet_mean) to mean daily precipitation (p_mean_i)	-		
	p_seasonality_i	seasonality and timing of precipitation (product *i*) estimated using sine curves to represent the annual temperature and precipitation cycles; positive (negative) values indicate that precipitation peaks in summer (winter); values close to 0 indicate uniform precipitation throughout the year)	-		
	frac_snow_i	fraction of precipitation (product *i*) falling as snow (i.e., on days colder than 0 ◦C)	-		
	high_prec_freq_i	frequency of high precipitation days (≥ 5 times mean daily precipitation) for product *i*	days yr^{-1}		
	high_prec_dur_i	average duration of high precipitation events (number of consecutive days ≥ 5 times mean daily precipitation), for product *i*	days		
	high_prec_timing_i	season during which most high precipitation days (≥ 5 times mean daily precipitation) occur	season		
	low_prec_freq_i	frequency of dry days (< 1 mmday^{-1}), for product *i*	days yr^{-1}		
	low_prec_dur_i	average duration of dry periods (number of consecutive days <1 mmday^{-1}), for product *i*	days		
	low_prec_timing_i	season during which most dry days (< 1 mmday^{-1}) occur, for product *i*	season		
Hydrological signatures (computed for 1 April 1990 to 31 March 2010)	q_mean	mean daily discharge	mm day^{-1}	Streamflow records collected from http://explorador.cr2.cl	Table 3 in A17
	runoff_ratio_i	runoff ratio (ratio of mean daily discharge to mean daily precipitation), for product *i*	-		
	stream_elas_i	streamflow precipitation elasticity (sensitivity of streamflow to changes in precipitation at the annual timescale, using the mean daily discharge as reference and precipitation product *i*)	-		
	slope_fdc	slope of the flow duration curve, FDC (between the log- transformed 33rd and 66th streamflow percentiles)	-		
	baseflow_index	baseflow index (ratio of mean daily baseflow to mean daily discharge, hydrograph separation performed using the Ladson et al., (2013) digital filter)	-		
	hdf_mean	mean half-flow date (date on which the cumulative discharge since 1 April reaches half of the annual discharge)	day of the year		
	Q5	5% flow quantile (low flows)	mm day^{-1}		
	Q95	95% flow quantile (high flows)	mm day^{-1}		
	high_q_freq	frequency of high-flow days (> 9 times the median daily flow)	days yr^{-1}		
	high_q_dur	average duration of high-flow events (number of consecutive days >9 times the median daily flow)	days		
	low_q_freq	frequency of low-flow days (< 0.2 times the mean daily flow)	days yr^{-1}		
	low_q_dur	average duration of low-flow events (number of consecutive days <0.2 times the mean daily flow)	days		
	zero_q_freq	percentage of days with Q=0	%		
	swe_ratio	ratio of peak of snow water equivalent to mean annual discharge	-	SWE product developed by Cortés and Margulis (2017)	
Intervention	sur_rights_n	total number of granted surface rights within the catchment	-		

	sur_rights_flow	annual flow calculated for consumptive permanent continuous surface water rights	$m^3 s^{-1}$	Water Atlas developed by the DGA (DGA, 2016a)	Sections 3.1.10 and 3.2.6
	interv_degree	intervention degree defined as the annual flow of surface water rights (consumptive permanent continuous), normalised by mean annual streamflow.	-		
	gw_rights_n	total number of granted groundwater rights within the catchment	-		
	gw_rights_flow	annual flow calculated for consumptive permanent continuous groundwater water rights	$m^3 s^{-1}$		
	large_dam	0 if there is no dam within the catchment, 1 if there is at least one dam classified as "large"	-	(http://www.ide.cl/descarga/capas/item/embalses-2016.html)	

Table 4: Evaluation metrics of PET gridded products (PET_{har} and PET_{mod}). Pearson correlation coefficients (r) and the ratios between gridded PET and observation-based PET ($INIA_{har}$ and $INIA_{ETO}$) were spatially averaged within the macro-zones.

Macro-zone	PET_{har} compared with $INIA_{har}$		PET_{mod} compared with $INIA_{ET0}$	
	r	ratio	r	ratio
Far North	0.76	0.96	0.19	1.66
Near North	0.92	0.92	0.91	1.68
Central Zone	0.95	1.00	0.96	1.79
Southern Zone	0.97	1.07	0.96	1.58
Austral Zone	0.97	1.08	0.95	1.14
Southern Patagonia	0.98	0.93	0.96	1.06