Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-23
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 February 2018

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

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], Alvaro Ayala[9,10]

[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]Laboratory of Hydraulics, Hydrology and Glaciology (VAW), ETH Zurich, Zurich, Switzerland
[10]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 data set for large sample studies in Chile (South America). The data set includes 516 catchments and provides catchment boundaries, daily streamflow records and basin-averaged time series of the following hydrometeorological variables: 1) daily precipitation retrieved from four gridded sources; 2) daily maximum, minimum and mean temperature; 3) daily potential evapotranspiration (PET); 4) 8-day accumulated PET; and 5) daily snow water equivalent. In addition to the hydro-meteorological time series, we use diverse data sets to extract key landscape attributes characterizing climatic, hydrological, topographic, geological and land cover features. We also describe the degree of anthropic intervention within the catchments by relying on publicly available water rights data for the country. The information is synthetized in 64 catchment attributes describing the landscape and water use characteristics of each catchment. To facilitate the use of the dataset presented here 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 dataset (CAMELS dataset) created for the United States, and proposed several others. Following this nomenclature, we named our dataset CAMELS-CL, which stands for CAMELS dataset in Chile. Based on the constructed dataset, we analysed the main spatial patterns of catchment attributes and the relationships between them. In general, the topographic attributes were explained by the Andes Cordillera; climatic attributes revealed the basic features of Chilean climate; and hydrological signatures revealed the leading patterns of catchment hydrologic responses, resulting from complex, non-linear process interactions across a range of spatiotemporal scales, enhanced by heterogeneities in topography, soils, vegetation, geology and other landscape properties. Further, we analysed human influence in catchment behaviour by relating hydrological signatures with a novel human intervention attribute. Our findings reveal that larger human intervention results in decreased annual flows, runoff ratios, decreased elasticity of runoff with respect to precipitation, and decreased flashiness of runoff, especially in drier catchments.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-23
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 February 2018

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

CAMELS-CL provides unprecedented information in South America, a continent largely underrepresented in large-sample studies. The proximity of the Andes means that this dataset includes high-elevation catchments, which are generally poorly represented world-wide due to data-scarcity. The CAMELS-CL dataset can be used to address a myriad of applications, including catchment classification and regionalization studies, the modelling of water availability 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 data sets freely available for the community.

1 Introduction

Large-sample hydrology has been recognized as a fundamental piece 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 smaller number of catchments. A common approach for the analysis of large numbers of catchments is to explore interrelationships between catchment attributes describing landscape, climate and hydrologic behaviour. These attributes are usually calculated based on 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, regionalization studies use 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 disentangle the interplay between landscape, climate and hydrologic behaviour, and thus providing insights on suitable model structures.

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. The motivation of this paper is to contribute with advancing hydrological science by introducing a large-sample dataset including 516 catchments in Chile. Chile extends over 4,300 km along the west side of South America (17.8°S to 55.0°S), and encompasses 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. Hydro-meteorological data includes 1) daily streamflow (at catchment outlet), and catchment-averaged time series of 2) daily precipitation coming from three different gridded sources (re-analysis and satellite-based); 3) daily maximum, minimum and mean temperature; 4) daily potential evapotranspiration (PET) based on the Hargreaves formula (Hargreaves and Samani, 1985); 5) 8-day accumulated PET based on MODIS imagery; and 6) daily snow water equivalent based on a high-resolution reanalysis (Cortés and Margulis, 2017).

To facilitate and encourage the use of the dataset presented here, and promote common standards and formats in large-sample studies, we computed 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 uses the meteorological and streamflow data dataset collated by Newman et al., (2015) and provides quantitative estimates of a wide range of attributes for 671 catchments in the contiguous United States. The CAMELS dataset has already been used for 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 Conterminous United States (CONUS), among others. Following this nomenclature, we named our dataset CAMELS-CL, which stands for CAMELS dataset in Chile. Importantly, we added an attribute class not covered by A17: the degree of human intervention in each catchment. This novel information is valuable since anthropogenic activities can have major impacts on catchment behaviour, but human influence is often difficult to quantify, especially for hundreds of catchments.

We used CAMELS-CL dataset to advance the understanding of hydrological systems by analysing the spatial distribution of catchments attributes. In particular, we addressed the following research question: What are the dominant spatial patterns of physical, climatic and hydrological catchment attributes in Chile? Furthermore, under the hypothesis that anthropic intervention affects catchments hydrological response, we addressed the following research question: Can the human influence on catchment behaviour be detected and explored using the information on surface water rights provided by CAMELS-CL? We used hydrological signatures to describe catchment behaviour and information on surface water rights to characterize the level of human intervention. To the best of our knowledge, the effects of human intervention within a catchment on its hydrological signatures has not been investigated.

The CAMELS-CL dataset is freely available at the Center for Climate and Resilience Research website (www.cr2.cl), providing unprecedented information in the region. The research questions that can be addressed by using CAMELS-CL may vary from very general ones (e.g., what are the main controls on runoff generation?) to very specific ones (e.g., what are the impacts of forest cover types in water availability? what are the impacts of water use allocation in summer flows?).

This paper is structured as follows: Sect. 2 describes the study area; Sect. 3 provides descriptions of the datasets collected and the variables computed at the catchment scale; Sect. 4 describes the catchment attributes obtained from these datasets and

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-23
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 February 2018

analyses their spatial distribution; Sect. 5 presents the analysis on the effects of human activities on catchment behaviour; Sect. 6 summarizes 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 (Figure 1). The country lies on the Nazca and Antarctic plaques. 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 (DGA, 1986). The Andes range defines the east border with Bolivia and Argentina for most of the country. Featuring altitudes well above 3000 m a.s.l. in most of country with summits of up to 7,000 m a.s.l. (e.g. mountain Aconcagua or volcano Ojos del Salado), the Andes acts as an effective barrier for the atmospheric flows, leading to particularly high precipitation amounts and defining most of the regional hydroclimatic conditions (Garreaud, 2009).

Chile has 16 administrative regions (Fig. 2) that have been grouped in four macro-zones by the Chilean Water Directorate (DGA), based on their main 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 within this manuscript, 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. 2.

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) in 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 climate (Cfc) and cold steppe (BSk).

3 Datasets

3.1 Catchment boundaries

The first step in the CAMELS-CL database development was the delimitation of catchment boundaries and their storage as shapefile polygons (Fig. 3). It should be mentioned that there is an official database for Chilean hydrographic network that 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 101 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 DGA hydrographic network is that –given the methodology used for its implementation– the existing streamflow gauges do not necessarily correspond with catchment, sub-catchment or sub-sub-catchment outlets. Furthermore, DGA catchment boundaries are truncated at the administrative national border, even when –for some catchments– there are areas contributing with runoff located in Bolivian and Argentinian territory. Given that any hydrologic application within a controlled basin (i.e., with streamflow records) requires the total area contributing with the measured streamflow, different studies in Chile have had to delineate their own (most probably different) boundaries, which is a considerable source of errors and makes result comparison difficult.

To overcome this limitation, in CAMELS-CL we created our own catchment boundaries database, considering that the basin outlets were located at the position of the streamflow gauges selected in Sect. 3.2, and following only topographic-driven limits (not the administrative national border). A key challenge for this task –which represents an important source of uncertainty– is that the station geographic coordinates reported by the DGA did not coincide exactly with the location of the river, according to Google Earth imagery. For those cases, ancillary information was used to determine the most probable gauge location, such as the name of the gauge (usually containing information about the name of the river, nearby roads and/or bridges), road maps, and Google Earth imagery (Google, 2016).

After defining the catchment outlet, 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 – i.e., many headwater catchments drain into rivers flowing in larger catchments that are also part of this dataset. To account for this, we generated a catchment hierarchy that characterises the nested configuration within the dataset. This hierarchy was built as a logical matrix indicating the basins that are contained within each 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 regionalization.

3.2 Topography

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 polygon obtained in Sect. 3.1, 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 Fig. 7c).

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-23
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 February 2018

3.3 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 resoution of 1:1,000,000 and is the most complete and commonly used map for the country. For each catchment, we reported the most frequent 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 useful indicator of the presence of karstic systems.

3.4 Land cover

We used the 30-m resolution land cover dataset 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). The classes can be 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. 2): croplands; forests, grasslands; shrublands; wetlands; water bodies; impervious surfaces; barren lands; and snow and ice. Some of the classes are 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). 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).

3.5 Glaciers

Glaciers can play a relevant role in catchment hydrology by providing significant annual contributions to runoff, and shaping hydrological seasonality. In Chile, glaciers can be found over almost its entire territory and vary 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. Even though the land cover dataset of Zhao et al. (2016) identify areas of snow and ice, we decided to include a glacier inventory for calculating the degree of glacierization of the selected catchments. Glacier inventories have the advantages of using geomorphologic glacier-delineation techniques, and the recognition of debris-covered areas, which 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. 3). In CAMELS-CL, the RGI

6.0 was clipped within each catchment and two attributes were computed: the total glacierized area (km^2) and the percentage of glacierized area in the catchment (%) (further details in Sect. 4).

3.6 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. Finally, we discarded gauges located in artificial channels, ending up with 516 selected gauges for the CAMELS-CL database. Figure 4a presents the mean annual discharge for each station (computed for its complete period of record). Figure 5 illustrates the availability of daily streamflow records for different time periods (represented with different colours). Note that water years go from April 1st to March 31st. As expected, the number of stations decreases as we filter them with larger 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 period. When considering all stations with at most 30% of missing data, then 249 to 258 stations (48% and 50% of the total number of catchments, respectively) would meet this criterion (Fig. 5).

3.7 Precipitation

Precipitation is a key driver of the water cycle, and even in densely monitored regions, precipitation data is highly uncertain (Tian and Peters-Lidard, 2010; Woldemeskel et al., 2013). This limitation is aggravated in less-developed regions and areas difficult to access, where only a sparse network of meteorological stations is available. To have an indicator of the uncertainty of this key forcing data, 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 $precip_{cr2met}$, $precip_{tmpa}$, $precip_{chirps}$ and $precip_{mswep}$. The spread of these time series for a single catchment should be an indicator of the uncertainty of the variable.

The $precip_{cr2met}$ times series is derived from CR2MET, a spatially-distributed daily precipitation product for Chile, which is currently being used in the updated national water balance carried out by DGA (project that should finalise on 2019). This product is partly based on a statistically downscaled ERA-Interim reanalysis data (Balsamo et al., 2015). The method builds on multiple linear regression models used 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, are calibrated with are large network of quality-controlled rain-gauge records. Depending on the distance of a given grid-cell to neighbouring stations, the final product is obtained from merging downscaled precipitation and spatially interpolated in-situ observations. Further information about formulation, quality control and evaluation of the product can be found in the DGA water balance report.

The three satellite-based precipitation products included in the current version of CAMELS-CL were selected based on the exhaustive comparison and evaluation reported by Zambrano-Bigiarini et al., (2017) for the entire Chilean territory. The Tropical Rainfall Measuring Mission (TRMM; Huffman et al., 2007) Multi-satellite Precipitation Analysis (TMPA) is a joint mission of NASA and JAXA to provide "best" estimates of quasi-global precipitation, using infra-red and passive microwave data from a wide variety of satellite-borne precipitation-related sensors at relatively high spatial resolution (0.25°). TMPA products include a real-time (TRMM 3B42RT) and research (TRMM 3B42 version 7) products, with latency times of about seven hours from the observation time and two months, respectively. In this study, the research product TRMM 3B42v7 was used because it makes use of Global Precipitation Climatology Project (GPCP; Adler et al., 2003) and Climate Assessment and Monitoring System (CAMS) data to rescale its estimates on a monthly basis. The instruments on TRMM were turned off on 2015, and this product will be superseded by the Global Precipitation Measurement (GPM) mission product IMERG, launched on February 27th 2014. To overcome the current short data length of the new IMERG dataset, TMPA data will continue to be computed until IMERG is considered operational (expected in spring 2018, Huffman et al., 2015).

The Climate Hazards Group InfraRed Precipitation with Station data version 2 (CHIRPSv2; Funk et al., 2015) is a new long term (1981 to near-present), quasi-global (50°N46°S) daily, pentadal, and monthly SRE with a spatial resolution of 0.05°. CHIRPSv2 uses the Tropical Rainfall Measuring Mission Multi-Satellite Precipitation Analysis version 7 (TRMM 3B42v7) in order to calibrate global Cold Cloud Duration (CCD) rainfall estimates (Funk et al., 2015). CHIRPSv2 also uses observed rain gauge data in order to reduce biases of its estimates, based on public and private data archives at monthly scale. CHIRPSv2 was designed for monitoring agricultural drought and global environmental changes over land. Originally, this dataset spanned from 50ºN to 50ºS, but since November 2012 onwards data were not generated south of 46ºS. More information can be found in Funk et al. (2015).

The Multi-Source Weighted-Ensemble Precipitation (MSWEP, Beck et al., 2017) version 1.1, is a new global precipitation dataset released in June 2016, with a 3- hourly temporal and 0.25º spatial resolution, specifically designed for hydrological modelling. MSWEP was designed to improve the performance of satellite products in representing precipitation in mountainous, tropical, and snowmelt-driven regions. It is based on observed rain gauge data, satellite observations and reanalysis data to provide reliable precipitation estimates over the entire globe. A detailed description of this dataset can be found in Beck et al., (2017). For this work, only daily data from version 1.1 are used, but newer versions of the product (v2.0 is already available while v2.1 is upcoming) will be included in CAMELS-CL after validation with ground measurements in Chile.

3.8 Temperature

Daily time series of minimum (T_{min}), maximum (T_{max}) and mean (T_{mean}) temperature for each catchment were also derived from CR2MET dataset, which is currently being used in the updated national water balance carried out by DGA. The daily minimum and maximum temperatures in CR2MET (CR2MET/T_{max} and CR2MET/T_{min}, respectively) were mapped for the period 1979-2016 with a different approach than that followed for precipitation. In this case, the method is nurtured by land-

surface temperature (LST) estimates from Moderate Resolution Imaging Spectro-radiometer (MODIS) satellite retrievals, in addition to near surface temperature provide by ERA-Interim. Multiple regression models for both CR2MET/T_{max} and CR2MET/T_{min} are computed to fit local observations using LST as explanatory data. Given the data gaps and relatively short period available for LST, the final product is extrapolated back in time using ERA-Interim information. To get daily mean temperatures (CR2MET/T_{mean}), the long-term CR2MET/T_{max} and CR2MET/T_{min} products are used to adjust the 3-hourly near surface temperature of ERA-Interim. The adjusted 3-hourly data is finally averaged to derive CR2MET/T_{mean} (Boisier et al., *in prep*). Gridded daily mean, minimum and maximum temperatures products (0.05° lat-lon resolution) were then clipped and averaged within the catchment boundaries, resulting in a daily time series for each catchment, named T_{mean}, T_{min} and T_{max}, respectively.

3.9 Potential evapotranspiration

We processed catchment-scale potential evapotranspiration (PET) from two different sources. Having two different estimates of the same variable may be used to quantify the uncertainty of PET within the catchments. 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 T_{min} and T_{max} (described in Sect. 3.8) to generate a gridded PET estimate, which was then clipped and averaged within the catchment boundaries to generated a daily averaged PET time series– called pet_{har} hereinafter.

The second PET product is based on the MODIS instrument operating on both the Terra and Aqua spacecrafts. We used the PET product MOD16 collection 5 (Mu et al., 2005), which is processed from different sources of information, including the collection 5 FPAR/LAI (MOD15A2) (Myneni et al., 2002), the collection 4 land cover type 2 (MOD12Q1) (Friedl et al., 2002), the albedo collection 5 MCD43B2 and MCD43B3 (Jin et al., 2003; Lucht et al., 2000), and the daily meteorological reanalysis data from NASA's MERRA GMAO (GEOS-5). MOD16 is calculated based on 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. We clipped and averaged MOD16 pixels within the catchment boundaries to obtain an 8-day time series for each catchment – called hereinafter pet_{modis}.

The two PET products generated for CAMELS-CL catchments have different advantages and disadvantages. In particular, pet_1 has a lower spatial resolution (5x5 km^2) compared to pet_2 (1×1 km^2), but covers a longer period (1979-2015, the same as T_{min} and T_{max}). Furthermore, pet_2 has a more complex formulation based on surface energy balance including parameters such as albedo and FPAR/LAI, which in theory should bring a more robust estimation than pet_1, which is calculated from an empirical approach based on air temperature. In order to assess the quality of gridded pet_{har} and pet_{modis} products, we compared them (before calculating the average within the catchment boundaries) against an independent set of PET point values calculated from meteorological stations of the Chilean National Institute of Agricultural Research (INIA). With the meteorological information, we computed PET for each meteorological station based on the Hargreaves approach (Hargreaves-INIA, which

is used to evaluate pet_{har}) and based on FAO Penman-Monteith reference crop evapotranspiration (ET_0) (ET_0-INIA, which is used to evaluate pet_{modis}).

The comparison between pet_{har} and Hargreaves-INIA provided Pearson correlation coefficients greater than 0.75 throughout the national territory. Pearson correlation coefficients spatially-averaged within the macro-zones were 0.76, 0.99, 0.93, 0.83 and 0.98 for the Far and Near North, the Central Zone, the Southern Zone, the Austral Zone and Southern Patagonia, respectively. The ratio between mean pet_{har} and mean Hargreaves-INIA ranged from 0.92 to 1 for the different macro-zones, indicating high agreement. The ratio between standard deviations ranged between 0.97 and 1.03 per macro-zone. These results are in agreement with the coincident formula used to computed pet_{har} and Hargreaves-INIA.

The comparison between pet_{modis} and the ET_0-INIA led to a Pearson correlation coefficient greater than 0.80 except for the Far North, where the correlation was below 0.25 in the three available stations. Regarding the ratio between the means of pet_{modis} and the ET_0-INIA, the comparison indicates that MOD16 (pet_{modis}) systematically overestimates weather stations estimates. The results were, on average, 1.66 times in the Far North, 1.68 in the Near North, 1.79 in Central Zone, 1.58 in Southern Zone, 1.14 in Austral Zone and 1.06 in Patagonia Southern. The ratios between the standard deviation of the product data and that obtained from meteorological stations were 0.59, 1.03, 1.25, 1.57, 1.23 and 1.19, respectively. The biases found here may be explained by the theoretical differences between ET_0 and the PET calculated in MOD16. ET_0 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_{modis} product that includes a more complete parametrization of those variables according to vegetation characteristics.

Since the INIA records were used differently for evaluating pet_{har} and pet_{modis}, a direct comparison between the two evaluations is not possible, although they provide valuable information about the quality of (the same) PET product across the territory. A more detailed discussion on the evaluation of MOD16 products by using ground meteorological data is provided in Zambrano-Bigiarini et al. (*in prep*).

3.10 Snow water equivalent

Snow water equivalent (SWE) data was processed from the daily SWE product generated by Cortés and Margulis (2017), which has a 180x180 m^2 spatial resolution. For SWE estimations, Cortés and Margulis (2017) used SWE ensemble estimates obtained from forward modelling "prior" values, which were then conditioned via data 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 of the 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 model used to generate the prior ensembles estimates is the SSiB3 LSM (Yang et al., 1997) combined with a Snow Depletion Curve model (SDC; Liston, 2004). Verification of the reanalysis framework has been performed for the Sierra Nevada using in-situ sensor data (Margulis et al., 2016) and for the Andes (Cortés et al., 2016) using more than 2000 snow survey points obtained from 2009 to 2015 and 350 site-years of peak annual snow pillow and snow course SWE data from 1985 to 2015. 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 as well 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.

3.11 Water rights

To account for the degree of human intervention within the catchments, we compiled and processed granted water rights available from the Water Atlas developed by the DGA (DGA, 2016a). This dataset includes information about the water source (surface or groundwater), the type of right (consumptive or non-consumptive), its use (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 (permanent and continuous, permanent and discontinuous, permanent and alternated, eventual and continuous, eventual and discontinuous, or eventual and alternated). A detailed explanation of these water right types classification can be found in Carey (2014). A key limitation of this dataset is that it provides information about granted rights and not about the actual use of those rights (Larraín, 2006). Additionally, some water right records have incomplete information (e.g., missing coordinates, water volume assigned and temporal allocation).

To illustrate water allocations within the territory, Fig. 6 displays the surface and groundwater rights (all types) over a large portion of the country (the water rights dataset was processed for the complete territory though). It is clear that groundwater rights dominate in the Near North and Central Zone (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 the Southern Zone, especially within high elevation areas towards the Andes.

4 Catchment attributes

A total of 64 catchment attributes were computed and grouped in six attribute classes (Table 2). To motivate the use of common standards in the development of large sample catchment datasets, we computed 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 climatic indices and hydrological signatures were fully adopted from A17. Soils characteristics were not computed at this stage since there is no national publicly available dataset. Nonetheless, global maps (such as the Soilgrids250m dataset (Hengl et al., 2017) will be included in the CAMELS-CL dataset in the near future (see concluding remarks). Given the differences in input datasets, some of the location and topography, geologic characteristics, and land cover characteristics attributes were not computed here. Several new attributes were computed for the classes location and topography (Sect. 4.1), land cover characteristics (Sect. 4.3) and hydrological signatures (Sect. 4.5). Finally, we added a new class to describe the degree of intervention within the catchments. These intervention attributes are calculated based on water use data (Sect. 3.11) and described in Table 3 and Sect. 4.6.

The description of the 64 computed attributes and the corresponding data source used to compute them is presented in Table 3. To ensure the reproducibility of our results, the reference for the explicit formulation of climatic indices and hydrological signatures is provided in Table 3. In the following sections, we discuss the spatial distribution of the catchment attributes presented in Table 3, separated by class: location and topography (Sect. 4.1); geology (Sect. 4.2); land cover (Sect. 4.3); climate indices (Sect. 4.4); hydrological signatures (Sect. 4.5); and water use (Sect. 4.6).

4.1 Location and topography

For visualisation purposes, six (out of 14) catchment attributes within the class location and topography are displayed in Fig. 7. Figure 7a 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. 7f) with three categories: coastal (or low elevation), foothills and altiplano catchments, defined by gauge elevations lower than 50 m a.s.l., between 900 and 1000 m a.s.l., and above 3,500 m a.s.l., respectively. Different categories may be defined, depending on application requirements. In facts, as it was defined, the location_type attribute does not properly classify the catchments in Southern Patagonia, where the catchments feature low elevations at the outlets, but do not necessarily drain to the sea (i.e., they are not coastal).

Figure 7b displays basin-averaged mean elevations, revealing smoother N-S and E-W gradients compared with gauge elevation gradients (Fig. 7a). This is because the mean elevation calculated for downstream catchments includes nested catchments (located at higher altitudes). The spatial distribution of mean catchment slope follows different patterns depending on the macro-zone (Fig. 7c). The Far North – dominated by the flat Altiplano Plateau – exhibits relatively small variations in mean slopes (with relatively low mean slope 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. 7d), which is consistent with smaller head-water 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 north-to-south drainage direction, or a portion of their total contributing area in Argentina (beyond the east national border). Other 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), the number of nested basins within each catchment (Fig. 7e) is not necessarily correlated with the entire contributing area (Fig. 7d). For example, some small catchments might be highly instrumented (i.e., with many stream gauges, because of –for example– water allocation priorities), and thus having a large number of nested basins, while large but poorly instrumented catchments might not have inner basins defined.

4.2 Geology

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 large presence of Pyroclastics, Siliclastic sedimentary rocks and Intermediate volcanic rocks (Figures 8a and 8b), which can result in the connection of groundwater systems through fractured volcanic rocks (DGA, 1986). This means that there could be a difference between the surface catchment boundary (based on topography) and the groundwater system, which should be taken into consideration when analyzing the hydrologic response of these catchments. Figure 8a also indicates that there can be strong geological differences between neighbour catchments. Furthermore, there tends to be a high geological variability within the catchments. For the large majority of them, the dominant geological class covers less than half of the catchment, as indicated by the histogram of Figure 8b. The effect of these diverse and quickly-varying geological conditions on hydrological behaviour will be the focus of future research based on the CAMELS-CL data sets. The occurrence of carbonate sedimentary rock is particularly low (Figure 8e), with only 24 catchments with at least 10% of carbonate sedimentary 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).

4.3 Land cover

As summarised in Table 2, five land cover attributes in A17 were not computed since the map developed by 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). Because of the topographic configuration of CAMELS-CL region (with very high elevations and latitudes) and the importance of glaciers in the water balance, we added two attributes based on the area of each catchment covered by glaciers, calculated from the glaciers inventory described in Sect. 3.5. We found that 255 catchments (48% of the total) have some degree of glacierization, reaching up to 62% in the Geike River catchment in the Southern Patagonia. Selected glaciers cover a total area of 7,321 km^2, corresponding to almost a quarter of the glacierized area in the Southern Andes (RGI Consortium, 2017). The catchments with the largest degree of glacierization

(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.

A limitation of the land cover map developed by Zhao *et al.*, (2016) is that it spans exclusively over the national territory, although several catchments (almost 50) have a portion of their areas in Argentina (eastern boundary). To account for this limitation, we generated an attribute indicating the percentage of the catchment contained within the land cover map. Therefore, this index also serves as a quality flag for basin-averaged land cover characteristics.

Figure 9 illustrates a sub-set of the land cover attributes listed in Table 3. Fig. 9a shows the forested (native forests and forest plantation types) area within the catchments, which is concentrated in the Southern Zone, Austral Zone and Southern Patagonia. In forested catchments, exotic forest plantations dominates the coastal areas of the Central and Southern Zones, with forest plantation indices up to one (Fig. 9b). This distribution is due 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 and forest plantations lands (Armesto et al., 2010; Miranda et al., 2015). This conversion has had dramatic impacts in forest ecosystem services such as water provision (Jones et al., 2017; Lara et al., 2009).

Figures 9c and 9d show the Far and Near North Zones with more homogeneous land cover type, where shrublands and impervious lands span in more than 60% of the catchments. Towards the southern areas, the coverage of the dominat classes decreases substantially, transitioning towards a mosaic of different land cover types. The missing land cover data is presented in Fig. 9e, which should be accounted for if the land cover attributes of the affected catchments (the ones with portions in Argentina, as shown in Fig. 2) are used to characterise them in applications such as catchment classification or parameter regionalisation.

4.4 Climatic indices

To allow a direct comparison between CAMELS (A17) and CAMELS-CL, the climatic indices listed in Table 3 were computed for the same period of record used in A17, i.e., water years 1990 to 2009, corresponding to 1 April 1990 to 31 March 2010 for Chile. If these indices are required for different record periods, the formulae provided in the references from Table 3 can be used with the raw hydro-meteorological time series (available at CAMELS-CL website). The complete spatial and temporal coverage of the meteorological variables allow computing the climatic indices for all 516 catchments (in contrast to hydrological signatures that were computed for a sub-set of catchments, as explained in Sect. 4.5). Precipitation and PET-based attributes were calculated for the four precipitation products (Sect. 3.7) and the daily PET product (Sect. 3.9), respectively.

The climatic attributes presented in Figs. 4 and 10 reveal basic features of the Chilean climate, 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 3000 mm over western Patagonia (Fig. 4b). Such marked precipitation gradient obeys to the varying influence of the semi-permanent anticyclone over the subtropical Southeast Pacific and the westerly wind belt at mid-latitudes. The frequency of high precipitation events also increases southward, with a maximum in south-central Chile

(Fig. 4d). The high terrain of the Andes in the Far North is influenced by the monsoonal regime developing over the interior of the continent and receives about 300 mm/year over 4000 m a.s.l. Superimposed on its north-south 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 the lowlands and the windward slopes; Viale and Garreaud, 2014). Potential evapotranspiration has a more restricted range (800-1000 mm/year; Fig. 4d), so the aridity index (PET/P, Fig. 10c) tends to be less (more) than one to the north (south) of Santiago (33°S). The positive values of precipitation seasonality (Fig. 10a) in the northern part of the country indicate precipitation peaks in summer (djf), which is consistent with high precipitation days the Far North during this season (Fig. 10f). Negative seasonality values indicate a winter maximum (jja) for most of the country (Fig. 10a), which is consistent with high precipitation days concentrate in winter (Fig. 10f) for all the macro-zones, except the Far North and Southern Patagonia. The seasonality values close to zero in Southern Patagonia (Fig. 10a) indicate uniform precipitation throughout the year. The zero-temperature isotherm during winter storms ranges between 1500 and 4000 m a.s.l., so that most of the precipitation along the coast and interior valleys falls as rain, with snow prevailing in high-elevation basins.

4.5 Hydrological signatures

In concordance with climatic indices, the hydrological signatures listed in Table 3 were computed for the period 1 April 1990 to 31 March 2010 for Chile. To compute these attributes, we selected a sub-set of 222 catchments with valid daily streamflow records in at least 85% of the period considered. It should be noted that, despite of the calculation of hydrological indices for a subset of catchments, raw daily time series for all 516 catchments are included in the publicly available CAMELS-CL database. These time series and the formulae provided in the references from Table 3may be used if the signatures are required for different time periods.

Figure 11 illustrates the spatial distribution of 12 (out of 14) hydrological signatures summarised in Table 3, revealing the leading patterns of catchment hydrologic responses. To exclude anthropic intervention effects that might be influencing hydrologic responses, we plotted the attributes for 135 catchments that feature a low anthropic intervention degree. This selection was made based on interv_degree (Table 3) values lower than 3%, however, based on the analysis presented in Sect. 5, a different threshold value could be applied.

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. 10c). 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. 7b) snow-dominated (Fig. 10b) basins in Central Chile – where the most frequent season for low precipitation days is Dec-Feb (Fig. 10i).

The slope of the mid-segment of the FDC (Fig. 11d) – a signature to quantify 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. 11e) and low discharge precipitation elasticity (Fig. 11f). 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. 11g-i), similar spatial distributions of low flow frequencies (Fig. 11j) and mean low flow durations (Fig. 11k) are obtained across the country. Q95 (Fig. 11i) and Q5 (Fig. 11l) provide generally similar patterns to those of mean daily discharge (Fig. 11a), 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. 11g-l and climatic indices in Fig. 10d-i highlights 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 disconnection is observed between the duration of low precipitation (Fig. 10h) and low flow (Fig. 11k) 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. 11) – in contrast to generally smooth patterns in climate indices (Fig. 10) – 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. As discussed in A17, careful attention should be paid to such interactions, as well as uncertainties involved in the calculation of hydrological signatures (Westerberg et al., 2016; Westerberg and McMillan, 2015).

4.6 Intervention

Figure 12 summarises water rights records used to assess the intervention degree within the catchments. We can see that the number of surface rights (Fig. 12a) increases from north to south, while the number of groundwater rights (Fig. 12d) increases from east to west. These values do not provide information about allocated volumes, but they show how many water rights holders need to 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 needed.

In terms of allocated surface and groundwater flows (Fig. 12b and Fig. 12e, respectively), we only considered the consumptive permanent continuous water rights for calculation. From these specific water right types, we used only those expressing annual flows in units of volume per time. Water rights expressed as “shares” are not provided with their corresponding conversion into volume units (DGA, 2016b). “Shares” rights correspond to a 6% of the national water rights database, and are the oldest (allocated prior to the 2005 water code reform), thus probably representing a majority of the rights within the Central Zone.

The above limitations may lead to an underestimation of the allocated flow, due to (at least) the following reasons: non-consumptive rights may have their restitution points outside the catchment boundaries (however, they were not considered allocated flow calculation); shares rights are disregarded; there are missing pieces of information thus some rights may be omitted (Sect. 3.11). 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 the unauthorised use of surface and groundwater extractions.

Despite the limitations of this dataset and the attributes presented in Fig. 12, water rights information is still critical to quantify human intervention within a catchment, and it has not been officially processed at the catchment scale in Chile. To quantify the intervention degree within a catchment, we propose to use the attribute interv_degree (described in Table 3 and illustrated in Fig. 12c), calculated as the ratio between the annual surface flow allocated within a catchment, and the catchment 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. To select a sub-set of catchments based on their degree of intervention, a threshold value could be applied to interv_degree attribute (low threshold values would indicate a low intervention degree).

5 Impacts of human activities on catchment behaviour

This analysis uses hydrological signatures to describe catchment behaviour and the interv_degree attribute (Sect. 4.6) to characterize the level of human intervention. Figure 13 (top panel) displays scatter plots between four hydrological signatures and the logarithm of the intervention degree index –which accounts for consumptive, continuous surface water rights. We used colors to indicate the aridity of each catchment, as aridity is a major driver of hydrological behaviour. The Spearman correlation coefficient and p-values (at 95% confidence interval) are also presented.

Larger human intervention results in decreased annual flows and runoff ratios, especially in drier catchments. Interestingly, a larger number of consumptive surface rights is reflected on decreased elasticity of runoff with respect to precipitation, and decreased flashiness of runoff, supported by low p-values. Note that these scatter plots do not allow for the separation of the effects of aridity and human intervention. To address this, we binned the data and used boxplots to disentangle the two effects. The middle panel in Fig. 13 shows the boxplots of these four hydrological signatures for the catchments classified, binned depending on their aridity (humid, medium and dry) and their degree of human intervention (low and high intervention). 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 degrees. In contrast, the dry catchments (zoomed view in bottom row of Fig. 13) reveal significant differences in hydrological signatures for high and low human intervention degree. Consistently with the scatter plots, the annual flows show a substantial decrease in catchment with larger number of consumptive surface rights, which is expected given the withdrawal of water within water-scarce regions. Another important behaviour found here is that drier catchments become less sensitive to precipitation when human intervention is high (represented by significantly lower streamflow elasticity).

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-23
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 February 2018

6 Concluding remarks

We introduced the CAMELS-CL dataset, which provides novel information in a region that is largely underrepresented in large-sample studies. CAMELS-CL includes daily streamflow and a suite of hydrometeorological variables (precipitation, temperature, potential evapotranspiration, and snow water equivalent) for 516 catchments in Chile. The dataset also includes shapefiles polygons of the catchment boundaries, generated based on the location of the streamflow gauges. This catchment boundaries dataset overcomes the main limitations of the official national hydrographic network (DGA; CIREN, 2014), where streamflow gauges do not necessarily correspond with catchment outlets and the catchment boundaries are truncated at the administrative national border. Additionally, we synthesized diverse and complementary data sets to quantify key geophysical attributes that shape catchment behaviour. We computed a total of 64 catchment attributes describing topography, soils, geology, land cover, climate, hydrology, and intervention degree for each catchment.

The main spatial patterns of catchment attributes and the relationships between them were analysed across the 4,300 km covered by the dataset, which includes high altitude catchments, and five different primary climatic regimes. We identified the elevation and slope gradients from the 516 catchments, which are mainly dominated by the Andes position along the country. A high geological variability was observed between neighbour catchments and within the catchments. Glacier-based attributes indicate that 255 catchments (48% of the total) have some degree of glacierization, with the most glacierized catchments (more than 15% of their areas covered by glaciers) located in the Austral and Southern Patagonia macro-zones, followed by the Olivares and Volcan river catchments (about 14%) in the Central Zone. The land cover attributes show the main land cover types within the catchments, indicating larger heterogeneity towards the southern regions. The climatic attributes reveal basic features of the Chilean climate, with a marked precipitation N-S gradient that obeys to the varying influence of the semi-permanent anticyclone over the subtropical Southeast Pacific and the westerly wind belt at mid-latitudes, superimposed with a E-W gradient due to orographic enhancement over the windward slope of the Andes cordillera. Hydrological signatures reveal the leading patterns of catchment hydrologic responses, with strong correlations between runoff (daily flows and runoff ratios) and mean annual precipitation and aridity index. The attributes observed in the northern regions (Far North and Near North) reveal a substantial subsurface and groundwater contributions to total runoff. In general, we observe sharper variations in hydrological signatures compared to patterns in climate indices. This is due to complex, non-linear process interactions across a range of spatiotemporal scales, enhanced by heterogeneities in topography, soils, vegetation, geology and other landscape properties.

The analysis on the impacts of human activities on catchment behaviour leads to several important insights. We showed that larger human intervention results in statistically significant decreased annual flows, runoff ratios, decreased elasticity of runoff with respect to precipitation, and decreased 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 this type of anthropic index to predict shifts in hydrological systems.

We discussed some of the main limitations of each dataset and its corresponding attributes, which should be considered when using CAMELS-CL for selecting catchments and interpreting results. The first source of uncertainty introduced in CAMELS-CL is stream gauge location, which affects catchment boundary delineation. Catchment polygons are used to process all basin-scale meteorological and physical variables, so uncertainties in basin delineation are directly propagated to the rest of the database. Hence, any errors detected from incoherent hydrological analysis results or new ancillary information about stream gauge location will be flagged and corrected in future version of the CAMELS-CL data set.

Another source of uncertainty in hydrological applications is the estimation of precipitation, which is particularly challenging in scarce data and mountainous regions. To enable users to evaluate the reliability of precipitation estimates within a catchment, we processed four different products that can be used to explore their uncertainties. A precipitation product inter-comparison is beyond the scope of this paper, but we see it as potential research avenue to be pursued using CAMELS-CL. Similarly, potential evapotranspiration was estimated using two different products: (i) a temperature-based PET product whose daily resolution meets the requirements of many applications, and (ii) an 8-day accumulated PET product that has better spatial resolution and a more process-based formulation accounting for surface characteristics.

An essential contribution of this study is that we characterize the degree of human intervention in each catchment. Anthropogenic activities can have a major influence on the storage and transport of water across the landscape, hence they have to be accounted for when exploring catchment behaviour. The intervention attributes we derived based on Water Atlas developed by the Chilean Water Directorate enable a first classification of the basins based on their degree of intervention. These attributes however should be interpreted while keeping their limitations in mind (Sect. 4.6). We expect to refine those attributes based on new sources of information, including dam location and dimensions, inventory of national water demands (project currently being developed by DGA), and technical knowledge from DGA advisors. Along this line, 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, further –new and more detailed– information about the Chilean cryosphere should be included. For example, complementing the global inventory processed here with national inventories of Chile and Argentina. Furthermore, studies focusing on hydrological projections of glacierized catchments require not only the glacier surface area but also an estimate of the total amount of water stored in the glaciers. Therefore, a database of ground-penetrating radar measurements of Chilean glaciers and results from the further development of approaches for estimating ice thickness remotely (Farinotti et al., 2017) will be included in future versions of CAMELS-CL.

The contributions of this paper are to advance the understanding of hydrological systems by the analysis on the spatial distribution of catchments attributes; to demonstrates that catchment behaviour changes with human intervention; and to advance hydrological science by providing a dataset that can be used to address a wide range of research questions. The research questions that can addressed with CAMELS-CL may be related with –but not limited to– catchment classification, similarity and regionalization, 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. The time series of streamflow, meteorological variables, and the catchment attributes that constitute CAMELS-CL are available from the Center for Climate and Resilience Research website (http://www.cr2.cl/recursos-y-publicaciones/bases-de-datos/datos-informacion-integrada-por-cuencas/).

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.

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.

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.

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.

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

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

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.

Farinotti, D., Brinkerhoff, D. J., Clarke, G. K. C., Fürst, J. J., Frey, H., Gantayat, P., Gillet-Chaulet, F., Girard, C., Huss, M., Leclercq, P. W., Linsbauer, A., Machguth, H., Martin, C., Maussion, F., Morlighem, M., Mosbeux, C., Pandit, A., Portmann, A., Rabatel, A., Ramsankaran, R., Reerink, T. J., Sanchez, O., Stentoft, P. A., Singh Kumari, S., Van Pelt, W. J. J., Anderson, B., Benham, T., Binder, D., Dowdeswell, J. A., Fischer, A., Helfricht, K., Kutuzov, S., Lavrentiev, I., McNabb, R., Hilmar Gudmundsson, G., Li, H. and Andreassen, L. M.: How accurate are estimates of glacier ice thickness? Results from ITMIX, the Ice Thickness Models Intercomparison eXperiment, Cryosphere, 11(2), 949–970, doi:10.5194/tc-11-949-2017, 2017.

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.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S. and Kempen, B.: SoilGrids250m: Global gridded soil information based on machine learning, PLoS One, 12(2), e0169748, doi:10.1371/journal.pone.0169748, 2017.

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.

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.

Huffman, G. J., Bolvin, D. T. and Nelkin, E. J.: Integrated Multi-satellitE Retrievals for GPM (IMERG) Technical Documentation, NASA/GSFC Code, 612, 47, doi:10.1136/openhrt-2016-000469, 2015.

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.

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.

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.

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.

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

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

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.

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

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.

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.

USGS: Global 30 Arc-Second Elevation (GTOPO30)., 1996.

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.

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.

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.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-23
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 February 2018

Figures

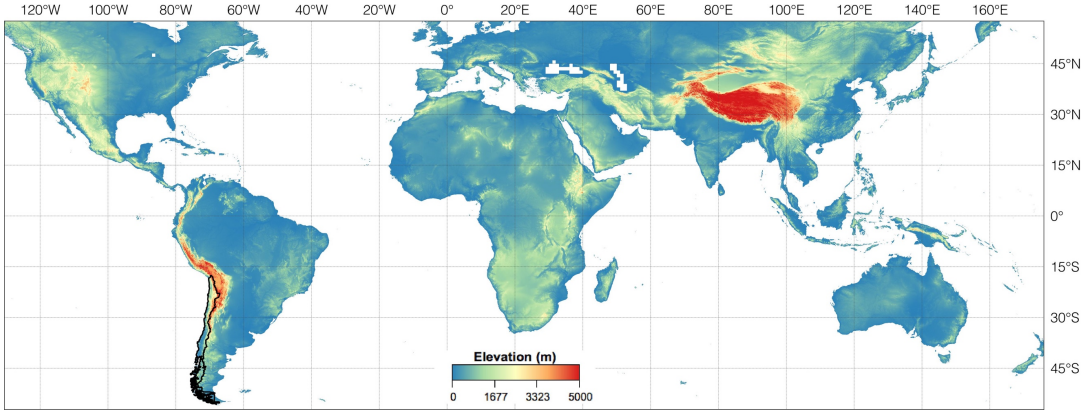

Figure 1: World map with topography at 1 km resolution (USGS, 1996). The Chilean boundary is highlighted in black.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-23
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 February 2018

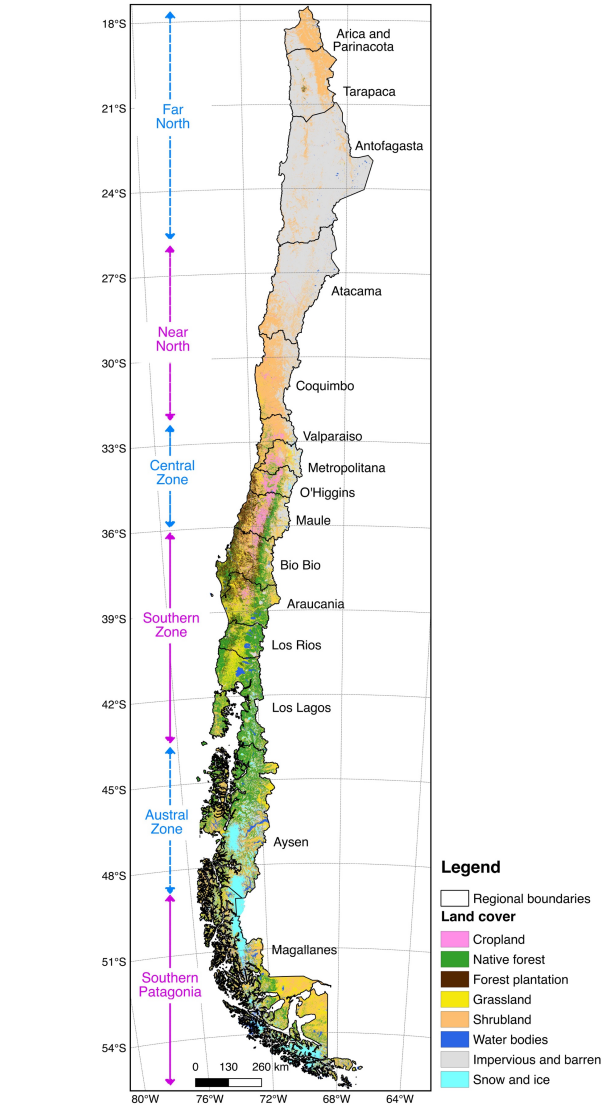

Figure 2: Chilean regional boundaries and names, and the six defined macro-zones (blue and magenta arrows). The background colour correspond to the main land cover classes, obtained from Zhao et al., (2016).

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-23
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 February 2018

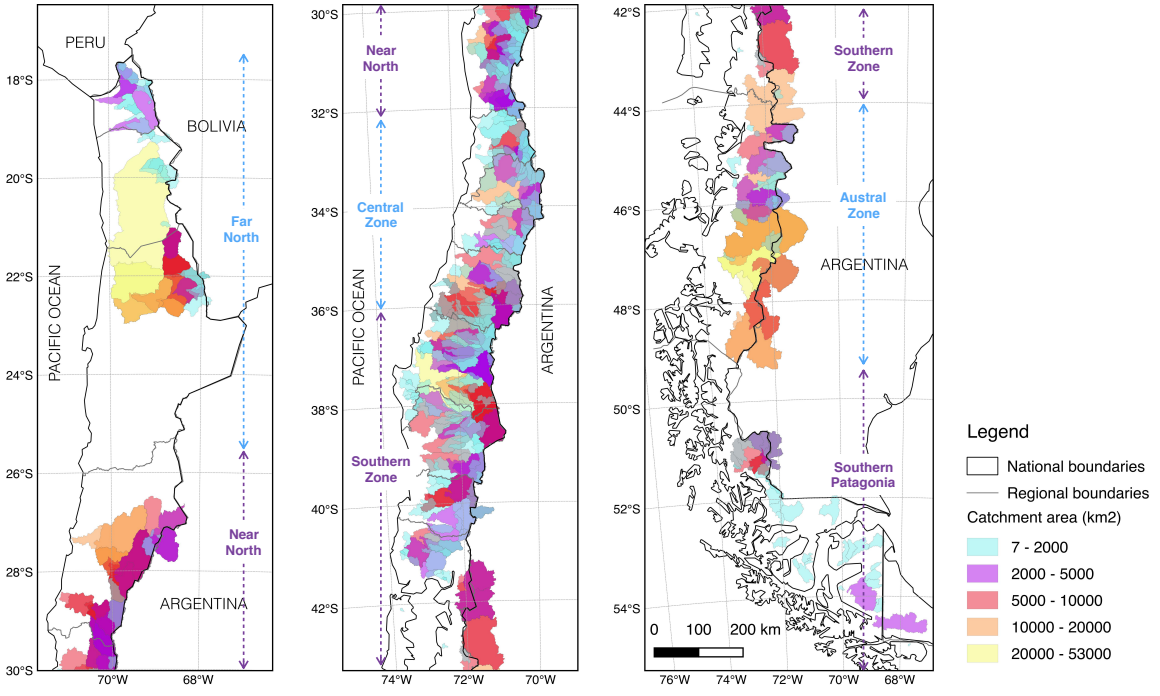

Figure 3: Catchment boundaries and contributing areas (km^2) of the 516 watersheds included in this study. The six defined macro-zones are indicated in blue and purple arrows.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-23
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 February 2018

Hydrology and Earth System Sciences
Discussions

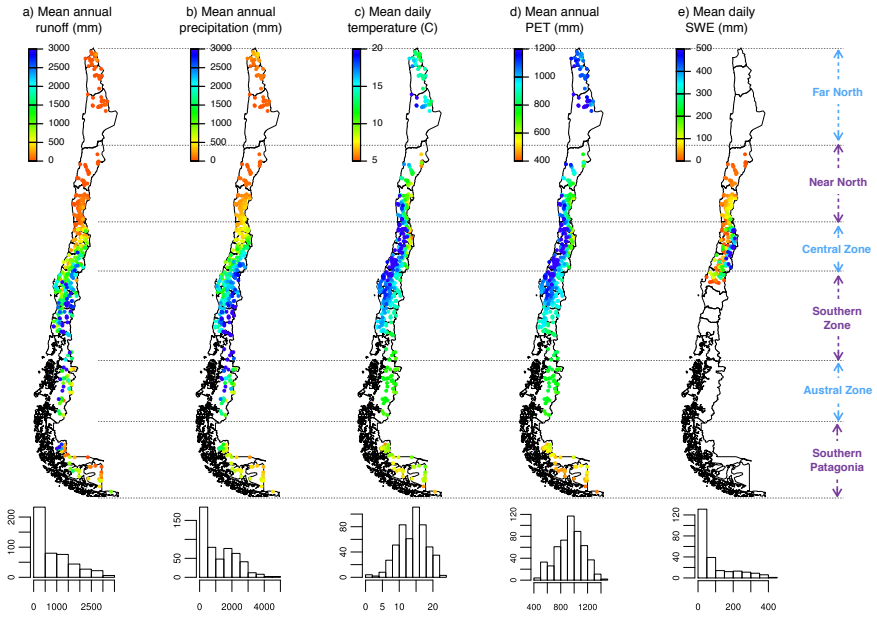

Figure 4: Mean annual hydrometeorological variables, calculated for the complete recording period of each variable. Panels b and d were generated with, $precip_{cr2met}$ and pet_{har} products, respectively. The histograms indicate the number of catchments (out of 516) in each bin. The points represent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

5

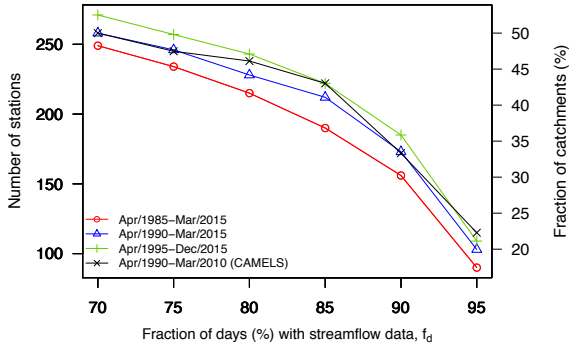

Figure 5: Number of stations (left y-axis) having at least f_d % of days with daily streamflow records, for different periods. The right y-axis shows the percentage of catchments (out of 516) that meets the criterion. We include the same period used in the CAMELS dataset (Addor et al., 2017) as a reference.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-23
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 February 2018

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

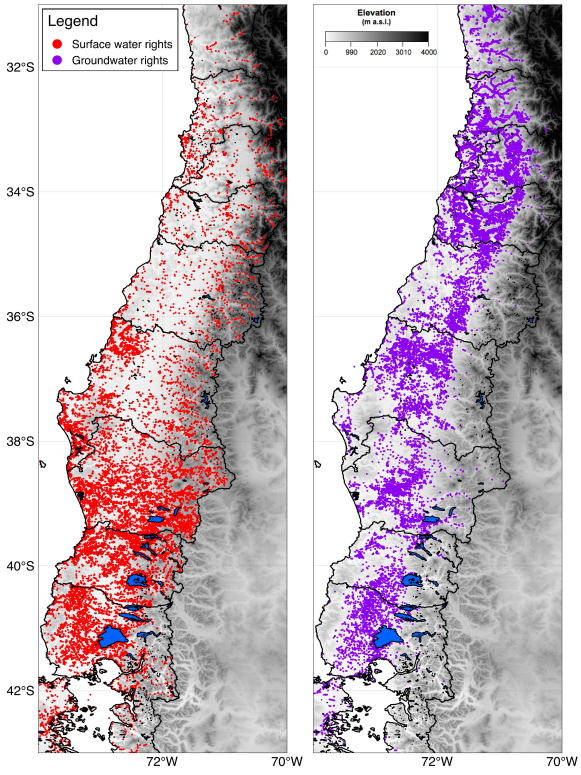

Figure 6: Surface (left panel) and ground (right panel) water rights (consumptive, non-consumptive, permanent, eventual and alternated) granted by the Chilean Water Directorate (DGA) for a portion of the country. Background colors represent topography (greyscale) and the main water bodies (highlighted in blue).

5

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-23
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 February 2018

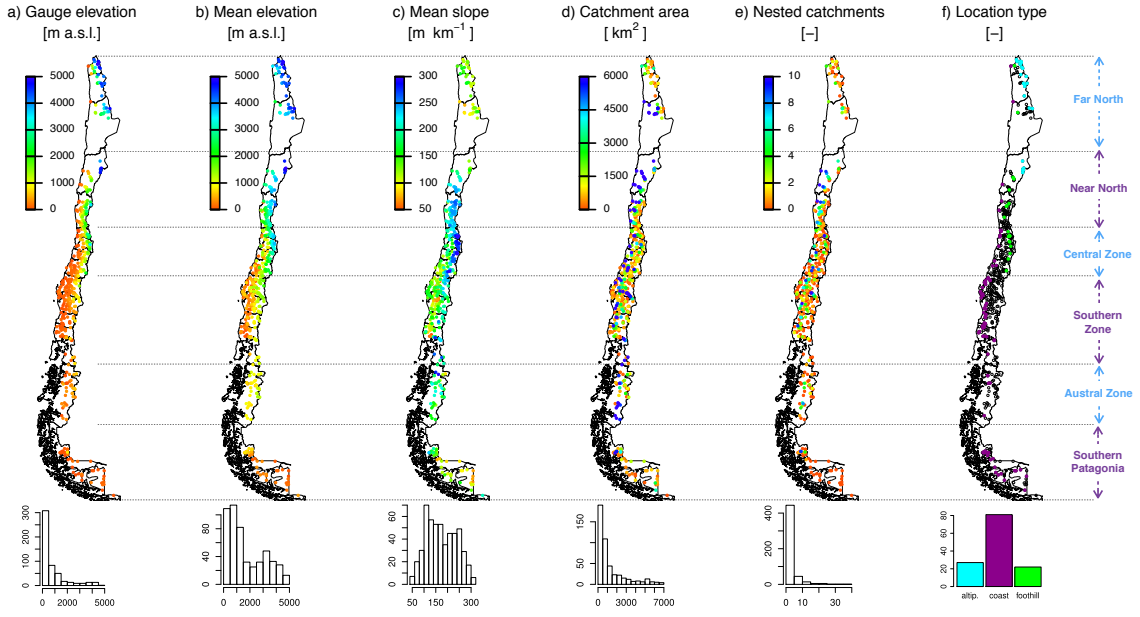

Figure 7: Location and topography. For visualization purposes, catchment areas (panel d) are shown up to their 90th percentile. The histograms indicate the number of catchments (out of 516) in each bin.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-23
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 February 2018

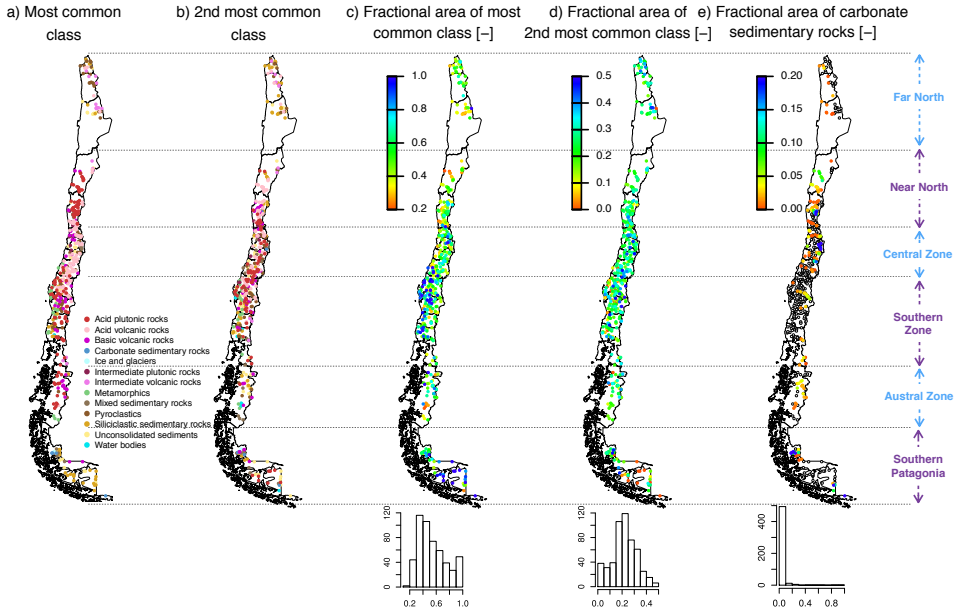

Figure 8: Geology attributes. The histograms indicate the number of catchments (out of 516) in each bin.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-23
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 February 2018

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

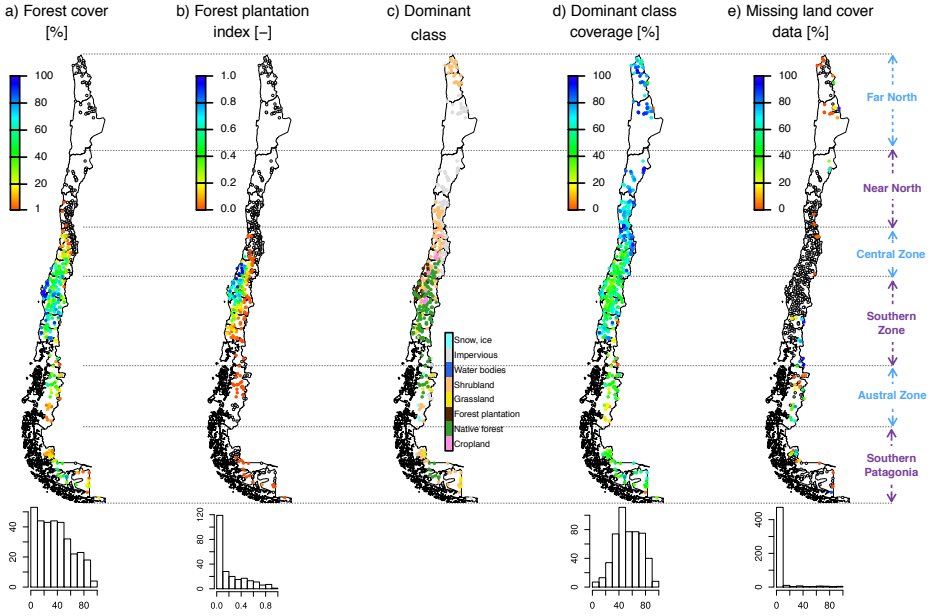

Figure 9: Land cover characteristics. Values below color bar lower limits are shown blank (i.e., for panel e, this means that there is no missing land cover data within those catchments). The histograms indicate the number of catchments (out of 516) in each bin.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-23
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 February 2018

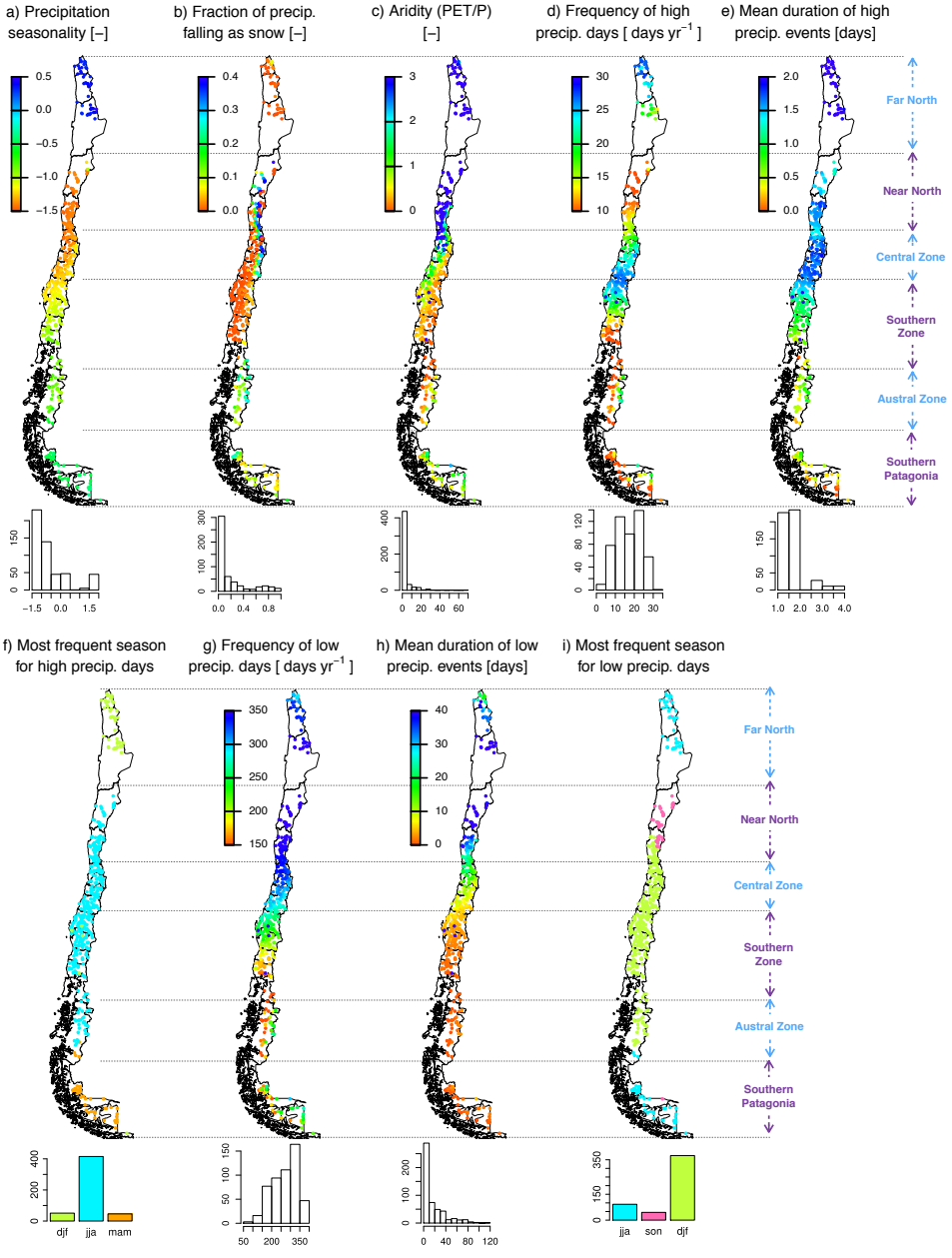

Figure 10: Climatic indices (calculated from precip$_{cr2met}$ product). For visualization purposes, attributes below the lower color bar value are blank and above the upper colobar value are blue. The histograms indicate the number of catchments (out of 516) in each bin.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-23
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 February 2018

Hydrology and Earth System Sciences
Discussions

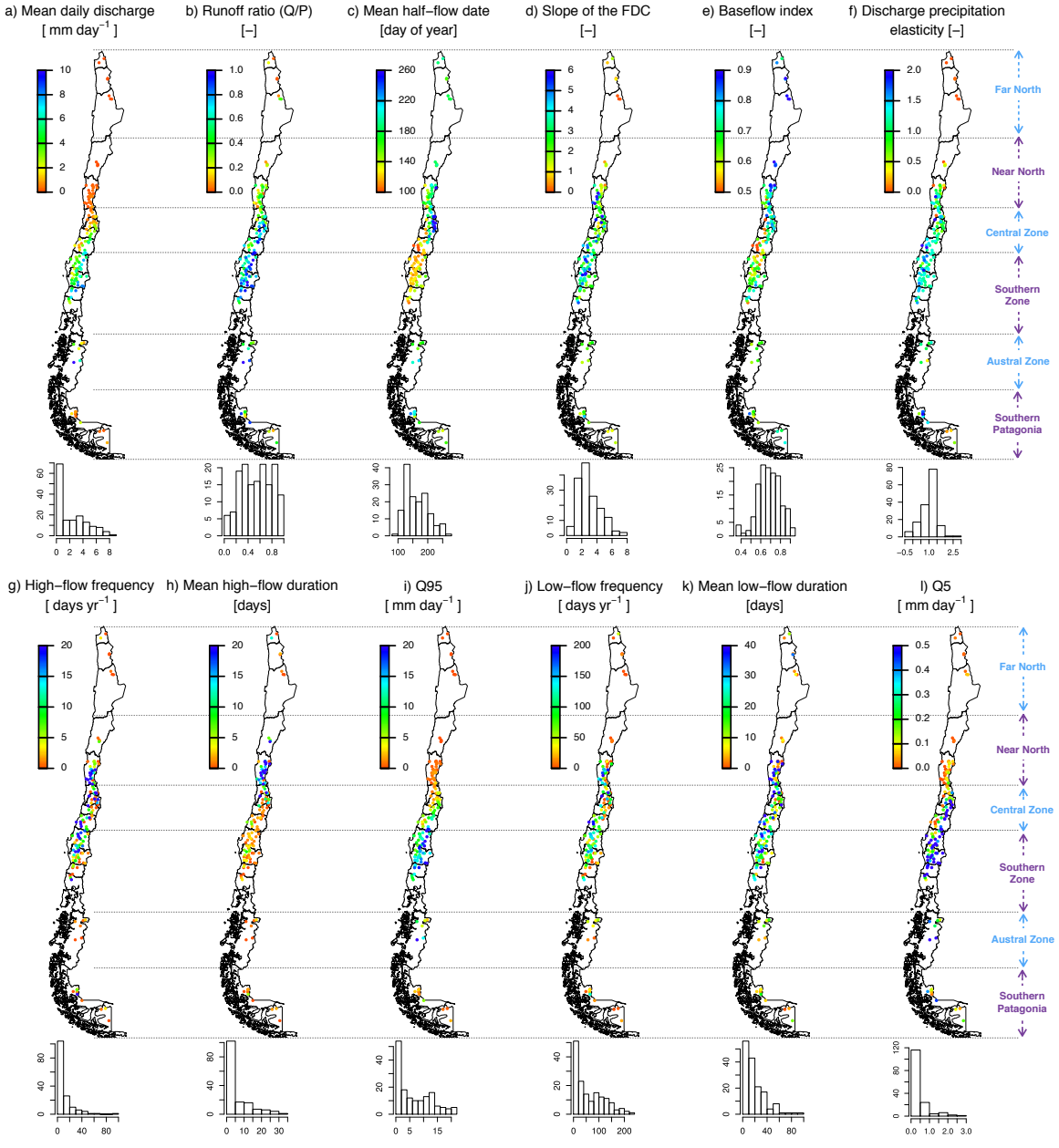

Figure 11: Hydrologic signatures. For visualization purposes, attributes below the lower color bar value are blank and above the upper colobar value are blue. The histograms indicate the number of catchments (out of 153) in each bin.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-23
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 February 2018

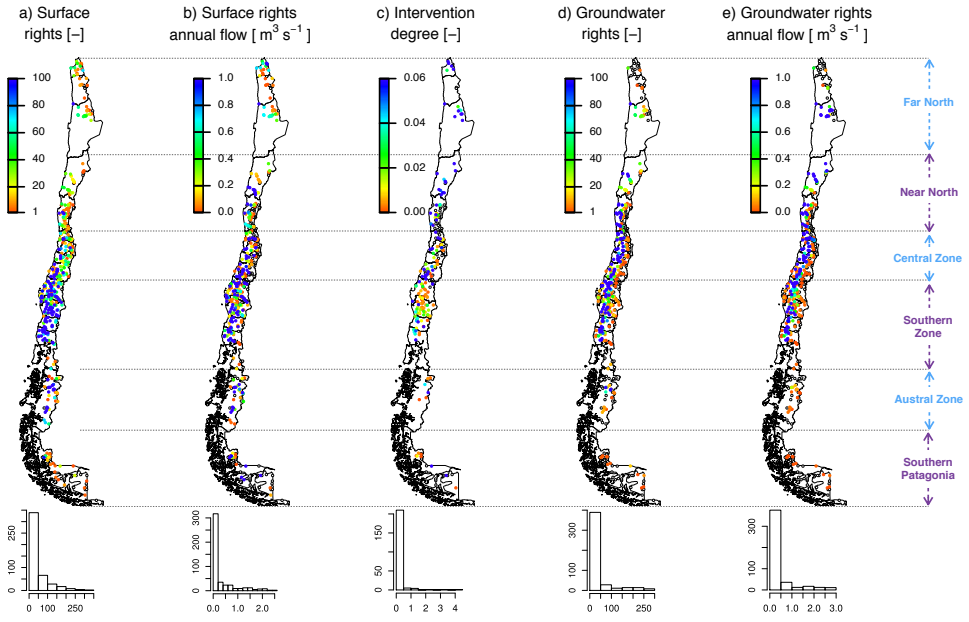

Figure 12: Water rights attributes. For visualization purposes, the attributes in panels a, b, d and e are shown up to their 90th percentile. Attributes below the lower color bar value are blank and above the upper color bar value are blue. The histograms indicate the number of catchments (out of 516) in each bin, except for panel c, which shows a total of 222 catchments (based on availability of streamflow records).

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-23
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 February 2018

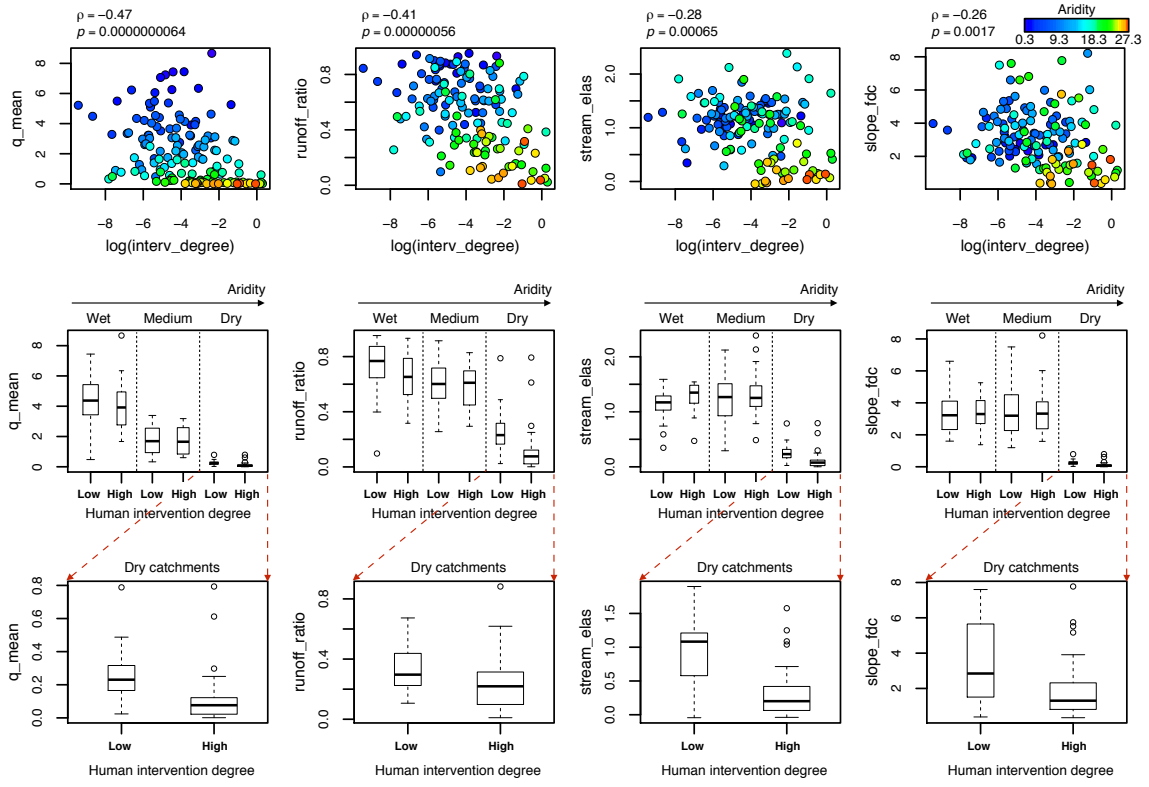

Figure 13: The top panel presents the relation between four hydrological signatures and the log-transformed human intervention degree ("interv_degree" from Table 3). The spearman rank correlation coefficients and their p-values at 95% confidence are shown for each plot. The color corresponds to the aridity index ("aridity" from Table 3). The middle panel shows the boxplots (box widths are proportional to the number of catchments in each box) of the same hydrological signatures for the catchments classified by their aridity index (wet: aridity below 0.6, medium: aridity between 0.6 and 2, and dry: aridity above 2) and by their human intervention degree (low: interv_degree below 3%, and high: interv_degree greater than 3%). The bottom panel presents a zoomed view of the dry catchments from the middle panel.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-23
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 February 2018

Hydrology and Earth System Sciences
Open Access
Discussions

Tables

Table 1: Precipitation products

Name	*Description*	*Spatial resolution*	*Temporal resolution*	*Period of record*
$precip_{cr2met}$	Obtained from de CR2METv1.3 product. Based on quality-controlled station-based rainfall records, this product uses a combination of statistical downscaling and interpolation techniques (Boisier et al., *in prep*)	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 12 additional attributes
Climatic indices	11 attributes	11 attributes adopted from A17
Hydrological signatures	13 attributes	13 attributes adopted from A17 1 additional attribute
Intervention degree	not computed	5 attributes
Total number of attributes	**59**	**38 adopted from A17** **26 introduced**

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-23
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 February 2018

Hydrology and Earth System Sciences
Discussions

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)	Section 3.2
	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)	-	-	Section 3.1
	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)	-		
	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 4.1
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 3.4 and 4.3
	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 and 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).	%		

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-23
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 February 2018

Hydrology and Earth System Sciences
Discussions

	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 3.5 and 4.3
	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. 3.6, 3.7 and 3.8, respectively.	Table 2 in A17
	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	-	Water Atlas developed by the DGA (DGA, 2016a)	Sections 3.11 and 4.6
	sur_rights_flow	annual flow calculated for consumptive permanent continuous surface water rights	$m^3\ s^{-1}$		
	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	$m^3\ s^{-1}$		