LamaH | *La*rge-Sa*m*ple D*a*ta for *H*ydrology and Environmental Sciences for Central Europe

Christoph Klingler, Karsten Schulz, Mathew Herrnegger

Institute for Hydrology and Water Management, University of Natural Resources and Life Sciences, Vienna, 1190, Austria

5 *Correspondence to*: Christoph Klingler (christoph.klingler@boku.ac.at)

Abstract. Very large and comprehensive datasets are increasingly used in the field of hydrology. Large-sample studies provide insights into the hydrological cycle that might not be available with small-scale studies. LamaH (*La*rge-Sa*m*ple D*a*ta for *H*ydrology) is a new dataset for large-sample studies and comparative hydrology in Central Europe. It covers the entire upper Danube to the state border Austria / Slovakia, as well as all other Austrian catchments including their foreign upstream areas.

10 LamaH covers an area of 170 000 km² in 9 different countries, ranging from lowland regions characterized by a continental climate to high alpine zones dominated by snow and ice. Consequently, a wide diversity of properties is present in the individual catchments. We represent this variability in 859 observed catchments with over 60 catchment attributes, covering topography, climatology, hydrology, land cover, vegetation, soil and geological properties. LamaH further contains a collection of runoff time series as well as meteorological time series. These time series are provided with daily and also hourly resolution.

15 All meteorological and the majority of runoff time series cover a span of over 35 years, which enables long-term analyses, also with a high temporal resolution. The runoff time series are classified by over 20 different attributes including information about human impacts and indicators for data quality and completeness. The structure of LamaH is based on the well-known CAMELS datasets. In contrast, however, LamaH does not only consider headwater basins. Intermediate catchments are also covered, allowing, for the first time within a hydrological large sample dataset, to consider the hydrological network and river

20 topology in applications. We discuss not only the data basis and the methodology of data preparation, but also focus on possible limitations and uncertainties. Potential applications of LamaH are also outlined, since it is intended to serve as a uniform basis for further research. LamaH is available at https://doi.org/10.5281/zenodo.4525244 (Klingler et al., 2021).

1 Introduction

Hydrology and hydrological processes are characterized by high spatiotemporal variability. Runoff generation in small-scale,

25 alpine catchments with steep and complex topography is dominated by different processes compared to lowland rivers with flat topography. The water balance in an energy-limited, humid catchment in Europe is completely different than, for example, in a water-limited catchment in dry (semi-) arid regions in Africa or Australia. A water droplet flowing via the Russian Lena into the Arctic Sea has a completely different biography than a water droplet from Rwanda in Central Africa, which reaches the Mediterranean Sea via the Nile after more than 6 600 km. Boundary conditions and major drivers for the differences are

30   the catchment properties, which can be described by characteristics regarding topography, hydro-climate, land cover, geology and soil conditions.

In order to deepen our understanding of the hydrological process and further increase the reliability of (hydrological) models, it is necessary to account for this spatiotemporal variability in our approaches. A number of international initiatives (e.g. Distributed Model Intercomparison Project (DMIP; Smith et al., 2004); Inter-Sectoral Impact Model Intercomparison Project

35   (ISI-MIP; Warszawski et al., 2014); Model Parameter Estimation Project (MOPEX; Duan et al., 2006) or Hydrological Ensemble Prediction Experiment (HEPEX; Schaake et al., 2007)) have been launched in recent decades with the aim to advance the prediction of hydrologic variables through comprehensive model benchmarking in different regions of the world. New efforts strive for creating homogeneous and consistent datasets, which serve as a solid basis towards the development of new modelling approaches.

40   In this context, a trend towards more complete and extensive datasets is apparent: 1) Remote sensing has enabled consistent and global mapping of Earth's atmosphere and surface. The Sentinel (ESA, 2021; Malenovsky et al., 2012), Landsat (NASAa. 2021; Irons et al., 2012) and MODIS (NASAb, 2021; Barnes et al., 2003) missions are probably known to a broader public. 2) New software platforms or applications for obtaining and processing these mostly very data-intense (e.g. regarding data volumes) remote sensing products also facilitate the applicability. Examples of such platforms are "Google Earth Engine"

45   (GEEa, 2021; GEEb, 2021; Gorelik et al., 2017; Klingler et al., 2020), the "Copernicus Open Access Hub" (COPa, 2021) or the "Copernicus Climate Data Store" (COPb, 2021). 3) There is growing awareness that our understanding of the complex hydrological processes can be deepened through "large-sample" studies (Gupta et al., 2014). Large-sample hydrology (LSH) includes information from a broad range of different watersheds in order to derive robust conclusions (Addor et al., 2019). Several research groups in different areas of hydrology have already focused on LSH for this reason (e.g. Berghuijs et al.,

50   2014; Blöschl et al., 2019a; Döll et al., 2016; Gudmundsson et al., 2019; Luke et al., 2017; Kuentz et al., 2017; Singh et al., 2014; Van Lanen et al., 2013). 4) Finally, data-driven models and deep learning approaches have recently gained significant attention in hydrology (Sit et al., 2020). Independent from the fact that these developments are controversially discussed (Nearing et al., 2020), their excellent performance in time series prediction, also in an ungauged setting (e.g. Kratzert et al., 2019a), is related to the ability of machine learning to identify patterns and relationships in data (Kratzert et al., 2019b). These

55   approaches however strongly depend on the availability of large-sample datasets (e.g. Kratzert et al., 2019a; 2019b; Kratzert et al., 2018).

Given the workload and scope of large-sample studies, it is reasonable to differentiate between the preparation of the data basis and the subsequent investigation and to publish the findings separately. A selection of previously published "large-sample" datasets can be found in Table 1 in Gupta et al. (2014). Other datasets for hydrologic applications include the Global Runoff

60   Reconstruction (Ghiggi et al., 2019), Global Streamflow Indices and Metadata Archive (Do et al., 2018; Gudmundsson et al., 2018), HydroATLAS (Linke et al., 2019), HydroSHEDS (Lehner et al., 2008), and the CAMELS (CAtchment Attributes and MEteorology for Large-sample Studies; Addor et at., 2017a; and references in the following section) collection. The CAMELS datasets are characterized by a consistent data preparation, consistent structure and potential limitations as well as uncertainties

are also discussed in detail. However, CAMELS only include data for individual, independent (headwater) catchments and not
for an interconnected river network (Addor et al., 2019). The first CAMELS dataset was published by Addor et al. (2017a)
and Newman et al. (2015) for the contiguous territory of the United States, containing data for 671 watersheds. Here, the basins
are mostly small- to medium sized (median basin size of 330 km², with a range of 4 km² to 25 800 km²). Further CAMELS
datasets for Chile (Alvarez-Garreton et al., 2018; 516 catchments), Brazil (Chagas et al., 2020; 897 catchments), and the United
Kingdom (Coxon et al., 2020; 671 catchments) followed later. CAMELS datasets always represent a composite of
hydrometeorological time series and static catchment attributes aggregated to observed (headwater) basins. The question, how
reasonable and applicable are meteorological and catchment properties, when aggregated to several thousands of square
kilometers as is the case, is critical.

LamaH (*La*rge-*Sa*mple *Da*ta for *H*ydrology) is a new dataset for LSH (859 observed catchments) in Central Europe and is
generally based on the structure of the CAMELS datasets. Consequently, LamaH also includes runoff time series,
meteorological forcings and static catchment attributes. However, LamaH offers some innovations. For example, LamaH also
implements a catchment delineation that represents the inter-catchment area (difference area or intermediate catchments) of
neighboring gauges, in addition to the usual catchment delineation used in CAMELS datasets, which is equivalent to the
orographic catchment area of the individual gauges. Supplementary attributes such as the gauge topology, as well as the flow
length and gradient between two adjacent gauges, are added to specify the interconnected hydrologic network. This enables,
for example, to model the local runoff generation in the intermediate catchments and the river routing separately. A further
novelty of LamaH is the finer resolution of the provided hydrometeorological time series (daily and hourly). Time series with
hourly resolution are crucial for a reliable result when modelling, for instance, the river routing or snow or glacier driven
processes, where the observed signal in runoff shows a distinct diurnal pattern.

This paper is organized as follows: After a description of the project area (section 2) and included basins and catchments
(section 3), the preparation of the hydrometeorological time series are described in section 4. Section 5 is about static catchment
properties and shows their spatial distribution. Additionally, uncertainties, limitations and restrictions of the used data sources
are discussed. Finally, section 6 includes a summary and an outlook of possible applications of LamaH.

2 Domain of coverage

LamaH covers an area of 170 000 km² in 9 different countries in Central Europe (Austria, Germany, Czech Republic,
Switzerland, Slovakia, Italy, Liechtenstein, Slovenia and Hungary; sorted by descending contributing area). Its scope includes
the upper Danube to the Austrian / Slovakian border, as well as all other catchment areas in Austria, including their adjacent
upstream areas in neighboring countries. The Piz Bernina with 4 049 m a.s.l. represents the highest point within the project
area, while the lowest point at about 130 m a.s.l. is located at the most downstream gauge of the Austrian Danube. The dominant
river is the Danube (ICPDR, 2020; Prohaska et al., 2020), which has its source in the far west of the project area near
Donaueschingen (Fig. 1, 8.15°E / 48.1°N). The catchments of Danube's main tributaries serve to divide the project area into

18 different river basins (Table 1). An overview of the domain covered in LamaH with the river regions and the 882 runoff gauges with their elevations is illustrated in Fig. 1. The difference to the 859 catchments defined in basin delineation A (see chapter 3) can be explained by the fact that 23 gauges, which mostly do not have a clearly definable catchment area (e.g. gauges at artificial channels or below large karst springs, see also chapter 5.8), were not considered in basin delineation. All

100 river regions in the project except regions 1 and 11 are part of the Danube´s catchment. River region 1 covers the upper catchment of the Rhine from its sources to Lake Constance and region 11 covers the Austrian catchment area of the Vltava, which is the largest tributary of the Elbe.

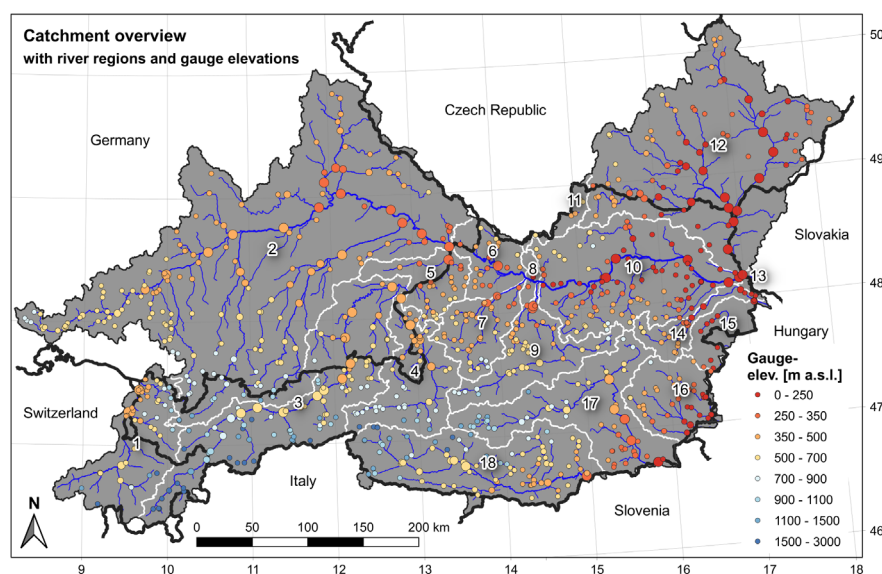



Fig. 1: Overview of the area covered in LamaH (colored in grey), and the 882 runoff gauges (the size of the circles is proportional to
105 **the catchment area). LamaH is divided into different river regions, which are bordered by the white lines. The black numbers are abbreviations for the individual regions, which are indicated in Table 1. The national borders are shown as thick black lines. Source of stream net: HydroATLAS (Linke et al., 2019).**

Table 1: Overview of the river regions.

No.	Name of river region	Area [km²]	Important headwaters and tributaries
1	Rhine	7 705	Anterior Rhine, Posterior Rhine, Plessur, Landquart, Ill, Frutz
2	Danube above Inn	50 486	Iller, Brenz, Wörnitz, Lech, Altmühl, Naab, Regen, Isar, Ilz, Vils
3	Inn above Salzach	15 735	Sanna, Ötztaler Ache, Sill, Ziller, Brandenberger Ache, Brixentaler Ache, Mangfall, Alz
4	Salzach	6 734	Krimmler Ache, Fuscher Ache, Gasteiner Ache, Großarlbach, Lammer, Saalach
5	Inn under Salzach	3 537	Mattig, Mühlheimer Ache, Rott, Pram

6	Danube between Inn and Traun	3 248	Erlau, Große Mühl, Innbach
7	Traun	4 051	Ischl, Ager, Krems
8	Danube between Traun and Enns	587	Gusen
9	Enns	6 091	Salza, Steyr
10	Danube between Enns and Morava	14 112	Aist, Naarn, Ybbs, Erlauf, Melk, Pielach, Kamp, Schwechat, Fischa
11	Vltava	997	Maltsch
12	Morava	26 666	Bečva, Thaya
13	Danube between Morava and Leitha	239	-
14	Leitha [a]	2 176	Schwarza
15	Rabnitz [a]	2 430	Einser-Kanal
16	Raab [a]	4 801	Lafnitz, Pinka
17	Mur [a]	10 438	Taurach, Pöls, Liesing, Mürz, Kainach, Sulm
18	Drava [a]	12 201	Isel, Möll, Lieser, Gail, Gurk, Lavant

[a] River joins the Danube outside the project area in Hungary / Croatia.

3 Basin and sub-basin delineations

Meteorological time series and catchment properties are usually based on global datasets, which are provided either in raster or vector form. In LamaH, a catchment property or time step of a meteorological time series usually represents the mean of the orographic catchment of a gauge. Consequently, aggregation of spatially distributed information of the corresponding source datasets was required. As aggregation areas are catchment boundaries (Fig. 2) from the Digital Hydrological Atlas of Austria (HAO, 2007) and from HydroATLAS (Linke et al., 2019) used. The sub-basins of both data sources were combined and, if necessary, adjusted to reflect the gauges catchment area at their downstream end. In a next step, all (smaller) sub-catchments, which belong to the respective upstream catchment area of an individual gauge, were combined. Each gauge is therefore assigned an aggregation area representing the overall orographic catchment area. We refer to this method of catchment / basin delineation in the further text and the dataset as "basin delineation A". Plausibility of this type of basin delineation was checked by calculating the ratio between the aggregation area and the officially, e.g. in the metadata of the gauges, declared catchment area (attribute "area_ratio" in Table A1). The median basin size over all 859 basins of LamaH applying basin delineation A is 178 km², with a range of 3.9 km² to 131 000 km². Basin delineation A is known from the CAMELS datasets. The advantage of basin delineation A is the independency between the basins, since the aggregation area fully represents the orographic catchment area of a gauge. However, for gauges with larger catchments, aggregation with basin delineation A leads to a significant loss of information, as variability as well as small-scale characteristics are lost. Another drawback is the multiple, but differently weighted, mapping of local features (see overlap in Fig. 2a). Therefore, basin

delineation A is supplemented by a form of delineation (catchment delineation B, also 859 catchments) where the orographic catchment area of the next upstream gauge (may be none, one, or more) is subtracted from that of the current gauge (Fig. 2b). This results in the representation of intermediate catchments, which become part of a large connected river network. The

130 dependency among these intermediate catchments requires a catchment or gauge hierarchy (Fig. 2b, attribute "HIERARCHY" in Table A1), as well as information regarding the upstream-downstream relationship ("NEXTUPID" or "NEXTDOWNID" in Table A1). The median basin size applying basin delineation B is 114 km², with a range of 1.3 km² to 2 496 km². The third basin delineation provided in LamaH (further referred as basin delineation C in the text and dataset) is similar to basin delineation B, but only includes catchments with no or only low anthropogenic influence (454 catchments) (Fig. 2c). The

135 intention is to provide modelers in subsequent studies with an easy to use data basis of hydrological patterns that are near to natural conditions. Anthropogenic influences in the catchments and runoff data is described more detailed in section 5.8.

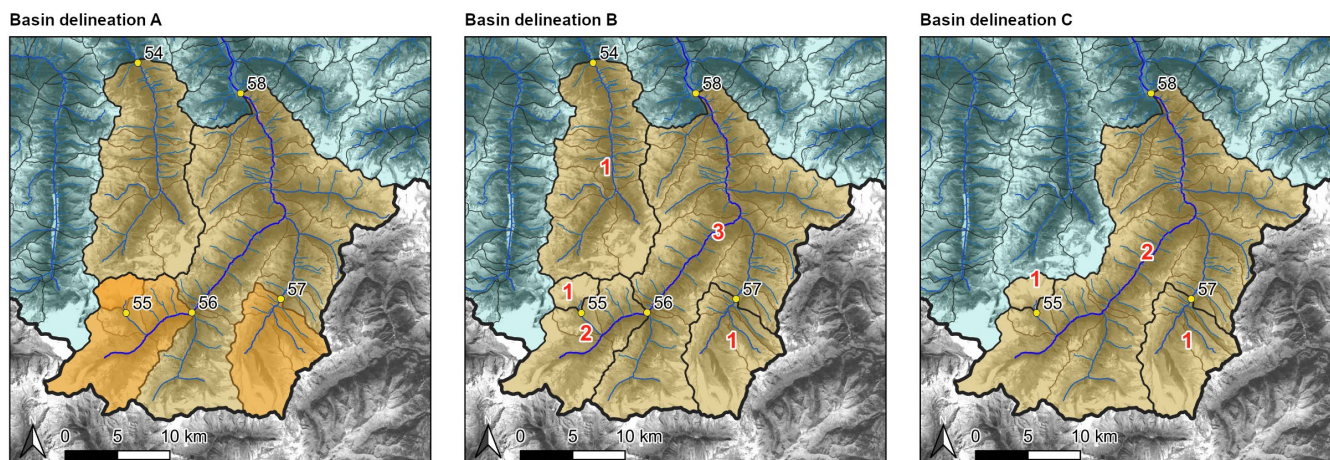



Fig. 2: Types of basin delineations shown with an example. Unaggregated sub-basins within the project area are marked with a blue tone and bordered with a thin black line. Aggregated basins are displayed with an orange/brown tone and are surrounded by a
140 **medium-thick black line. The thickest black line shows the border of the project area. The numbers in black font indicate the gauge ID, while the numbers in red show the gauge hierarchy (plot b and c). Basin delineation A (similar to the well-known CAMELS datasets): The aggregation area corresponds to the orographic catchment area of a gauge. There are overlaps with the aggregation areas of the downstream gauges. In plot A, therefore, the aggregation area of gauges 56 and 57 overlaps with that of gauge 58, and the aggregation area of gauge 55 overlaps with that of gauges 56 and 58 (indicated by the different colour tones). Basin delineation**
145 **B: The aggregation areas in this method considers the difference area (intermediate catchments) between the orographic catchment area of the respective gauge and the catchment area of the next upstream gauges. Consequently, there are no overlaps, but a gauge hierarchy becomes necessary. The hierarchy of the gauges 54, 55 and 57 is 1, because there is no upstream gauge (headwater basins). Gauge 56 has hierarchy 2, because gauge 55 with hierarchy 1 is upstream. Hierarchy 3 is assigned to gauge 58, because there is at least one gauge with hierarchy 2 (gauge 56) in the upstream area. Basin delineation C: Similar to basin delineation B, but only**
150 **uninfluenced or low-influenced gauges / catchments (see chapter 5.8) are considered. In plot C, it is assumed that gauges 54 and 56 are strongly influenced. Consequently, these two gauges are excluded from the basin delineation. The aggregation area of gauge 58 (now hierarchy 2) includes the intermediate catchment area of gauge 56. Source of background satellite image: Google © 2020 TerraMetrics, Kartendaten © 2020, source of stream network: TYROL (2020).**

4 Hydrometeorological time series

155 ### 4.1 Runoff data

LamaH contains daily and hourly runoff time series for 882 different gauges, located in 4 different countries (Austria, Germany, Switzerland and Czech Republic). The main provider of runoff time series was the Hydrographic Central Bureau of Austria (HZB, 2020), which contributed data for 609 gauges located in Austria. The hydrographical services of the German federal states Bavaria (GKD, 2020) and Baden-Württemberg (LUBW, 2020) provided 125, respectively 61 runoff time series.

160 25 runoff time series came from the hydrological office of Switzerland (BAFU, 2020), while time series for 61 gauges were provided by the Czech Hydrometeorological Institute (CHMI, 2020). The format of all obtained time series was unified, enabling much easier data processing. The various gauge attributes and metadata are listed and described in Table A1. The unit of discharge is $m^3 s^{-1}$ for both daily and hourly resolution. Conversion to runoff heights can be performed using the catchment area provided (attribute "area_gov" in Table A1).

165 Runoff time series are in most cases derived by water level - discharge relationships (rating curves). Changes in channel profile, e.g. after floods with strong bedload transport, can lead to an incorrect runoff calculation. However, attempts are usually made to minimize this source of error by periodically adjusting the rating curve. The adjustment frequency of these rating curves is not publicly available, but only gauges in the highest quality class (quality classes are declared in Bavaria and the Czech Republic) were included in LamaH. Runoff time series with daily resolution are often provided with longer observation periods

170 than those with hourly resolution. Therefore, daily and hourly runoff time series can be obtained separately from the listed hydrological offices. However, we normally requested only the time series with hourly resolution and derived the daily time series from them. This approach was chosen for the runoff data from Austria, Germany and Switzerland, since those time series with hourly resolution mostly include quite long recording periods. Fig. 3a and the included histogram show that most gauges have continuous data recording since the late 1970`s. In contrast, the time series from the Czech Republic were

175 requested with both daily and hourly resolution, as the continuous (hourly) time series here only starts after 2005. The runoff time series in LamaH were limited to the period 1981 - 2017, because 1981 was the starting year of the meteorological ERA5-Land forcings (see section 4.2), and 2017 was the last year for quality-controlled runoff data from Austria at the point of request.

Although the exact scope of data verification by the staff of the various hydrological services is not further specified, we have

180 added an attribute describing the check status (attribute "checked" in Table A1) to each time step of the runoff time series. The Austrian, Czech and Swiss runoff data are provided exclusively checked, while the runoff data from the Bavarian hydrographic service is in most cases quality controlled until years 2014 - 2016. Data from the German federal state Baden-Württemberg is often checked only from the year 2010 onwards. Some time series contain gaps, also after checking by the hydrological services. In order to reduce the number of gaps, up to 6 consecutive gaps (6 hours) were filled with linear interpolation during

185 our processing. Any remaining gaps in the time series were marked with the number -999. The fraction of remaining gaps in the continuous runoff time series is declared by the attribute "gaps" (Table A1) and illustrated in Fig. 3b. It is shown that those

gauges with very few gaps (< 0.1‰) are mostly located in Austria, Czech Republic and Switzerland. About 80% of the 882 gauges have no gaps in their continuous time series after our processing. The time steps with gaps before our processing are listed in separate files, attached to the dataset. The spatial distribution of the gauge hierarchies from basin delineation B (see

190   Fig. 2b) is mapped in Fig. 3c., where 50% of all gauges have a hierarchy of 1 and thus represent headwater catchments. The highest hierarchy (26) is found for the very last downstream gauge of the Austrian Danube (ID 399).

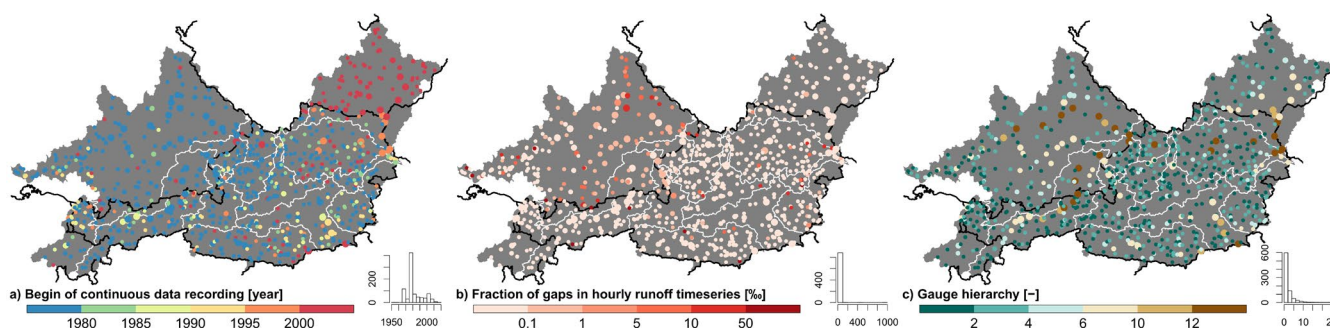



Fig. 3: Maps showing a selection of gauge-referenced attributes. The size of the circles is proportional to the respective catchment area. The histograms indicate the number of gauges (out of 859) in each category.

195   **4.2 Meteorological data**

Given the extent of the ECMWF (European Centre for Medium-Range Weather Forecasts) ERA5-Land dataset with global coverage (Muñoz Sabater et al., 2021), it was possible to obtain gap-free time series with daily and hourly resolution for 15 meteorological variables and 39 years (Table A2). ERA5-Land is a derivative of the ERA5 climate reanalysis (Hersbach et al., 2020), but only covers the terrestrial components. Further developments compared to ERA5 include an interpolation package

200   for a finer temporal resolution, an additional sea level adjustment of the meteorological fields, as well as more efficient possibilities for the import of updates (Muñoz Sabater et al., 2021; Yang and Giusti, 2020). ERA5-Land has a spatial resolution of 0.1 arc degrees (about 9 x 11 km at the latitudes of the project area) compared to the grid size of ERA5 of 0.25 arc degrees. The temporal resolution of ERA5-Land is 1 hour, while ERA5 only has a 3-hour resolution. There is no data assimilation (fitting to observations) applied for ERA5-Land, but observations are indirectly implemented via the assimilated atmospheric

205   fields of ERA5 (Hennermann and Guillory, 2020; Yang and Giusti, 2020). In accordance to ECMWF regulations, an uncertainty estimate for ERA5-Land will be released (Muñoz Sabater, 2019b; Muñoz Sabater, 2017), but was not available at the time of writing (January 2021).

The meteorological time series were determined for all 3 forms of basin delineation (A/B/C in chapter 3). In order to do so, we intersected the aggregation areas of the various basin delineations with the gridded data of ERA5-Land in order to get area

210   fractions for an area-weighted aggregation (arithmetic averaging). As already mentioned in the introduction, we would also like to point out possible uncertainties of the published data. We therefore determined the components of the water balance for the period 01.10.1989 to 30.09.2009 and plotted them (Fig. 4a). Values of catchments influenced by cross-basin water transfers, water withdrawals or intakes, large karstic springs or high infiltration (see chapter 5.8) are not shown in Fig. 4a/c to allow a

more objective interpretation. In case of long-term water balances, it is usually feasible to neglect artificial storage in the
215 catchment. The difference between long-term mean precipitation (P) and runoff height (Q) should be equal to the total
evapotranspiration (ETA) in a fulfilled water balance. This would be shown by having all points in Fig. 4a on the 1:1 line,
which is not the case. Reasons for the rather strong scatter (Pearson correlation R = 0.30) may be an insufficient representation
of precipitation or total evapotranspiration by ERA5-Land, an inaccurate recording of runoff (e.g. strong, unrecorded
groundwater flow or change in river profile at the gauging station and thus inadequate water level - discharge relationship), a
220 significant deviation between orographic and hydrographic catchment area (subsurface inflows and outflows, especially in
karstic areas), or lastly, in case of existing glaciers, a negative mass balance (Lambrecht and Kuhn, 2007; Kuhn, 2004;
Kobolschnig and Schöner, 2011; Oerlemans et al., 1998; WGMS, 2005). Using other precipitation datasets for the same
evaluation does not result in a significantly more compliant long-term water balance. CHIRPS Daily v2 (Funk et al., 2015)
resulted in a correlation R between (P - Q) and ETA of 0.34 or MSWEP v2.2 (Beck et al., 2017; 2019) in even a lower R-value
225 of 0.26. Even if we cannot resolve the issue at hand, the total evapotranspiration from ERA5-Land and its dependence on
elevation seems quite plausible compared to other studies (Fig. 4a; HAO, 2007, Map 3.3; Herrnegger et al., 2012, Fig. 20).
Negative differences of mean precipitation and runoff height (Fig. 4a) and thus runoff coefficients > 1.0 (Fig. 4c, 32 of 594
catchments) are mainly present in higher terrain (negative mass balance of glaciers, Fig. 8d) and in catchments with a high
fraction of carbonate sedimentary rocks (indicator of karst, Fig. 11c). Since ERA5-Land indirectly incorporates in-situ
230 observational data via the assimilated atmospheric fields of ERA5 (Yang and Giusti, 2020), systematic measuring error of a
terrestrial station being used could also explain insufficient mean precipitation (Herrnegger et al., 2018). The individual
components of the water balance are attached to the dataset for every catchment, since this evaluation might be useful for
explaining any deviations in a later modelling.

The Budyko curve (Fig. 4b; Budyko 1974) describes the relationship of the ratio current evapotranspiration / precipitation
235 (ETA/P) to the ratio potential evapotranspiration / precipitation (PET/P) and indicates whether evapotranspiration of a
catchment is limited by energy or water. Ideally all points should lie in the proximity of the Budyko curve. The deviation from
this ideal case can primarily be explained by a significantly too high PET of ERA5-Land over nearly the entire range of
elevation. For example, 98% of all 859 watersheds show mean annual PET sums above 1000 mm. As these PET sums are not
realistic at the latitudes of the project area (HAO, 2007, Map 3.2; Herrnegger et al., 2012, Fig. 17), we did not include the
240 potential evapotranspiration of ERA5-Land in the LamaH dataset. The runoff coefficient (Q/P) as a function of the ratio mean
precipitation / total evapotranspiration (P/ETA) is shown in Fig. 4c. The altitudinal dependency can be clearly seen in Fig. 4c,
while catchments with lower mean elevation show less scatter. Fig. 4d shows the contrast of the long-term precipitation of
ERA5-Land with those of ERA5 (Pearson correlation R = 0.936). ERA5 indicates systematic surplus of precipitation at
catchments with mean altitudes above 2000 m a.s.l. compared to ERA5-Land, while at catchments with mean altitudes between
245 800 and 1200 m a.s.l. the opposite is more likely to be the case. The correlation between the long-term precipitation of ERA5-
Land and those of the dataset CHIRPS daily v2 (Funk et al., 2015) is 0.916 (Fig. 4e). Further, the mean precipitation sums of
CHIRPS daily v2 tend to be lower than those of ERA5-Land over the whole range of altitudes. More scatter (R = 0.841)

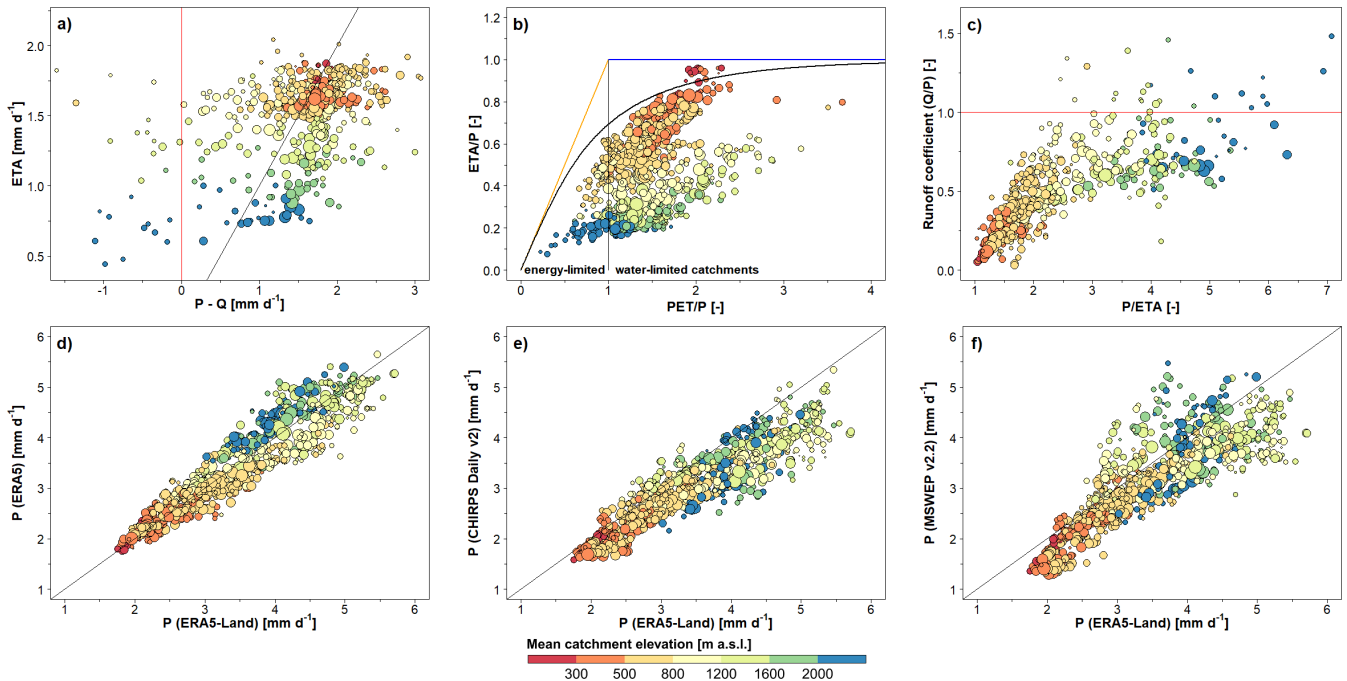
appears especially in catchments with higher elevations when comparing the long-term precipitation sums of ERA5-Land and MSWEP v2.2 (Beck et al., 2017; 2019; Fig. 4f).

250

Fig. 4: Analysis regarding the long-term water balance, evapotranspiration and comparison of the ERA5-Land`s mean precipitation with other datasets for the hydrological years 1990 - 2009 and basin delineation A. a) Total evapotranspiration (ETA) from ERA5-Land as a function of the difference between precipitation (P) from ERA5-Land and observed runoff depth (Q). b) Budyko curve indicates if ET of a catchment is limited by energy (PET/P < 1) or by water (PET/P > 1). c) shows the runoff coefficient (ratio of Q
255 **and P) as a function of the fraction of ERA5-Land´s precipitation and total evapotranspiration. In figures a), b) and c) only basins are plotted, where the corresponding gauge has data for the period 01.10.1989 - 30.09.2009. Further, in figures a) and c) only basins are plotted, which are not affected by artificial water input or withdrawal, karstic springs or high infiltration (see chapter 5.8). Figures d) / e) and f) illustrates the relationship between ERA5-Land´s precipitation compared to the datasets "ERA5", "CHIRPS Daily v2.0" and "MSWEP v2.2" in 882 basins. The diagonal black line in a), d), e) and f) is the 1:1 line. The red lines in a) and c)**
260 **show physical constraints. The sloped orange line in b) indicates the energy-limit, while the horizontal blue line represents the water-limit. The curved black line in b) represents the Budyko curve. The size of the symbols in all plots is proportional to the catchment area, while the color indicates the mean elevation of the catchment (see legend at bottom).**

5 Catchment attributes

The various physio-geographical characteristics of a catchment, as well as their interactions, are essential for water storage
265 and transport on and below earth's surface (Blöschl et al., 2013). The spectrum of influencing catchment characteristics includes topography, climate, hydrology, land cover, vegetation, soil, geology, as well as the type and degree of (anthropogenic) impact on runoff processes. Furthermore, catchment attributes are crucial to determine interrelationships among different watersheds along several gradients (Addor et al., 2017a; Falkenmark and Chapman, 1989; Fan et al., 2019).

In most cases, we used freely available datasets with global or at least European coverage for deriving the different catchment
270 attributes. Aggregation of the spatially distributed information of the basic datasets is performed for each of the 3 different

basin delineations (A/B/C, according to chapter 3) and by default by calculating the area-weighted arithmetic mean (otherwise indexed). Furthermore, the upscaling is performed by two different approaches: 1) In the first approach (referred as "upscaling approach: 1") the aggregation is based on all the raster cells, which centroids are located inside the catchment, whether the catchment completely covers them or not. Especially at small catchments it is possible that no or only one raster centroid

275   intersects the basin delineation. In this case the aggregation is calculated from all contributing raster cells using an area-weighted mean. Upscaling approach 1 is mainly used for relatively fine-gridded data sources (< 1 km), since it is not that computing-intensive and potential inaccuracies are negligible. 2) The second approach of upscaling (referred as "upscaling approach: 2") exclusively performs the aggregation area-weighted. Upscaling approach 2 is used for coarser gridded and vectoral data sources. The applied approach is indicated in the corresponding tables in the appendix.

280   The data basis for LamaH, methods of processing, possible uncertainties as well as the spatial distribution of catchment properties (Addor et al., 2017b) are discussed in more detail in the following subsections. It is clear that due to the large dataset this account is far from complete. The individual attributes are listed in tabular form in the appendix with a more detailed description, units and with reference to the data sources.

5.1 Topographic indices

285   We calculated 10 topographic attributes, which are listened in Table A3. The attribute "area_calc" describes the calculated aggregation (catchment) area, depending on the applied method of basin delineation (see chapter 3). Basin delineation A shows that about 34% of all 859 basins (aggregation areas) are smaller than 100 km^2, 50% are between 100 and 1 000 km^2, 14% are between 1 000 and 10 000 km^2, and about 2.8% are larger than 10 000 km^2. Large catchment areas are especially present for the gauges at the Danube and its larger tributaries (Fig. 5a). One reason for using multiple basin delineations is the reduction

290   of aggregation areas and thus providing a more representative representation of local conditions and maintain natural variability. When applying basin delineation B, about 45% of all 859 aggregation areas have an area of less than 100 km^2, 52% between 100 and 1 000 km^2 and only 2.3% have an area above 1 000 km^2.

A key factor for hydrologic processes is elevation, as it affects numerous other catchment characteristics including climate, land cover, vegetation, or soil development (Addor et al., 2017a). We derived the mean catchment elevation (Fig. 5b,

295   "elev_mean" in Table A3), the median elevation ("elev_med"), standard deviation within a catchment ("elev_std"), the elevation range (maximum-minimum elevation in the catchment, Fig. 5c, "elev_ran"), as well as the mean catchment slope (Fig. 5d, "slope_mean") from NASA's SRTM dataset (Farr et al., 2007). SRTM features a grid size of 30 m and provides a maximum global absolute error of 16 m at a 90% confidence interval, while accuracy decreases with increasing elevation and slope (Farr et al., 2007). The slope was calculated with the algorithm of Horn (1981) using the terrain elevation from SRTM.

300   High mean catchment elevations and slopes are most apparent in the Eastern Alps, which extend from the southwest to the central east of the project area. This high elevated area is mainly surrounded by the flatter Alpine foothills and regions with older geological zones (Fig. 5b).

The shape of the catchment area and the stream network also influence runoff formation. The direction of precipitation in relation to the longitudinal axis of the catchment is of major interest in case of flood situations, especially in larger catchments. For this reason, we also specified the angle between the north direction and the longitudinal axis ("mvert_ang") in addition to the distance of the longitudinal axis of a catchment ("mvert_dist"). In combination with the two wind components of ERA5-Land ("10m_wind_u", "10m_wind_v" in Table A2) it is possible to derive the relative rainfall trajectory. The shape of the catchment is also relevant for the rise of the flood wave. The attribute of length elongation according to Schumm (1956) (Fig. 5e, "elon_ratio") is an indicator regarding the "roundness" (the higher, the rounder) of the catchment. But irregularities like large notches in catchment's shape may reduce the significance of this attribute. Stream density (Fig. 5f, "strm_dens") is a function of several characteristics (e.g. climate, relief, soil properties, geology, vegetation, land use, glaciation or karstification) and can therefore be an informative indicator for comparing watersheds (Olden and Poff, 2003). The EU-Hydro-River Network Database (EEA, 2019) is used for calculating the stream density, since it is consistent and a fine-resolved dataset, which is important in this respect.

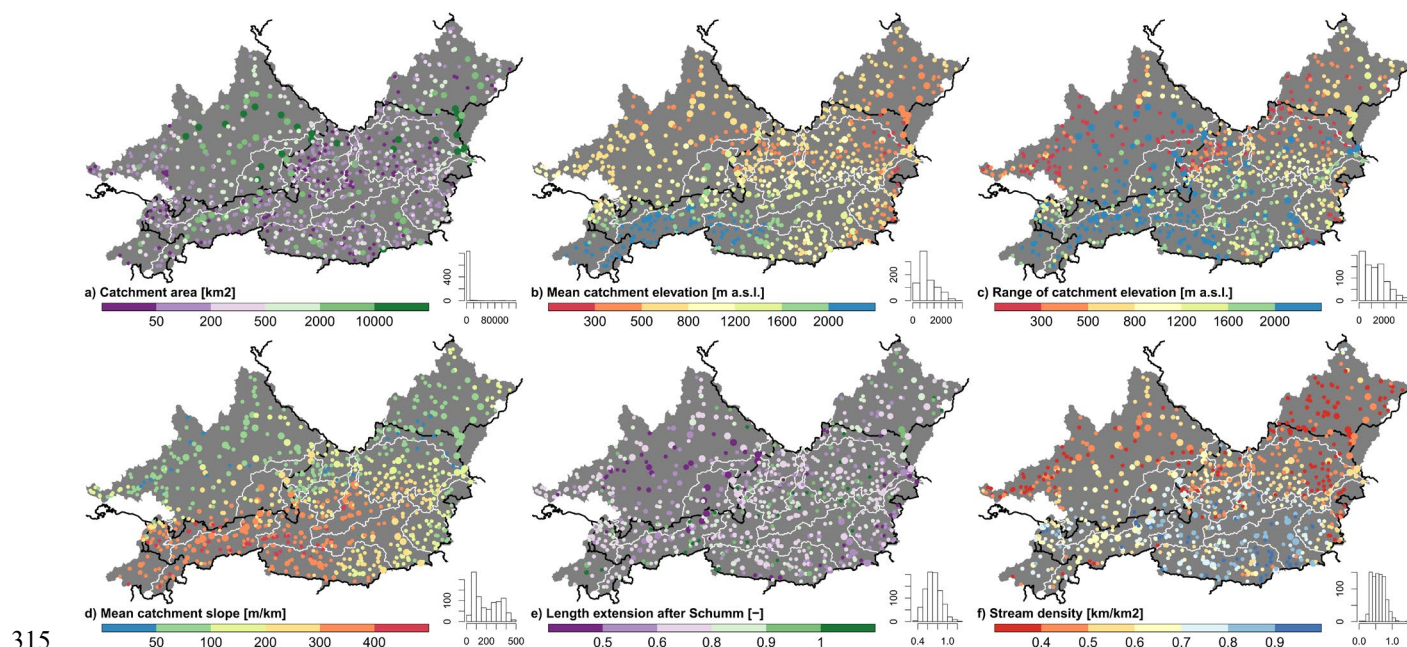



Fig. 5: Spatial distribution of a selection of topographic attributes representing the characteristics of the entire orographic catchment (catchment delineation A, Fig. 2a). The histograms indicate the number of basins (out of 859) in each category. The size of the circles is proportional to the catchment area.

5.2 Climatic indices

LamaH includes 12 different attributes reflecting aspects of climatic characteristics (Table A4). These attributes were calculated mainly from the meteorological time series of ERA5-Land for the period 01.10.1989 to 30.09.2009 (Addor et al., 2017b). As an alternative to ERA5-Land´s potential evapotranspiration, which shows unrealistic high values (see section 4.2), the Reference Evapotranspiration (ET0) from the "Global Aridity Index and Potential Evapotranspiration (ET0) Climate

Database v2" (Trabucco and Zomer, 2019), which was computed for the period 1970 to 2000, is provided. ET0 describes the

325 atmosphere's capacity for evapotranspiration given defined vegetation characteristics. Potential evapotranspiration (PET) can

be derived from ET0 using correction factors for vegetation and soil properties (Allen et al., 1998; Hargreaves, 1994).

Long-term climatic characteristics are described by long-term daily precipitation (Fig. 6a, "p_mean" in Table A4), reference

evapotranspiration (Fig. 6b, "et0_mean"), total evapotranspiration ("eta_mean"), and the aridity index (Fig. 6c, "aridity"),

which is defined as the ratio between ET0 and precipitation. The spatial pattern of long-term precipitation sums (Fig. 6a)

330 clearly shows an elevation gradient and blocking effects along the northern Alps. The west of the project area is characterized

by higher mean precipitation due to the stronger influence of oceanic climate. The relationship between mean catchment

elevation (Fig. 5b) and long-term reference evapotranspiration (Fig. 6b, Pearson correlation R = -0.79), aridity (Fig. 6c, R = -

0.67), or the fraction of precipitation falling as snow (Fig. 6g, R = 0.96) show similar spatial patterns. About 14% of all

catchments, which are exclusively located in the eastern part of the project area, have aridity (ET0/P) greater than 1.

335 Attributes characterizing seasonality are the fraction of precipitation falling as snow (Fig. 6g, "frac_snow") and the seasonality

index, which relies on sinusoids to describe the precipitation cycle over the year (Fig. 6d, "p_season"). A higher positive

seasonality index indicates higher precipitation sums during summer, while values near 0 show a more balanced precipitation

distribution throughout the year.

While long-term and seasonal indices describe general climatology, they provide little or no information about relatively short-

340 term events such as drought or heavy rainfall. Consequently, we also calculated attributes representing the frequency of high

precipitation days (days per year with at least 5 times mean daily precipitation; Fig. 6e, "hi_prec_fr") and dry days (days per

year with max. 1 mm d^{-1} precipitation; Fig. 6h, "lo_prec_fr"), their mean duration (Fig. 6f, "hi_prec_du" / Fig. 6i,

"lo_prec_du"), and the most likely season of occurrence ("hi_prec_ti" / "lo_prec_ti"). The reason for the higher frequency of

high precipitation days in the south-eastern part of the project area (Fig. 6e) is primarily the combination of relatively rainfall-

345 rich convective precipitation events during the summer months and relatively low precipitation sums during the rest of the

year (Fig. 6d). For both, the mean frequency of dry days (Fig. 6h, R = -0.62) as well as their mean duration (Fig. 6i, R = -0.57),

a negative spatial correlation with the mean catchment elevation (Fig. 5b) can be observed. Most common season for high

precipitation is for 89% of all 859 basins summer (June, July and August), while winter (December, January and February) is

the most common season for dry days in 89% of the basins.

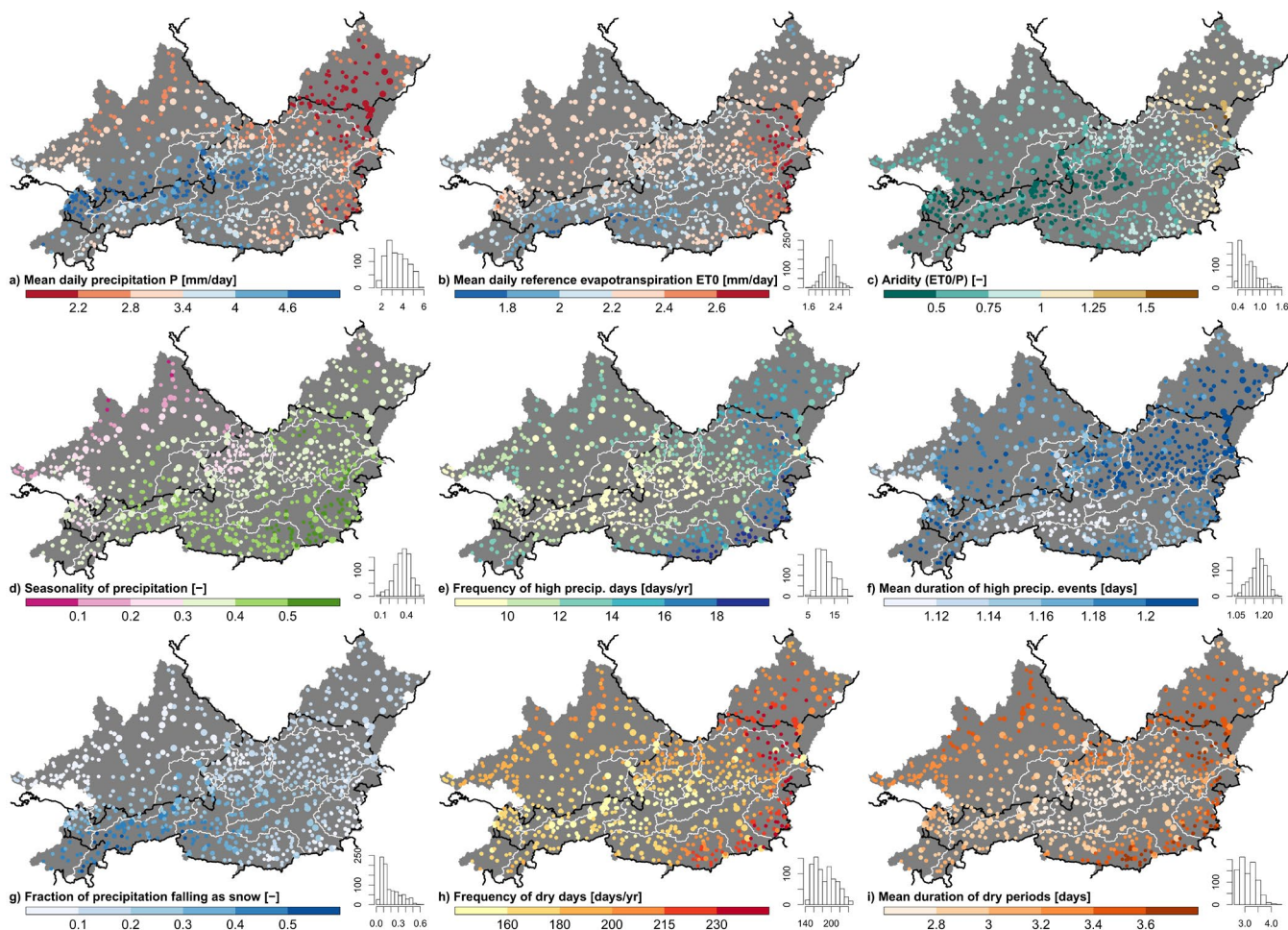



Fig. 6: Spatial distribution of a selection of climate indices representing the characteristics of the entire orographic catchment (catchment delineation A, Fig. 2a). The histograms indicate the number of basins (out of 859) in each category. The size of the circles is proportional to the catchment area.

5.3 Hydrological indices

355 The runoff time series are characterized by 14 different attributes (Table A5), which were calculated for the period 01.10.1989 to 30.09.2009 (Addor et al., 2017b). The indices were computed for those gauges, which cover the whole period of investigation (717 gauges). However, the evaluations for the entire period of record (first 01. October to 30. September 2017) are additionally made available within the dataset. Hydrological attributes can be divided into those describing long-term characteristics, seasonality, and more short-term situations such as high and low flow.

360 Aridity by itself can be a good predictor for runoff occurrence in a catchment (Arora, 2002; Blöschl et al., 2013; Budyko, 1974). This is also shown by the similar spatial pattern of long-term runoff height (Fig. 7a, "q_mean" in Table A5, R = -0.70), and runoff coefficient (Fig. 7b, "runoff_ratio", R = -0.65) compared to those of aridity (Fig. 6c). The runoff coefficient (Q/P) is the fraction of precipitation that drains a surface after deducting evapotranspiration, groundwater flow or change in storage

in the long-term. Explanations for runoff coefficients greater than 1 are given in chapter 4.2. The ratio of baseflow to total
365 runoff can be a useful indicator for watershed classification (Sawicz et al., 2011; Fan, 2015), and is further referred as the
baseflow index ("baseflow_index"). It should be noted that this index is highly dependent on the method used to separate the
hydrograph (Beck et al., 2013; Chapman, 1999; Eckhardt, 2008). For this reason, we used the Ladson filter (Landson et al.,
2013) and the approach of Tallaksen and Van Lanen (2004) for hydrograph separation. The runoff-precipitation elasticity
("stream_elas") characterizes the inertia of change in mean runoff given a change in mean precipitation (Sankarasubramanian
370 et al., 2001). For example, a value of 3 would indicate a change in runoff of 3% given a change in precipitation of 1%. High
runoff-precipitation elasticity is especially present in the eastern part of the project area (Fig. 7f). The fraction of days without
discharge (not shown, "zero_q_freq") may indicate strong infiltration (e.g. Danube Sinkhole; Hötzl, 1996), artificial water
withdrawal, or ceasing baseflow.

The seasonality of runoff is expressed by the attribute "hfd_mean", which shows the number of days from the beginning of the
375 hydrologic year (01. October) to the date when half annual of the runoff volume is reached (Court, 1962). Higher number of
days in Fig. 7c can be explained primarily by water storage in form of snow (Fig. 6g) or glaciers (Fig. 8d). Variability in runoff
(Fig. 7d, "slope_fdc") is expressed within LamaH by the slope of the flow duration curve between the log-transformed 33rd
and 66th runoff percentiles (Sawicz et al., 2011). High values are indicative for high runoff variability over the year, which can
be caused by seasonal water storage in the form of snow (Fig. 6g) or a strong response of runoff to precipitation (Yokoo and
380 Sivapalan, 2017). Extreme runoff events such as high or low flow are described by indices representing mean frequency (Fig.
7g, "high_q_freq" / Fig. 7j, "low_q_freq"), duration (Fig. 7h, "high_q_dur" / Fig. 7k, "low_q_dur") and magnitude. The
threshold for high flow (at least 9 times median daily discharge) is chosen according to Clausen and Biggs (2000), and that for
low flow (max. 0.2 times median daily discharge) according to Olden and Poff (2003). The magnitudes of extreme flows are
expressed by the 95th (high flow) and the 5th (low flow) runoff percentiles. Both Q_{95} (Fig. 7i) and Q_5 (Fig. 7l) percentiles exhibit
385 a spatial distribution similar to that of aridity (Fig. 6c, R_{Q95} = -0.62 / R_{Q5} = -0.63). The hydrological indices (Fig. 7) are spatially
less smoothly distributed compared to the climatic indices (Fig. 6). The reasons might be the influence of the (non-) linear
hydrological processes by locally heterogeneous catchment characteristics or uncertainties in runoff measurement (Addor et
al., 2017a; Westerberg et al., 2016).

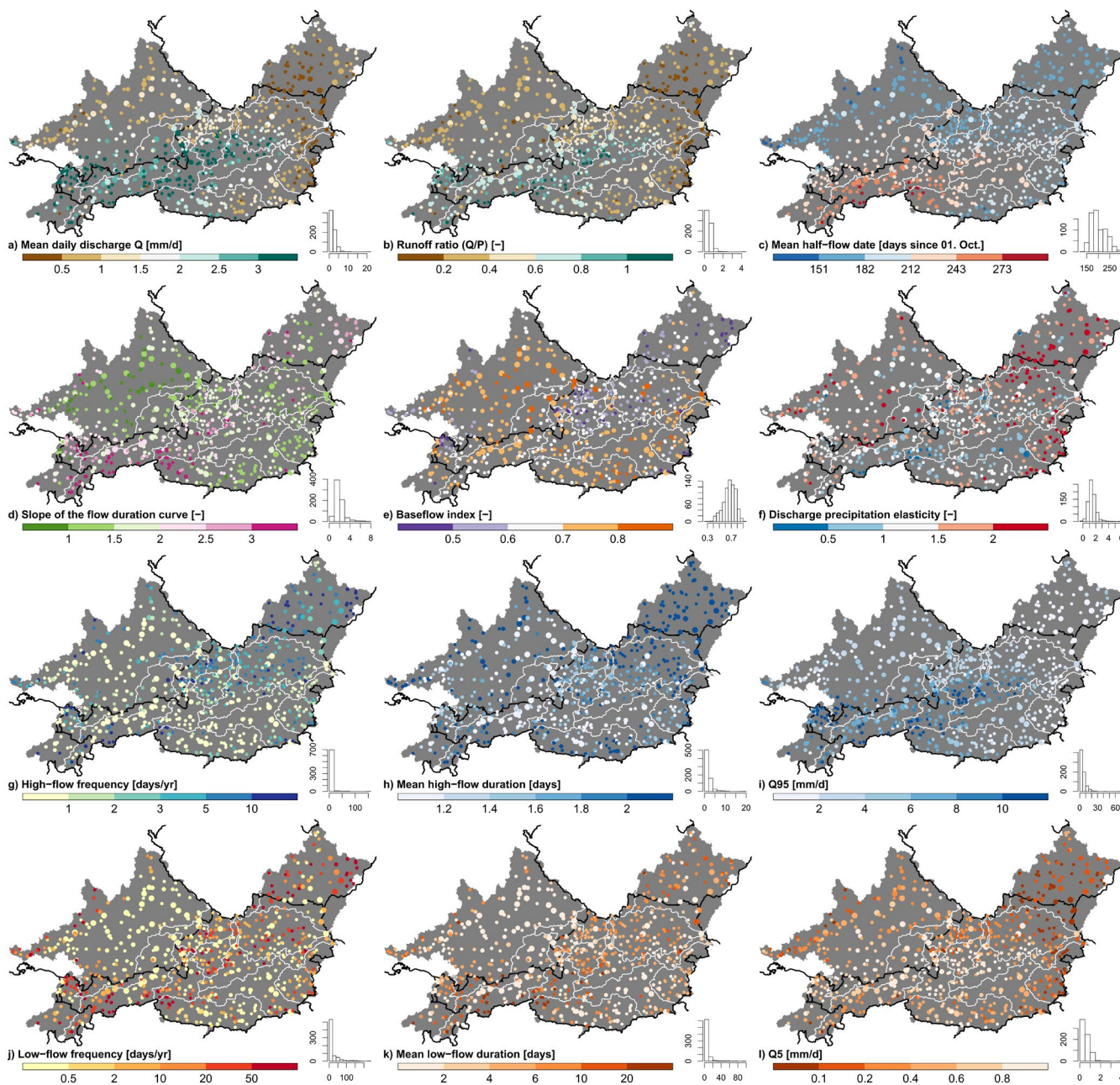



Fig. 7: Spatial distribution of a selection of hydrological indices. Only gauges are plotted, which include timeseries for the period 01. October 1989 to 30. September 2009. The histograms indicate the number of gauges (out of 717) in each category. The size of the circles is proportional to the catchment area.

5.4 Land cover characteristics

All attributes concerning land class (Table A6) are based on the CORINE Land Cover (CLC) 2012 raster dataset featuring a grid-size of 100 m (CORINE, 2012). CORINE is an initiative of the European Environment Agency with the aim to record

land cover of the European territory with a 6-year update cycle. The basic technical specifications like 44 land classes, 25 hectares minimum mapping unit (MMU) for areal phenomena and 100 m minimum width for linear phenomena have not changed since the beginning, facilitating comparisons over the years (CORINE, 2012). It should be noted that an MMU of 25 hectares prevents mapping of very small scaled structures. Other limitations might be the variability of satellite image quality

400 and contents, difficulties for setting up automatic conversion processes, the difference between human interpretation capacity and pixel-based classification (Bossard et al., 2000). However, the total reliability of the predecessor dataset CLC 2000 is 87.0 ± 0.8% according to a reinterpretation approach. The worst class-level reliability (< 70%) was found for sparse vegetation (CLC class 333) (Büttner and Maucha, 2006). The dominant land class within a basin delineation is derived by the majority of the intersecting raster centroids, while the fractions are derived by area share of the specific raster cells.

405 Agricultural land (Fig. 8a, "agr_fra" in Table A6) has high fractions in catchments with low mean slope (Fig. 5d, R = -0.89). The opposite occurs for fraction of bare areas (Fig. 8b, "bare_fra"), since the vegetation period is very short at high elevated terrain and a high terrain slope fosters gravitational erosion processes. Following the CAMELS datasets, no differentiation was made between deciduous and coniferous forests when calculating the forest share. The proportion of forest is highest in the central-eastern region of the project area (Fig. 8c, "forest_fra"), where agriculture and settlement are less prevalent and the

410 mountains are often lower than the forest line. Catchments with a relatively high proportion of glaciers (Fig. 8d, "glac_fra") are mainly located in the western Eastern Alps. The influence of glaciers upon the hydrological regime is primarily apparent in the upper parts of the river regions Inn (region 3 in Fig. 1 and Table 1), Salzach (region 4) and Drava (region 18). High proportions of water surface (Fig. 8e, "lake_fra") can be explained by large lakes, which were mostly formed at the end of the last great ice age about 10 000 years ago (mainly in the Alpine foothills), or by large artificial water reservoirs (mainly in the

415 Czech Republic). Catchments in the Vienna metropolitan area (eastern part of river region 10), as well as in the lower Rhine valley (northern part of river region 1) show quite high fractions of urban area (Fig. 8f, "urban_fra"). However, most catchments (about 74%) have less than 5% urban area.

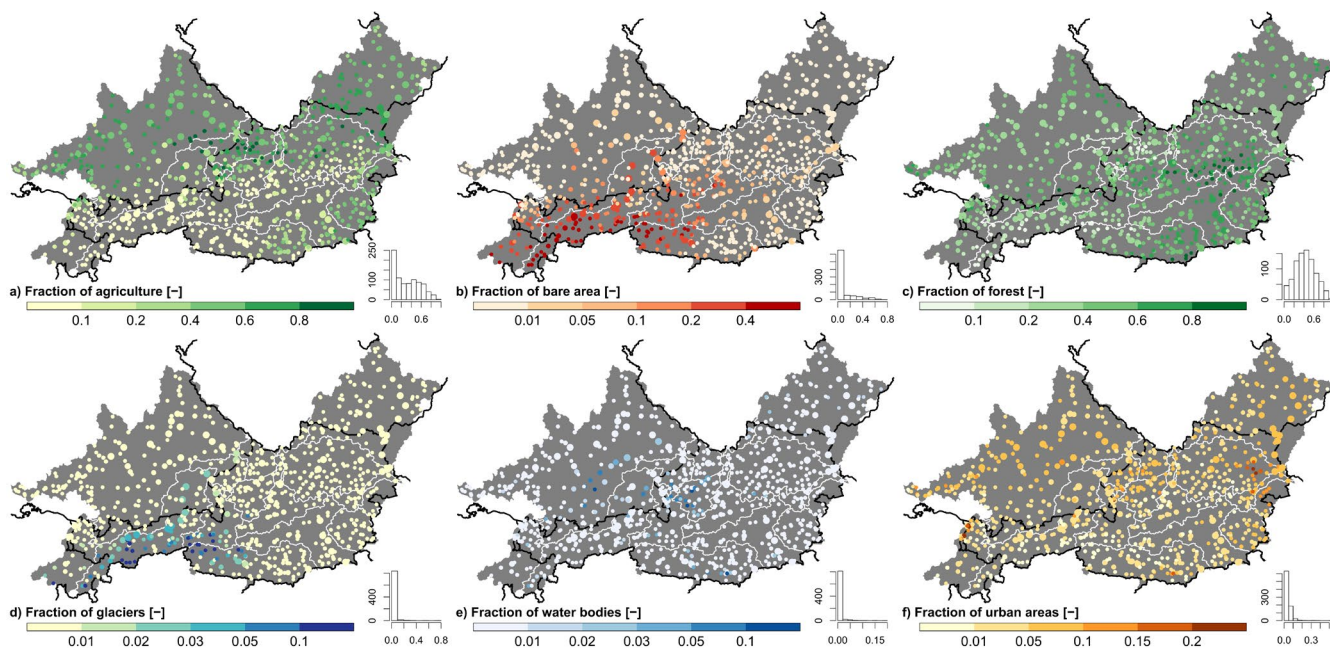



Fig. 8: Spatial distribution of land class fractions representing the characteristics of the entire orographic catchment (catchment
420 **delineation A, Fig. 2a). The histograms indicate the number of basins (out of 859) in each category. The size of the circles is**
proportional to the catchment area.

5.5 Vegetation indices

We calculated 6 different catchment attributes describing vegetation indices, which are based on Leaf Area Index (LAI),
Normalized Difference Vegetation Index (NDVI), and Green Vegetation Fraction (GVF) (Table A7). All vegetation indices
425 are based on long-term monthly means, using either the maximum, minimum, or difference between the maximum and
minimum monthly means (based on 12 monthly means). Processing of the remote sensing datasets was done using the Google
Earth Engine platform (GEEa, 2021; GEEb, 2021; Gorelick et al., 2017).

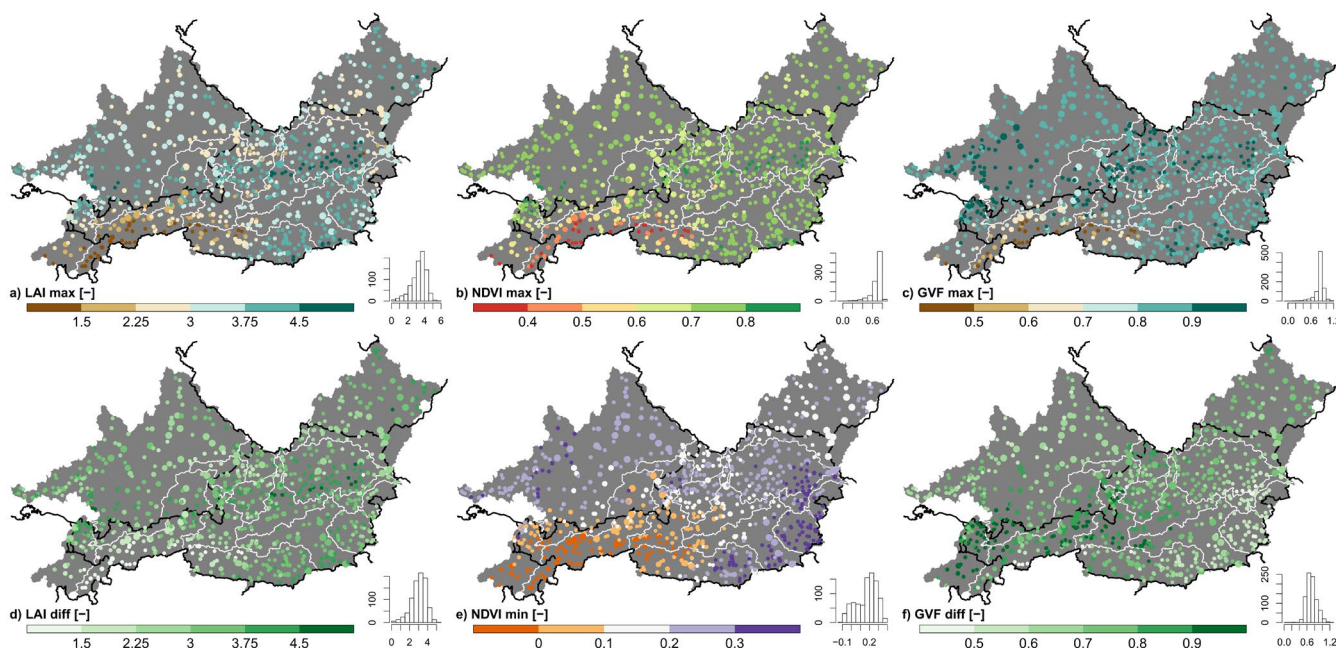
LAI represents vertical vegetation density and is defined as the sum of one-sided green leaf area per unit area for deciduous
forests and half of the total needle area per unit area for coniferous forests. LAI was derived from the "MODIS MCD15A3H"
430 dataset, which is a 4-day composition with 500 m grid resolution (Myneni et al., 2015). The maximum and minimum monthly
means were calculated for the period 01.08.2002 to 01.01.2020 using a cloud filter. The maximum monthly mean of LAI (Fig.
9a, "lai_max" in Table A7) and also the difference between maximum and minimum (Fig. 9d, "lai_diff") show a spatial
correlation with the forest fraction (Fig. 8c, R = 0.76 respectively 0.75). LAI_{diff} shows the same values as LAI_{max} for large parts
of the project area. Especially in regions characterized by a high proportion of coniferous forest the LAI_{diff} should be smaller
435 than LAI_{max} due to the permanent green cover. Snow cover during the winter months could be a possible reason for the non-
representative measurement of the minimum values of LAI.

NDVI is derived from the backscatter of 2 different spectral bands and is widely used for remote sensing-based vegetation
monitoring and classification (horizontal density, type, and physiological condition). The maximum and minimum monthly

NDVI is based on the "MODIS MOD09Q1" dataset with a temporal resolution of 8 days and a spatial resolution of 250 m
440 (Vermote, 2015). The calculation was performed for the period 01.04.2000 to 01.01.2020, also applying a filter on cloudy
satellite images. A negative correlation is apparent between the $NDVI_{max}$ (Fig. 9b, "ndvi_max", R = -0.78) or the $NDVI_{min}$
(Fig. 9e, "ndvi_min", R = -0.84) and the mean catchment elevation (Fig. 5b).

GVF (Green Vegetation Fraction) indicates the fraction of soil that is covered by green vegetation and can be derived from the
NDVI as follows in Eq. (1) (Broxton et al., 2014):

$$GVF = \frac{NDVI - NDVI_s}{NDVI_{c,v} - NDVI_s} \tag{1}$$

445 where NDVI represents the (maximum or minimum) monthly mean of NDVI, $NDVI_s$ the annual maximum NDVI of bare
ground, and $NDVI_{c,v}$ the annual maximum of vegetated ground surface as a function of IGBP land class (Table 1 in Broxton
et al., 2014). $NDVI_s$ was set to 0.09 in accordance to Broxton et al. (2014), while the spatial distribution of the IGBP land
classes was obtained from the "MODIS MCD12Q1" dataset of the year 2012 (Friedl and Sulla-Menashe, 2019). As the values
for $NDVI_s$ and $NDVI_{c,v}$ were derived for a global scale and thus do not necessarily correspond to conditions in the project area,
450 it is possible for GVF values to exceed the normal range between 0 and 1. In order to maintain consistency, we did not constrain
the GVF to the normal range, however. The spatial distribution of GVF_{max} (Fig. 9c, "gvf_max") shows similar spatial patterns
to those of LAI_{max} (R = 0.79) as well as $NDVI_{max}$ (R = 0.94), while GVF_{diff} (Fig. 9f, "gfv_diff") tends to be higher in regions
with a higher fraction of precipitation falling as snow (Fig. 6g).

455 **Fig. 9: Spatial distribution of vegetation indices representing the characteristics of the entire orographic catchment (catchment
delineation A, Fig. 2a). The histograms indicate the number of basins (out of 859) in each category. The size of the circles is
proportional to the catchment area.**

5.6 Soil characteristics

LamaH includes 10 different attributes to characterize soil properties (Table A8), where 8 of them are derived from the 1 km
460 grid sized "European Soil Database Derived Data" (ESDD; Hiederer, 2013a; 2013b). ESDD is based on the "European Soil
Database" (ESD; Panagos et al., 2012; Panagos, 2006), while the maximum available soil water content (TAWC) in ESDD
was calculated using pedotransfer functions (Hiederer, 2013a). ESDD provides soil attributes for a topsoil layer and a subsoil
layer having the boundary at 30 cm soil depth. Values from these two layers were therefore aggregated, weighted by the
available root depth ("root_dep" in Table A8), or in the case of TAWC also summed up. The attribute describing the depth to
465 bedrock "bdrk_dep" is based on the layer "average soil and sedimentary-deposit thickness" of the dataset "Global 1-km Gridded
Thickness of Soil, Regolith, and Sedimentary Deposit Layers" (GGT; Pelletier et al., 2016). GGT has a spatial resolution of
30 arc seconds (approximately 1 km) and is derived from landform-specific models (for upland, lowland, slope, and valley
floor) considering geomorphological principles, and incorporating data for topography, climate, and geology. Calibration and
validation in GGT were performed using independent borehole profiles (Pelletier et al., 2016). The 3D Soil Hydraulic Database
470 of Europe (3DSHD, Toth et al., 2017) dataset with a grid size of 250 m served as source for extracting the saturated hydraulic
soil conductivity ("soil_condu"). 3DSHD was derived using pedotransfer functions (Toth et al., 2015) incorporating attributes
from the SoilGrids250m dataset (SG250; Hengl et al., 2017), while SG250 is based on machine learning techniques including
data from about 150 000 soil profiles as well as remote sensing data for climate, vegetation, geomorphology, and lithology
(Hengl et al., 2017). Data within 3DSHD is provided for 7 different soil layers, so a depth-weighted harmonic averaging was
475 applied.

The provided soil attributes in LamaH may include large uncertainties and should therefore be considered with caution for
several reasons. First, the soil attributes from ESD are mainly based on extrapolated observations of soil profiles and expert
estimates (ESDB, 2004). Especially in the case of heterogeneous soil conditions and large distances between soil profiles the
reliability of ESD dataset must be cautioned. Data from soil profiles are also integrated in 3DSHD (Hengl et al., 2017; Toth et
480 al., 2017) and the dataset of Pelletier et al. (2016), but are rather used for calibration and validation. Toth et al. (2017) indicate
increased unreliability for 3DSHD above 1000 m a.s.l. (about 24.2% of the project area is above 1000 m a.s.l.). Furthermore,
the limitation of the soil depth at 1.5m in ESDD and 2.0 m in 3DSHD is another source of uncertainty (Boer-Euser et al.,
2016). As a last point, it must also be mentioned that much spatially distributed information is lost by aggregation to basin
scale.

485 Depth to bedrock (Fig. 10a, "bedrk_dep" in Table A8) shows similar spatial patterns as mean catchment slope (Fig. 5d, R = -
0.56), and mean elevation (Fig. 5b, R = -0.46). About 37 % of all 859 catchments have a mean depth to bedrock of more than
1.5 m. This depth represents the maximum root-available depth in ESDD (Fig. 10b, "root_dep"). The depth available for roots
tends to be higher in Germany and the Czech Republic than in other regions. If this is an indication of different measurement
methods across the countries is unclear. Low available rooting depths in Austria are, according to Fig. 10b, mainly present
490 where the fraction of carbonate sedimentary rocks (Fig. 11c) or glaciers (Fig. 8d) is high. 40% of all catchments exhibit a mean

organic soil content below 1%, while the highest organic contents are located in the southern German region (Fig. 10c, "oc_fra"). Further interrelationships of the various grain size fractions and the dominating bedrock are recognizable: 1) A high proportion of sand (Fig. 10d, "sand_fra") is especially prevalent where the fraction of metamorphic bedrock is also high (Fig. 11b, R = 0.47). 2) Moreover, the fraction of silt (Fig. 10e, "silt_fra") tends to be high on catchment level where a high fraction

495   of carbonate sedimentary rock (Fig. 11c, R = 0.52) is present. 3) Finally, we can observe an increase in clay content (Fig. 10f, "clay_fra") with increasing proportion of mixed sedimentary rock (Fig. 11d, R = 0.47). Soil porosity (Fig. 10g, "soil_poros") shows similar spatial patterns compared to sand fraction (R = -0.79), while saturated hydraulic conductivity (Fig. 10h, "soil_condu") tends to increase with decreasing mean catchment elevation (Fig. 5b, R = -0.63). The available soil water depth (TAWC, "soil_tawc") was determined in ESDD by including water content at field capacity, gravel content and root-available

500   depth (Hiederer, 2013a). That also explains the high correlation of TAWC (Fig. 10i) with the root-available soil depth (Fig. 10b, R = 0.94).

Earth System
Science
Data

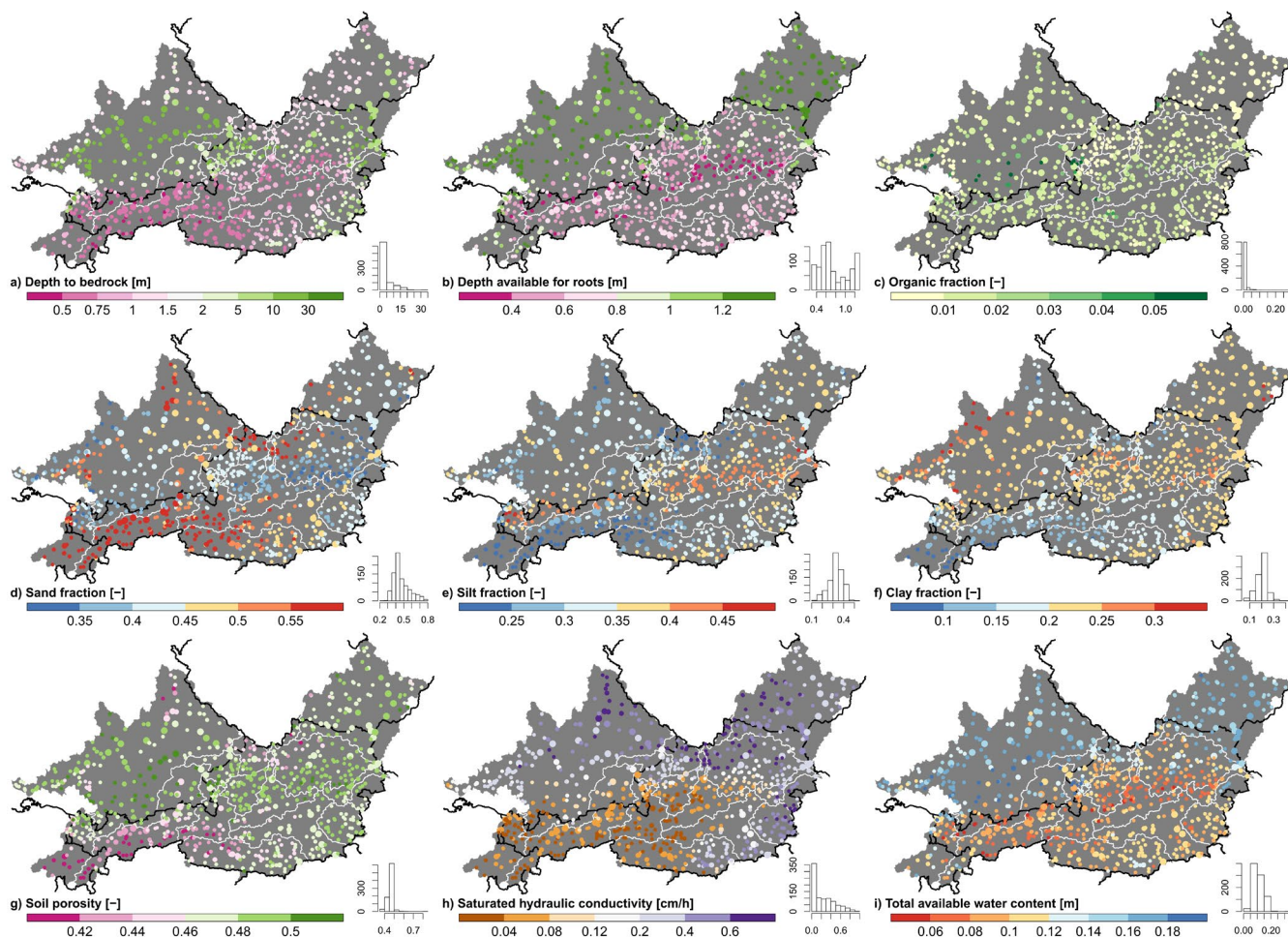



Fig. 10: Spatial distribution of soil attributes representing the characteristics of the entire orographic catchment (catchment delineation A, Fig. 2a). The histograms indicate the number of basins (out of 859) in each category. The size of the circles is proportional to the catchment area.

5.7 Geologic characteristics

We used the datasets GLiM (Hartmann and Moosdorf, 2012; Global Lithological Map) and GLHYMPS (Gleeson et al., 2014; Global Hydrogeology Maps) for deriving 16 different geologic attributes (Table A9). GLiM summarizes 92 regional geological maps in vector form and was used to extract the fractions of the different geological classes. GLiM offers 3 levels of detail, while the 1st level species the dominant lithologic class. The optional 2nd as well as 3rd level further specify, for example, the structure of the rock or local conditions (Hartmann and Moosdorf, 2012). For LamaH only the 1st level of GLiM was used, which contains 16 different geological classes. The classes "evaporites", "no data" and " intermediate volcanic rocks" do not occur within the project area. The 3 most common dominant geologic classes (Fig. 11a, "gc_dom" in Table A9) across all 859 catchments are metamorphites (mt, 35.1%), carbonate sedimentary rocks (sc, 27.4%), and mixed sedimentary rocks (sm, 21.2%). Metamorphic rocks (Fig. 11b, "gc_mt_fra") are predominant along the northern border of the project area (Bohemian

Massif), as well as in the more southern project area (Central Eastern Alps) and include mainly schist, gneiss and quartzite. From a hydrological point of view, the proportion of carbonate sedimentary rock is of particular interest, since a high fraction can be an indicator for karstic systems. High shares of carbonate sedimentary rocks are mainly found along the belt from the southwest to the central east of the project area (Northern Limestone Alps), the central southern border (Southern Limestone

520 Alps) and the north-eastern border (Swabian Alb) (Fig. 11c, "gc_sc_fra"). The flysch and molasse zone (Alpine foothills and central parts of the German project area) is basically characterized by a high fraction of mixed sedimentary rocks (Fig. 11d, "gc_sm_fra").

Attributes concerning permeability and porosity of the lithologic bedrock were extracted from GLHYMPS. There is a high spatial correlation between GLHYMPS and GLiM, as geologic classes of GLiM served as a starting point for assigning

525 hydraulic properties in GLHYMPS. Huscroft et al. (2018) declares that permeability in GLHYMPS is determined only for saturated conditions. GLHYMPS is only intended for regional-scale applications (i.e. spatial resolution greater than 5 km), as the influence of local heterogeneities such as fault zones can be neglected above this scale (Gleeson et al., 2014).

A high proportion of metamorphites or plutonites (mt, pa, pi in Fig. 11a) is commonly associated by low bedrock porosity (Fig. 11e, "geol_poros"). Catchments within the flysch and molasse zones in contrast exhibit relatively high porosity. High

530 bedrock porosity is not necessarily followed by high subsurface permeability ("geol_perme"), yielding a much more inhomogeneous spatial pattern in Fig. 11f than in Fig. 11e. The reason may be rock structure (2nd stage of GLiM), which can have different impacts on permeability and porosity (Table 1 in Gleeson et al., 2014).

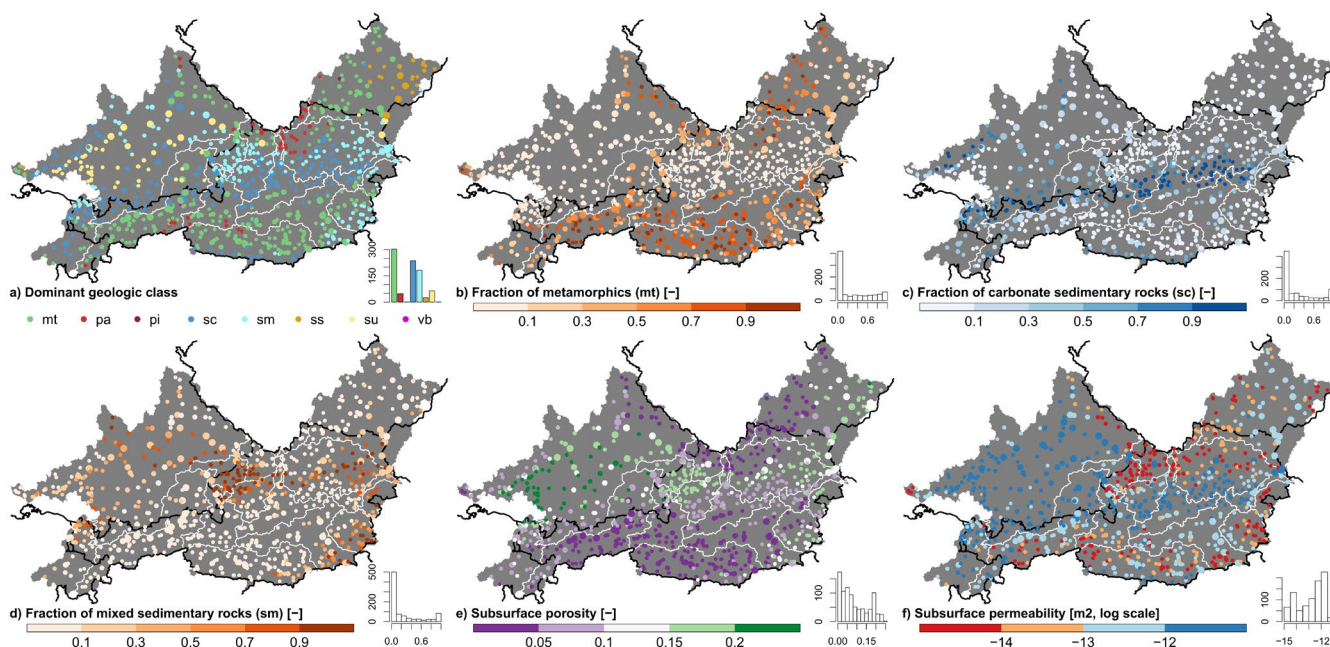



Fig. 11: Spatial distribution of geological attributes representing the characteristics of the entire orographic catchment (catchment
535 **delineation A, Fig. 2a). The histograms indicate the number of basins (out of 859) in each category. The size of the circles is proportional to the catchment area. Classes in plot a): mt – Metamorphites, pa – Acid plutonic rocks, pi – Intermediate plutonic**

rocks, sc – Carbonate sedimentary rocks, sm – Mixed sedimentary rocks, ss – Siliciclastic sedimentary rocks, su – Unconsolidated sediments, vb – Basic volcanic rocks.

5.8 (Anthropogenic) impact on runoff process and measurement

540   We provide 4 attributes (Table 2) in order to simplify filtering and evaluation of runoff gauges regarding any (anthropogenic) impact on runoff processes or its measurement. We have tried to represent the diversity of (human) impact by 13 different types of impact ("typimpact" in Table 2). The type of (human) impact on runoff or measurement was determined primarily from gauge-metadata declared by hydrographic services (BAFU, 2020; GKD, 2020; HZB, 2020; LUBW, 2020). Additionally, publicly available information, as well as manual aerial photo evaluations were used for determination. Typical types of human

545   impact in the project area are large water reservoirs often associated with hydropower plants and cross-basin water transfers. The following types of influence were not classified because the necessary information is not consistently available, or only with great effort: 1) icing, especially at smaller rivers in winter; 2) variable channel profiles leading to inaccurate rating curves; 3) high groundwater flow in the area around the gauge; and 4) subsurface in- or outflows especially in highly karstified areas. The hydrographs with hourly resolution in the months of January and July for the years 1990, 2005 and 2017 were additionally

550   manually evaluated regarding systematic diurnal variations ("diur_art" / "diur_glac"). Systematic fluctuations were further subdivided into those caused artificially (e.g. by storage power plants, power plants with swell operation or sewage treatment plants) and those caused naturally (snow or glacier melt). The degree of gauge impact ("degimpact") is determined using the classes 1) u – "no influence", 2) l – " low influence", 3) m – "moderate influence", 4) s – "strong influence", and 5) x – "not considered further", mostly based on the type of impact and any systematic diurnal variations. A low degree (l) was assigned

555   to those gauges characterized by impact type D (lake with unaffected outlet) or J (herding / vegetation at gauge). Larger (artificial) lakes are in case of flooding effective retention areas, and can therefore attenuate and delay the flood peak. A medium degree of impact (m) was assigned in case of impact type B (flood retention reservoir), C (lake with controllable outlet), F (emergency outlet of water reservoir), G (extreme events are influenced / not properly measured), K (fishing ponds), and L (high infiltration). An exception was made for 3 gauges at the upper Danube, which can be strongly (s) affected by full

560   seepage during the summer months (Hötzl, 1996). Gauges with impact type A (water reservoir with all-season water filling) were assigned in most cases a strong (s) degree of impact, in case of very large catchment area (selected Danube gauges) also a medium degree. An extent-based assignment of the degree of impact was finally made in case of impact type E (water withdrawals) or I (water intake). The hydrographic yearbook of Austria declares anthropogenic cross-basin water transfers by increasing or decreasing the natural catchment area of a gauge (BMLFUW, 2013). Given this information, it was possible to

565   calculate the relative anthropogenic change in catchment size for Austrian gauges, which allows a more comprehensive and objective classification. No influence (u) was assigned, if the relative artificial change of catchment size was less than 1%, while a low degree (l) of impact was noted in case of a change between 1 and 3%. Furthermore, a medium degree of impact (m) was attributed for changes up to 10%, and beyond a strong degree (s). There is no information regarding artificial changes in catchment size for gauges outside Austria. For this reason, the degree of influence was additionally determined there, and

570   also at Austrian gauges influenced by other kinds of water withdrawal (river branches, diversions or irrigation), on the basis
        of publicly available information, as well as aerial photo analyses. We thereby assigned mostly a strong level of impact, but in
        a few cases (e.g. withdrawals for drinking water) also a medium degree. Systematic diurnal variations of artificial origin (e.g.
        swell operation) were classified as "strongly influenced" ("diur_art" in Table 2). Obviously, a gauge or catchment area can be
        characterized by several types of impact. In such cases, the highest degree of impact was chosen. For gauges located in or

575   downstream of urban areas and without any type of influence, we applied a low degree (l) of impact. The reason is a potential
        influence by undetected water withdrawals or stormwater drains. If there is no obvious type of influence and the gauge is
        located above populated areas, the grade of impact was declared as unaffected (u). Finally, the degree of influence (x) was
        attributed for all those gauges, which 1) do not have a clearly assignable catchment area (e.g. gauges at artificial channels
        (impact type H) or below karstic springs), 2) are characterised by several time series (e.g. with or without consideration of mill

580   channels) and 3) have too many gaps (> 50%) in the time series. These gauges were subsequently assigned no meteorological
        time series or catchment characteristics due to lack of catchment delineation.

        The spatial, as well as the frequency distribution of the degree of impact is shown in Fig. 12. 3.5% of 882 gauges are not
        influenced (u), 48% show a low influence (l), 18.9% are moderately influenced (m) and 27% are strongly influenced (s), while
        2.6% belong to class (x). Low-influenced gauges are predominant in the northwest of the German project area, in the north of

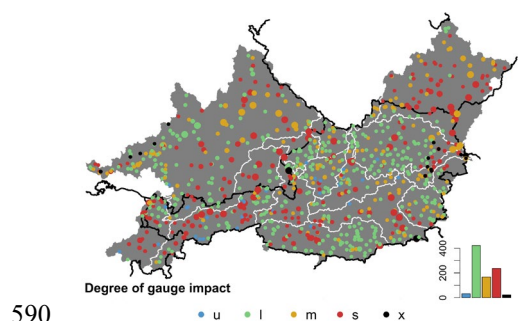
585   the Austrian central region (river region 5, 6, 7, 8, 9, 10 in Fig. 1), but also in the east (river region 16), as well as in the south
        of Austria (east of river region 18). Strongly influenced gauges are in contrast mainly prevalent, where large water reservoirs
        are in operation for hydropower generation (primarily in the Alpine region), but also for seasonal water balancing or flood
        protection (primarily in the Czech Republic and in the north of the German project area). It should be noted that gauges located
        far downstream of large reservoirs may still be strongly influenced by them.

590

Fig. 12: Degree of (anthropogenic) impact on gauges / catchments. The histogram indicates the number of gauges (out of 882) in each category. The size of the circles is proportional to the catchment area. Classes: u – no influence, l – low influence, m – moderate influence, s – strong influence, x – not considered in basin delineation.

Table 2: Attributes for (anthropogenic) gauge and catchment interference.

Attribute	Description	Unit	Data source

typimpact	Type of gauge impact, categorized in 13 classes: A – water reservoir with all-season water filling B – flood retention reservoir C – lake with controllable outlet (e.g. weir) D – lake with unaffected outlet E – water withdrawals F – emergency outlet of water reservoir G – extreme events are influenced / not properly measured H – artificial channel I – water intake (from neighbouring river or catchment) J – herding / vegetation at gauge K – fishing ponds L – high infiltration M – (karstic-) spring	-	BAFU (2020), CHMI (2020), GKD (2020), HZB (2020), LUBW (2020)
diur_art	Unnatural systematic variations in hydrograph with hourly resolution, binary classification (0 - no, 1 - yes)	-	see above
diur_glac	Systematic fluctuations in the hydrograph with hourly resolution, which are caused by glacier melt, binary classification (0 - no, 1 - yes)	-	see above
degimpact	Degree of gauge impact, classes: u – no influence l – low influence m – moderate influence s – strong influence x – not considered in basin delineation (mainly gauges at artificial channels or karstic springs)	-	see above

595 6 Summary and Conclusions

Hydrological studies often require an extensive foundation of data. In large-scale or cross-national projects, it is therefore often laborious and time-consuming to collect the required data and then to homogenize the usually different formats, definitions and conventions. Reasons are for instance the different organizational forms of the hydrographic authorities or communication barriers. LamaH provides a unique, homogeneous data base for hydrological and other environmental sciences, that can

600 overcome the mentioned barriers. Apart from the complete territory of Austria, LamaH also includes all neighboring upstream areas of the rivers flowing through Austria. LamaH contains runoff time series as well as 15 different meteorological time series (daily and hourly resolution) and over 60 attributes for 859 catchments. The 3 different catchment delineations allow investigations for individual (headwater) catchments, but also within an interconnected river network considering intermediate catchments. It is clear that LamaH also contains deficits and uncertainties, also due to the large number of data sources

605 included. These uncertainties have been addressed.

Blöschl et al. (2019b) highlighted numerous open hydrological challenges, such as runoff prediction in unobserved basins (PUB). Methods based on machine learning show promising results for time series prediction (e.g., Kratzert et al., 2019a; 2019b; Kratzert et al., 2018). However, uniformly structured "large-sample" datasets are helpful when applying these data-driven methods, because on the one hand the necessary preparatory work is drastically reduced and on the other hand the

610 exchange or comparability of the modelling results is considerably facilitated. Given the scope of LamaH, we hope that this dataset will serve as a solid data base for further investigations in various fields of hydrology and adjacent fields of environmental science. The high variability in the data in combination with the interconnected river network as well as the high temporal resolution of the time series could grant an improved understanding of processes in water transfer and storage, if appropriate modelling methods are used.

615

Data availability. LamaH is freely available at https://doi.org/10.5281/zenodo.4525244 (Klingler et al., 2021). The dataset is basically divided into 6 parts including basin delineation A/B/C, gauges, stream network, as well as an appendix. The first 4 parts mentioned contain shapefiles, various text files regarding the attributes as well as time series. The stream network is available with shapefiles, which contain numerous attributes, while the nomenclature of the CORINE Land Cover dataset

620 (chapter 5.4) is deposited in the appendix. The entire folder structure, supplementary information regarding the time series, and recommended citations to use are located in the folder "Info". The runoff time series of the German federal states Bavaria and Baden-Württemberg are retrospective checked and updated by the hydrographic services. Therefore, it might be appropriate to obtain more up-to-date runoff data from GKD (2020) or LUBW (2020).

Recommended additional citations when using LamaH. BAFU, 2020; CHMI, 2020; GKD, 2020; LUBW, 2020; BMLFUW,

625 2013; Broxton et al., 2014; CORINE, 2012; EEA, 2019; ESDB, 2004; Farr et al., 2007; Friedl and Sulla-Menashe, 2019; Gleeson et al., 2014; HAO, 2007; Hartmann and Moosdorf, 2012; Hiederer, 2013a; Hiederer, 2013b; Linke et al., 2019; Muñoz Sabater et al., 2021; Muñoz Sabater, 2019a; Muñoz Sabater, 2019b; Myneni et al., 2015; Pelletier et al., 2016; Toth et al., 2017; Trabucco and Zomer, 2019; Vermote 2015.

Author contributions. KS, MH and CK initiated the investigation and designed the study. CK requested the data base,
630 processed the time series and computed the various attributes. The manuscript was prepared by CK with contributions from all co-authors.

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

Financial support. This work was in parts funded by the Austrian Science Fund FWF, project number P 31213.

Acknowledgements. We would like to thank the hydrographic offices from the (federal) states Austria, Baden-Württemberg,
635   Bayern, Czech Republic and Switzerland for providing the runoff timeseries. Data processing was performed using the freely
available software packages R (R Core Team, 2020), Python (Python Software Foundation, 2020), and QGIS (QGIS
Development Team, 2020). Special thanks to all who have developed the numerous open source software, packages and
extensions or who share their experiences in the numerous online forums. LamaH would not have been possible without the
institutions, working groups and individuals who worked, in some cases for several years, on the used open-source datasets.

640   **Review statement.** We would also like to thank Frederik Kratzert, who did a lot of testing and contributed some proposals for
improving LamaH.

References

Addor, N., Do, H. X., Alvarez-Garreton, C., Coxon, G., Fowler, K., and Mendoza, P. A.: Large-sample hydrology: recent
progress, guidelines for new datasets and grand challenges, Hydrolog. Sci. J., 1–14,
645   https://doi.org/10.1080/02626667.2019.1683182, 2019.

Addor, N., Newman, A. J., Mizukami, N., and Clark, M.: The CAMELS data set: catchment attributes and meteorology for
large-sample studies, Hydrology and Earth System Science, 21, 5293–5313, https://doi.org/10.5194/hess-21-5293-
2017, 2017a.

Addor, N.: R scripts for reproducing the climatic and hydrological indices, as well as for creating the maps, GitHub,
650   available at: https://github.com/naddor/camels (last access: 02.03.2020), 2017b.

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop Evapotranspiration. Guidelines for Computing Crop Water
Requirements, FAO Irrigation and Drainage Paper 56, Food and Agriculture Organization (FAO) of the United
Nations, Rome, ISBN 92-5-104219-5, 300pp, 1998.

Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Puelma,
655   C., Cortes, G., Garreaud, R., McPhee, J., and Ayala, A.: The CAMELS-CL dataset: catchment attributes and
meteorology for large sample studies – Chile dataset, Hydrol. Earth Syst. Sci., 22, 5817–5846,
https://doi.org/10.5194/hess-22-5817-2018, 2018.

Arora, V. K.: The use of the aridity index to assess climate change effect on annual runoff, Journal of Hydrol., 265, 164–177,
https://doi.org/10.1016/S0022-1694(02)00101-4, 2002.

660   BAFU: Federal Office for the Environment – Hydrology Division, Bern, Switzerland (runoff data received: 23.09.2020),
2020.

Barnes, W. L., Xiong, X., and Salomonson, V. V.: Status of terra MODIS and aqua modis, Adv. in Space Res., 32(11),
2099–2106, https://doi.org/10.1016/S0273-1177(03)90529-1, 2003.

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., van Dijk, A. I. J. M., McVicar, T. R., and Adler, R. F.:
665   MSWEP V2 Global 3-Hourly 0.1 Precipitation: Methodology and Quantitative Assessment, B. Am. Meteorol. Soc.,
         100(3), 473–500, https://doi.org/10.1175/BAMS-D-17-0138.1, 2019.

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., Brocca, L., Pappenberger, F.,
         Huffman, G. J., and Wood, E. F.: Global-scale evaluation of 22 precipitation datasets using gauge observations and
         hydrological modeling, Hydrol. Earth Syst. Sci., 21, 6201–6217, https://doi.org/10.5194/hess-21-6201-2017, 2017.

670   Beck, H. E., Van Dijk, A. I. J. M., Miralles, D. G., De Jeu, R. A. M., Bruijnzeel, L. A., McVicar, T. R., and Schellekens, J.:
         Global patterns in base flow index and recession based on streamflow observations from 3394 catchments, Water
         Resour. Res., 49, 7843–7863, https://doi.org/10.1002/2013WR013918, 2013.

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, 5638–5661,
675   https://doi.org/10.1002/2014WR015692, 2014.

Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A. P., Parajka, J., et al.: Changing climate both increases and decreases
         European river floods, Nature, 573, 108–111, https://doi.org/10.1038/s41586-019-1495-6, 2019a.

Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., et al.: Twenty-three unsolved problems in
         hydrology (UPH) – a community perspective, Hydrolog. Sci. J., 64, 1141–1158,
680   https://doi.org/10.1080/02626667.2019.1620507, 2019b.

Blöschl, G., Sivapalan, M., Savenije, H., Wagener, T., and Viglione, A. (Eds.): Runoff prediction in ungauged basins:
         synthesis across processes, places and scales, Cambridge University Press, Cambridge, ISBN 9781107028180, 465pp,
         2013.

BMLFUW: Hydrographic Yearbook of Austria 2013, Federal Ministry of Agriculture, Regions and Tourism – Hydrographic
685   Central Office, Vienna, Austria, available at: https://www.bmlrt.gv.at/dam/jcr:bb100102-eac3-4b51-8d8d-
         f40b94e5d546/Jahrbuch_2013_Datenteil.pdf (last access: 16.12.2020), 2013.

Boer-Euser, T. de, McMillan, H. K., Hrachowitz, M., Winsemius, H. C., and Savenije, H. H. G.: Influence of soil and
         climate on root zone storage capacity, Water Resour. Res., 52, 2009–2024, https://doi.org/10.1002/2015WR018115,
         2016.

690   Bossard, M., Feranec, J., and Otahel, J.: CORINE land cover technical guide – Addendum 2000, Technical report No 40,
         European Environment Agency, Copenhagen, Denmark, 105pp, 2000.

Broxton, P. D., Zeng, X., Scheftic, W., and Troch P. A.: A MODIS-Based Global 1-km Maximum Green Vegetation
         Fraction Dataset, Journal of applied Meteorology and Climatology, 53, 1996–2004, https://doi.org/10.1175/JAMC-D-
         13-0356.1, 2014.

695   Budyko, M. I.: Climate and Life; Academic Press: New York, NY, USA, 1974.

Büttner, G. and Maucha, G.: The thematic accuracy of Corine land cover 2000 – Assessment using LUCAS (land use/cover area frame statistical survey), Technical report No 7/2006, European Environment Agency, Copenhagen, Denmark, ISBN 92-9167-844-9, 90pp, 2006.

Chagas, V. B. P., Chaffe, P. L. B., Addor, N., Fan, F. M., Fleischmann A. S., Paiva, R. C. D, and Siqueira V. A.: CAMELS-BR: hydrometeorological time series and landscape attributes for 897 catchments in Brazil, Earth Syst. Sci. Data, 12, 2075–2096, https://doi.org/10.5194/essd-12-2075-2020, 2020.

Chapman, T: A comparison of algorithms for stream flow recession and baseflow separation, Hydrol. Processes, 13(5), 701–714, https://doi.org/10.1002/(SICI)1099-1085(19990415)13:5<701::AID-HYP774>3.0.CO;2-2, 1999.

CHMI: Czech Hydrometeorological Institute, Brno, Czech Republic (runoff data received: 14.12.2020), 2020.

Clausen, B. and Biggs, B. J. F.: Flow variables for ecological studies in temperate streams: groupings based on covariance, J. Hydrol., 237, 184–197, https://doi.org/10.1016/S0022-1694(00)00306-1, 2000.

COPa: European Space Agency & European Commission, Copernicus Program, Copernicus Open Access Hub, available at: https://scihub.copernicus.eu/ (last access: 22.02.2021), 2021.

COPb: European Space Agency & European Commission, Copernicus Program, Copernicus Climate Data Store, available at: https://cds.climate.copernicus.eu/#!/home (last access: 22.02.2021), 2021.

CORINE: CORINE Land Cover 2012, European Environment Agency, Copenhagen, Denmark, available at: https://land.copernicus.eu/pan-european/corine-land-cover (last access: 02.03.2020), 2012.

Court, A.: Measures of streamflow timing, J. Geophys. Res., 67, 4335–4339, https://doi.org/10.1029/JZ067i011p04335, 1962.

Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., Howden, N. J. K., Lane, R., Lewis, M., Robinson, E. L., Wagener, T., and Woods, R.: CAMELS-GB: Hydrometeorological time series and landscape attributes for 671 catchments in Great Britain, Earth Syst. Sci. Data, 12, 2459–2483, https://doi.org/10.5194/essd-12-2459-2020, 2020.

Do, H. X., Gudmundsson, L., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of a daily streamflow archive and metadata, Earth Syst. Sci. Data, 10, 765–785, https://doi.org/10.5194/essd-10-765-2018, 2018.

Döll, P., Douville, H., Güntner, A., Schmied, H. M., and Wada, Y.: Modelling freshwater resources at the global scale: challenges and prospects, Surv. Geophys., 37, 195–221, https://doi.org/10.1007/s10712-015-9343-1, 2016.

Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H. V., Gusev, Y. M., Habets, F., Hall, A., Hay, L., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O. N., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T., and Wood, E. F.: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, Journal of Hydrology, 320, 3–17, https://doi.org/10.1016/j.jhydrol.2005.07.031, 2006.

Eckhardt, K.: A comparison of baseflow indices, which were calculated with seven different baseflow separation methods, J. of Hydrol., 352, 168–173, https://doi.org/10.1016/j.jhydrol.2008.01.005, 2008.

730 EEA: EU-Hydro – River Network Database, Version 1.2, European Environment Agency under the framework of the
     Copernicus program, available at: https://land.copernicus.eu/imagery-in-situ/eu-hydro/eu-hydro-river-network-
     database (last access: 22.10.2020), 2019.

ESA: European Space Agency, Copernicus Program, Sentinel Mission, Sentinel Online, available at:
     https://sentinels.copernicus.eu/web/sentinel/home (last access: 22.02.2021), 2021.

735 ESDB: The European Soil Database distribution version 2.0, European Commission and the European Soil Bureau Network,
     CD-ROM, EUR 19945 EN, 2004.

Falkenmark, M. and Chapman, T.: Comparative hydrology: An ecological approach to land and water resources Unesco,
     UNESCO, Paris, 1989.

Fan, Y., Clark, M., Lawrence, D. M., Swenson, S., Band, L. E., Brantley, S. L., Brooks, P. D., Dietrich, W. E., Flores, A.,
740   Grant, G., Kirchner, J. W., Mackay, D. S., McDonnell, J. J., Milly, P. C. D., Sullivan, P. L., Tague, C., Ajami, H.,
     Chaney, N., Hartmann, A., Hazenberg, P., McNamara, J., Pelletier, J., Perket, J., Rouholahnejad-Freund, E.,
     Wagener, T., Zeng, X., Beighley, E., Buzan, J., Huang, M., Livneh, B., Mohanty, B. P., Nijssen, B., Safeeq, M., Shen,
     C., Verseveld, W., Volk, J., and Yamazaki, D.: Hillslope Hydrology in Global Change Research and Earth System
     Modeling, Water Resour. Res., 55, 1737–1772, https://doi.org/10.1029/2018WR023903, 2019.

745 Fan, Y.: Groundwater in the Earth's critical zone: Relevance to large-scale patterns and processes: Groundwater at large
     scales, Water Resour. Res., 51, 3052–3069, https://doi.org/10.1002/2015WR017037, 2015.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L.,
     Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., and Alsdorf, D. E.: The shuttle
     radar topography mission: Reviews of Geophysics, 45(2), RG2004, available at:
750   https://doi.org/10.1029/2005RG000183, 2007.

Friedl, M. and Sulla-Menashe, D.: MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid
     V006 [Data set], NASA EOSDIS Land Processes DAAC, available at:
     https://doi.org/10.5067/MODIS/MCD12Q1.006 (last access: 12.03.2020), 2019.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A.,
755   and Michaelsen, J.: The climate hazards infrared precipitation with stations-a new environmental record for
     monitoring extremes, Scientific Data, 2, Article Number: 150066 (2015), https://doi.org/10.1038/sdata.2015.66, 2015.

GEEa; Google Earth Engine Platform, available at: https://earthengine.google.com/platform/ (last access: 22.02.2021), 2021.

GEEb: Google Earth Engine Data Catalog, available at: https://developers.google.com/earth-engine/datasets (last access:
     22.01.2021), 2021.

760 Ghiggi, G., Humphrey, V., Seneviratne, S. I., and Gudmundsson, L.: GRUN: an observation-based global gridded runoff
     dataset from 1902 to 2014, Earth Syst. Sci. Data, 11, 1655–1674, https://doi.org/10.5194/essd-11-1655-2019, 2019.

GKD: Bavarian State Office for the Environment – Hydrographic Service, Munich, Germany, available at:
     https://www.gkd.bayern.de/en/rivers/discharge/tables (runoff data downloaded: 15.09.2020), 2020.

Gleeson, T., Moosdorf, N., Hartmann, J., and van Beek, L. P. H.: A glimpse beneath earth's surface: GLobal HYdrogeology
765   MaPS (GLHYMPS) of permeability and porosity, Geophys. Res. Lett., 41, 3891–3898,
       https://doi.org/10.1002/2014GL059856, available at:
       https://dataverse.scholarsportal.info/dataset.xhtml?persistentId=doi:10.5683/SP2/DLGXYO (last access: 12.03.2020),
       2014.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R.: Google Earth Engine: Planetary-scale
770   geospatial analysis for everyone, Remote Sensing of Environment, 202, 18–27,
       https://doi.org/10.1016/j.rse.2017.06.031, 2017.

Gudmundsson, L., Leonard, M., Do, H. X., Westra, S., and Seneviratne, S. I.: Observed Trends in Global Indicators of Mean
       and Extreme Streamflow, Geophys. Res. Lett., 46, 756–766, https://doi.org/10.1029/2018GL079725, 2019.

Gudmundsson, L., Do, H. X., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM) –
775   Part 2: Quality control, time-series indices and homogeneity assessment, Earth Syst. Sci. Data, 10, 787–804,
       https://doi.org/10.5194/essd-10-787-2018, 2018.

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, 463–477, https://doi.org/10.5194/hess-18-463-2014,
       2014.

780 HAO: Hydrological Atlas of Austria (digHAO), 3. Delivery, Federal Ministry of Agriculture, Regions and Tourism –
       Hydrographic Central Office, Vienna, Austria, 2007.

Hargreaves, G. H.: Defining and Using Reference Evapotranspiration, Journal of Irrigation and Drainage Engineering,
       120(6), 1132–1139, https://doi.org/10.1061/(ASCE)0733-9437(1994)120:6(1132), 1994.

Hartmann, J. and Moosdorf, N.: The new global lithological map database GLiM: A representation of rock properties at the
785   Earth surface, Geochem. Geophy. Geosy., 13, 1–37, https://doi.org/10.1029/2012GC004370, 2012.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., et al.: SoilGrids250m: Global
       gridded soil information based on machine learning, edited by: Bond-Lamberty, B., PLoS ONE, 12, e0169748,
       https://doi.org/10.1371/journal.pone.0169748, 2017.

Hennermann, K. and Guillory, A.: ERA5: uncertainty estimation, CDS dataset documentation, European Centre for
790   Medium-Range Weather Forecasts (ECMWF), available at:
       https://confluence.ecmwf.int/display/CKB/ERA5%3A+uncertainty+estimation (last access: 30.11.2020), 2020.

Herrnegger, M., Senoner, T., and Nachtnebel, H. P.: Adjustment of spatio-temporal precipitation patterns in a high Alpine
       environment, J. Hydrol., 556, 913–921, https://doi.org/10.1016/j.jhydrol.2016.04.068, 2018.

Herrnegger, M., Nachtnebel, H. P., and Haiden, T.: Evapotranspiration in high alpine catchments – an important part of the
795   water balance!, Hydrol. Res., 43(4), 460–475, https://doi.org/10.2166/nh.2012.132, 2012.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al.: The ERA5 global reanalysis,
       Quarterly Journal of the Royal Meteorol. Soc., 146(730), 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hiederer, R.: Mapping Soil Properties for Europe - Spatial Representation of Soil Database Attributes, Luxembourg: Publications Office of the European Union, 47pp, EUR26082EN Scientific and Technical Research series, ISSN 1831- 9424, https://doi.org/10.2788/94128, 2013a.

Hiederer, R.: Mapping Soil Typologies - Spatial Decision Support Applied to European Soil Database, Luxembourg: Publications Office of the European Union, 147pp, EUR25932EN Scientific and Technical Research series, ISSN 1831–9424, https://doi.org/doi:10.2788/8728, 2013b.

Horn, B. K. P.: Hill shading and the reflectance map, Proceedings of the IEEE, 69(1), 14–47, https://doi.org/10.1109/PROC.1981.11918, 1981.

Hötzl, H.: Origin of the Danube-Aach system, Environmental Geology, 27, 87–96, https://doi.org/10.1007/BF01061676, 1996.

Huscroft, J., Gleeson, T., Hartmann, J., and Börker, J.: Compiling and Mapping Global Permeability of the Unconsolidated and Consolidated Earth: GLobal HYdrogeology MaPS 2.0 (GLHYMPS 2.0), Geophys. Res. Lett., 45, 1897–1904, https://doi.org/10.1002/2017GL075860, 2018.

HZB: Federal Ministry of Agriculture, Regions and Tourism – Hydrographic Central Office, Vienna, Austria (runoff data received: 08.09.2020), 2020.

ICPDR: Danube Basin Facts & Figures, available at: https://www.icpdr.org/flowpaper/viewer/default/files/nodes/documents/icpdr_facts_figures.pdf (last access: 21.09.2020), 2020.

Irons, J. R., Dwyer, J. L., and Barsi, J. A.: The next Landsat satellite: The Landsat Data Continuity Mission, Remote Sensing of Environment, 122, 11–21, https://doi.org/10.1016/j.rse.2011.08.026, 2012.

Klingler, C., Kratzert, F., Schulz, K., and Herrnegger, M.: LamaH | Large-Sample Data for Hydrology and Environmental Sciences in Central Europe – link to files, Zenodo, https://doi.org/10.5281/zenodo.4525244, 2021.

Klingler, C., Bernhardt, M., Wesemann, J., Schulz, K., and Herrnegger, M.: Lokale hydrologische Modellierung mit globalen, alternativen Datensätzen, Hydrologie & Wasserbewirtschaftung, 64(4), 166–187, https://doi.org/10.5675/HyWa_2020.4_1, 2020.

Kobolschnig, G. R. and Schöner, W.: The relevance of glacier melt in the water cycle of the Alps: the example of Austria, Hydrol. Earth Syst. Sci., 15, 2039–2048, https://doi.org/10.5194/hess-15-2039-2011, 2011.

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., and Nearing, G.: Toward improved predictions in ungauged basins: Exploiting the power of machine learning, Water Resources Research, 55, 11344–11354, https://doi.org/10.1029/2019WR026065, 2019a.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., and Nearing, G.: Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets, Hydrol. Earth Syst. Sci., 23, 5089–5110, https://doi.org/10.5194/hess-23-5089-2019, 2019b.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrol. Earth Syst. Sci., 22, 6005–6022, https://www.hydrol-earth-syst-sci.net/22/6005/2018, 2018.

Kuhn, M.: The Reaction of Austrian Glaciers and their Runoff to Changes in Temperature and Precipitation Levels,
835   Österreichische Wasser- und Abfallwirtschaft, 56(1–2), 1–7, 2004.

Kuenz, A., Arheimer, B., Hundecha, Y., and Wagener, T.: Understanding hydrologic variability across Europe through catchment classification, Hydrol. Earth Syst. Sci., 21, 2863–2879, https://doi.org/10.5194/hess-21-2863-2017, 2017.

Ladson, A., Brown, R., Neal, B., and Nathan, R.: A standard approach to baseflow separation using the Lyne and Hollick filter, Australian Journal of Water Resources, 17, 25–34, https://doi.org/10.7158/13241583.2013.11465417, 2013.

840   Lambrecht, A. and Kuhn, M.: Glacier changes in the Austrian Alps during the last three decades, derived from the new Austrian glacier inventory, Ann. Glaciol., 46, 177–184, https://doi.org/10.3189/172756407782871341, 2007.

Lehner, B., Verdin, K., and Jarvis, A.: New global hydrography derived from spaceborne elevation data, Eos, Transactions, AGU, 89(10), 93–94, https://doi.org/10.1029/2008EO100001, 2008.

Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S.,
845   Moidu, H., Tan, F., and Thieme, M.: Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution, Scientific Data, 6, 283, https://doi.org/10.1038/s41597-019-0300-6, 2019.

LUBW: State Agency for the Environment Baden-Württemberg – Hydrographic Service, Karlsruhe, Germany, available at: http://udo.lubw.baden-wuerttemberg.de/public/p/pegel_messwerte_leer (runoff data received: 04.09.2020), 2020.

Luke, A., Vrugt, J. A., AghaKouchak, A., Matthew, R., and Sanders, B. F.: Predicting nonstationary flood frequencies:
850   Evidence supports an updated stationarity thesis in the United States, Water Resour. Res., 53, 5469–5494, https://doi.org/10.1002/2016WR019676, 2017.

Malenovsky, Z, Rott, H., Cihlar, J., Schaepman, M. E., Garcia-Santos, G., Fernandes, R., and Berger, M.: Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land, Remote Sensing of Environment, 120, 91–101, https://doi.org/10.1016/j.rse.2011.09.026, 2012.

855   Muñoz Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J. N.: ERA5-Land: A state-of-the-art global reanalysis dataset for land applications, ESSD discussion, https://doi.org/10.5194/essd-2021-82, 2021.

Muñoz Sabater, J.: ERA5-Land hourly data from 1981 to present, Copernicus Climate Change Service (C3S) Climate Data
860   Store (CDS), available at: https://doi.org/10.24381/cds.e2161bac (last access: 22.10.2020), 2019a.

Muñoz Sabater, J.: First ERA5-Land dataset to be released this spring, ECMWF newsletter, number 159 - spring 2019, available at: https://www.ecmwf.int/en/newsletter/159/news/first-era5-land-dataset-be-released-spring (last access: 30.11.2020), 2019b.

Muñoz Sabater, J., Dutra, E., Balsamo, G., Hersbach, H., Boussetta, S., Dee, D., and Hirahara, S.: ERA5-Land: A new state-
865   of-the-art Global Land Surface Reanalysis Dataset, 31st Conference on Hydrology – 2017 AMS annual meeting,
         Seattle, US, 2017.

Myneni, R., Knyazikhin, Y., and Park, T.: MCD15A3H MODIS/Terra+Aqua Leaf Area Index/FPAR 4-day L4 Global 500m
         SIN Grid V006 [Data set], NASA EOSDIS Land Processes DAAC, available at:
         https://doi.org/10.5067/MODIS/MCD15A3H.006 (last access: 12.03.2020), 2015.

870   NASAa: National Aeronautics and Space Administration (NASA), Landsat Science, available at:
         https://landsat.gsfc.nasa.gov/ (last access: 22.02.2021), 2021.

NASAb: National Aeronautics and Space Administration (NASA), MODIS Mission, available at:
         https://modis.gsfc.nasa.gov/ (last access: 22.02.2021), 2021.

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., and Gupta, H. V.: What
875   Role Does Hydrological Science Play in the Age of Machine Learning?, Water Resour. Res., e2020WR028091,
         https://doi.org/10.1029/2020wr028091, 2020.

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,
880   Hydrol. Earth Syst. Sci., 19, 209–223, https://doi.org/10.5194/hess-19-209-2015, 2015.

Panagos, P., Van Liedekerke, M., Jones, A., and Montanarella L.: European Soil Data Centre: Response to European policy
         support and public data requirements, Land Use Policy, 29(2), 329–338,
         https://doi.org/10.1016/j.landusepol.2011.07.003, 2012.

Panagos, P.: The European soil database, GEO: connexion, 5(7), 32–33, 2006.

885   Pelletier, J. D., Broxton, P. D., Hazenberg, P., Zeng, X., Troch, P. A., Niu, G., Williams, Z. C., Brunke, M. A., and Gochis,
         D.: Global 1-km Gridded Thickness of Soil, Regolith, and Sedimentary Deposit Layers. ORNL DAAC, Oak Ridge,
         Tennessee, USA, https://doi.org/10.3334/ORNLDAAC/1304, 2016.

Prohaska, S., Brilly, M., and Kryžanowski, A.: Cooperation of hydrologists from the Danube River Basin, Hydrol. Earth
         Syst. Sci. Discuss., 1–10, https://doi.org/10.5194/hess-2020-66, 2020.

890   Python Software Foundation: Python Language Reference, available at https://www.python.org, 2020.

QGIS Development Team: QGIS Geographic Information System, Open Source Geospatial Foundation Project, available at:
         http://qgis.osgeo.org, 2020.

Olden, J. D. and Poff, N. L.: Redundancy and the choice of hydrologic indices for characterizing streamflow regimes, River
         Res. Appl., 19, 101–121, https://doi.org/10.1002/rra.700, 2003.

895   Oerlemans, J., Anderson, B., Hubbard, A., Huybrechts, P., Johannesson, T., Knap, W. H., Schmeits, M., Stroeven, A. P., van
         de Wal, R. S. W., and Wallinga, J.: Modelling the response of glaciers to climate warming, Clim. Dynam., 14/4, 267–
         274, https://doi.org/10.1007/s003820050222, 1998.

R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, available at: https://www.r-project.org, 2020.

900 Sankarasubramanian, A., Vogel, R. M., and Limbrunner, J. F.: Climate elasticity of streamflow in the United States, Water Resour. Res., 37, 1771–1781, https://doi.org/10.1029/2000WR900330, 2001.

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

905 Schaake, J. C., Hamill, T. M., Buizza, R., and Clark, M.: HEPEX: the hydrological ensemble prediction experiment, Bulletin of the American Meteorological Society, 88(10), 1541–1548, https://doi.org/10.1175/BAMS-88-10-1541, 2007.

Schumm, S. A.: Evolution of drainage systems and slopes in Badlands at Perth Amboy, New Jersey, GSA Bulletin, 67(5), 597–646, https://doi.org/10.1130/0016-7606(1956)67[597:EODSAS]2.0.CO;2, 1956.

Singh, R., Archfield, S. A., and Wagener, T.: Identifying dominant controls on hydrologic parameter transfer from gauged to

910 ungauged catchments – A comparative hydrology approach, J. Hydrol., 517, 985–996, https://doi.org/10.1016/j.jhydrol.2014.06.030, 2014.

Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., and Demir, I.: A comprehensive review of deep learning applications in hydrology and water resources, Water Sci. Technol., 82(12), 2635–2670, https://doi.org/10.2166/wst.2020.369, 2020.

915 Smith, M. B., Seo, D. J., Koren, V. I., Reed, S. M., Zhang, Z., Duan, Q., Moreda, F., and Cong, S.: The Distributed Model Intercomparison Project (DMIP): motivation and experiment design, J. Hydrol., 298, 4–26, https://doi.org/10.1016/j.jhydrol.2004.03.040, 2004.

Tallaksen, L. and Van Lanen, H. A. J.: Hydrological drought. Processes and estimation methods for streamflow and groundwater, Developments in Water Science, 48, ISSN: 0167-5648, Amsterdam: Elsevier.

920 Toth, B., Weynants, M., Pasztor, L., and Hengl, T.: 3D soil hydraulic database of Europe at 250 m resolution, Hydrological Processes, 31, 2662–2666, https://doi.org/10.1002/hyp.11203, 2017.

Toth, B., Weynants, M., Nemes, A., Makó, A., Bilas, G., and Toth, G.: New generation of hydraulic pedotransfer functions for Europe, European Journal of Soil Science, 66, 226–238, https://doi.org/10.1111/ejss.12192, 2015.

Trabucco, A. and Zomer, R.: Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2, CGIAR

925 Consortium for Spatial Information (CGIAR-CSI), published online, available at: https://doi.org/10.6084/m9.figshare.7504448.v3 (last access: 18.03.2020), 2019.

TYROL: Catalog Water Network Tyrol, Government of the Austrian federal state Tyrol, Innsbruck, available at: https://www.data.gv.at/katalog (downloaded: 17.10.2020), 2020.

Van Lanen, H. A. J., Wanders, N., Tallaksen, L. M., and Van Loon, A. F.: Hydrological drought across the world: impact of

930 climate and physical catchment structure, Hydrol. Earth Syst. Sci., 17, 1715–1732, https://doi.org/10.5194/hess-17-1715-2013, 2013.

Vermote, E.: MOD09Q1 MODIS/Terra Surface Reflectance 8-Day L3 Global 250m SIN Grid V006 [Data set], NASA
    EOSDIS Land Processes DAAC, available at: https://doi.org/10.5067/MODIS/MOD09Q1.006 (last access:
    12.03.2020), 2015.

935   Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., and Schewe, J.: The inter-sectoral impact model
    intercomparison project (ISI–MIP): project framework, Proceedings of the National Academy of Sciences, 111(9),
    3228–3232, https://doi.org/10.1073/pnas.1312330110, 2014.

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, 1847–1865,

940   https://doi.org/10.1002/2015WR017635, 2016.

Westerberg, I. K. and McMillan, H. K.: Uncertainty in hydrological signatures, Hydrol. Earth Syst. Sci., 19, 3951–3968,
    https://doi.org/10.5194/hess-19-3951-2015, 2015.

WGMS (World Glacier Monitoring Service), Glacier Mass Balance Bulletin, No. 8 (2002–2003), edited by: Haeberli, W.,
    Noetzli, J., Zemp, M., Baumann, S., Frauenfelder, R., Hoelzle, M., Department of Geography, University of Zürich,

945   100pp, 2005.

Woods, R. A.: Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks, Adv. Water
    Resour., 32, 1465–1481, https://doi.org/10.1016/j.advwatres.2009.06.011, 2009.

Yang, X. and Giusti, M.: ERA5-Land: data documentation, CDS dataset documentation, European Centre for Medium-
    Range Weather Forecasts (ECMWF), available at: https://confluence.ecmwf.int/display/CKB/ERA5-

950   Land%3A+data+documentation (last access: 30.11.2020), 2020.

Yokoo, Y. and Sivapalan, M.: Towards reconstruction of the flow duration curve: development of a conceptual framework
    with a physical basis, Hydrol. Earth Syst. Sci., 15, 2805–2819, https://doi.org/10.5194/hess-15-2805-2011, 2011.

Appendix A

Table A1: Gauge referred attributes.

Attribute	Description	Unit	References
ID	ID number, the same ID number is also assigned to the corresponding basin for all aggregation modes, so this variable enables the link between runoff time series, meteorological time series and the catchment attributes	-	-
govnr	Official gauge number from the associated governments	-	BAFU (2020), CHMI (2020), GKD (2020), HZB (2020), LUBW (2020)

name	Official name of the runoff gauge	-	see above
river	Name of the belonging river	-	see above
area_gov	Catchment area obtained from the administration	km²	see above
elev	Elevation of the gauge`s zero point	m a.s.l.	see above
lon	Longitude in LAEA Europe grid (EPSG 3035)	m	see above
lat	Latitude in LAEA Europe (EPSG 3035)	m	see above
country	ISO 3166 ALPHA-3 code for country	-	see above
fedstate	Abbreviation for federal state, water management administration is executed in the given countries at the stage of federal states [a]	-	see above
region	River-based region [b]	-	see above
obsbeg_day	Start of daily runoff time series in the dataset	year	see above
obsbeg_hr	Begin of continuous (hourly) runoff data recording	year	see above
obsend	End of continuous (hourly) runoff data recording, number 0 indicates an up-to-date recording, daily time series might last longer	year	see above
gaps	Fraction of gaps in the hourly runoff timeseries over the whole time series, gaps are indicated by the value -999 in the time series	‰	see above
area_ratio	Ratio between variable "area_calc" of basin delineation A (Table A3) and "area_gov"	-	BAFU (2020), CHMI (2020), GKD (2020), HZB (2020), LUBW (2020), HAO (2007), HydroATLAS (Linke et al., 2019)
HIERARCHY	Gauge hierarchy [c]	-	see above
NEXTUPID	ID of the next upstream gauges (can be one or more), 0 indicates no upstream gauges [c]	-	see above
NEXTDOWNID	ID of the next downstream gauge (only one), 0 indicates no downstream gauge [c]	-	see above
dist_hup	Horizontal stream length from the most distant beginning of a watercourse within the basin to the gauge [d]	km	BAFU (2020), CHMI (2020), GKD (2020), HZB (2020), LUBW (2020), EU-Hydro - River Network Database (EEA, 2019)

dist_hdn	Horizontal stream length from the actual gauge to the next downstream gauge [c]	km	see above
elev_diff	Elevation difference from the actual`s gauge zero point to the next downstream`s gauge zero point [c]	m	see above
strm_slope	Slope of the actual gauge to the next downstream`s gauge, fraction of "elev_diff" and "dist_hdn" [c]	m km^{-1}	see above
checked	Declares if a timestep in the runoff time series was checked by the staff of the hydrological services or not, the specific scope of the check is not described [e]	TRUE or FALSE	BAFU (2020), CHMI (2020), GKD (2020), HZB (2020), LUBW (2020)

955 [a] List of abbrev. for attribute fedstate: Austria (BLD – Burgenland, CRN – Carinthia, LAT – Lower Austria, SBG – Salzburg, STY – Styria, TYR – Tyrol, UAT – Upper Austria, VBG – Vorarlberg, VIE – Vienna); Germany (BAV – Bavaria, BWT – Baden-Württemberg); Switzerland (GRI – Grisons, STG – Saint Gallen); Liechtenstein (LIE - Liechtenstein); Czech Republic (OLM – Olomouc, SBO – South Bohemian, SMO – South Moravian, VYS – Vysočina, ZLN - Zlin). [b] List of abbrev. for attribute region: 1 – Rhine, 2 – Danube above Inn, 3 – Inn above Salzach, 4 – Salzach, 5 – Inn under Salzach, 6 – Danube

960 between Inn and Traun, 7 – Traun, 8 – Danube between Traun and Enns, 9 – Enns, 10 – Danube between Enns and Morava, 11 – Vltava, 12 – Morava, 13 – Danube between Morava and Leitha, 14 – Leitha, 15 – Rabniz, 16 – Raab, 17 – Mur, 18 – Drava. [c] Only for basin delineation B and C. [d] Only for basin delineation A. [e] Visible in the daily and hourly runoff time series.

Table A2: Meteorological variables from ERA5L dataset (Muñoz Sabater, 2019a).

Variable hourly	Daily aggregation	Description	Unit
DOY	unchanged	Day of year	-
HOD	omitted	Hour of day	-
2m_temp	max, mean, min	Air temperature at a height of 2 m above Earth surface	°C
2m_dp_temp	max, mean, min	Dewpoint temperature at a height of 2 m above Earth surface	°C
10m_wind_u	mean	Horizontal speed of air moving towards the east at a height of 10 m above Earth surface	m s^{-1}
10m_wind_v	mean	horizontal speed of air moving towards the north at a height of 10 m above Earth surface	m s^{-1}
fcst_alb	mean	Forecast albedo, fraction of solar (shortwave) radiation reflected by Earth's surface (direct and diffuse)	-
lai_high_veg	mean	One-half of the total green leaf area per unit horizontal ground surface area for high vegetation type	-

lai_low_veg	mean	One-half of the total green leaf area per unit horizontal ground surface area for low vegetation type	-
swe	mean	Water equivalent of snow	mm
surf_net_solar_rad	max, mean	Amount of solar radiation (shortwave radiation) reaching the Earth`s surface (direct and diffuse) minus the amount reflected by the Earth's surface (governed by albedo), positive sign is indicator for radiation to the Earth	W m^{-2}
surf_net_therm_rad	max, mean	Net thermal radiation at the Earth´s surface, positive sign is indicator for radiation from the Earth	W m^{-2}
surf_press	mean	Surface pressure	Pa
total_et	sum	Total evapotranspiration, positive values indicate evapotranspiration, negative values condensation	mm
prec	sum	Total amount of precipitation (liquid and frozen)	mm
volsw_123	mean	Fraction of water in top soil layer, 0 to 100 cm depth	m^3 m^{-3}
volsw_4	mean	Fraction of water in sub soil layer, 100 to 289 cm depth	m^3 m^{-3}

965

Table A3: Topographic indices.

Attribute	Description	Unit	Data source
area_calc	Calculated basin area	km^2	HAO (2007), HydroATLAS (Linke et al., 2019)
elev_mean	Mean catchment elevation [a]	m a.s.l.	NASA JPL SRTMGL1 V3 Digital Elevation 30m (Farr et al., 2007)
elev_med	Median catchment elevation [a]	m a.s.l.	see above
elev_std	Standard deviation of elevation in catchment [a]	m a.s.l.	see above
elev_ran	Range of catchment elevation (max. – min. elev.) [a]	m a.s.l.	see above
slope_mean	Mean catchment slope; Horn (1981) [a]	m km^{-1}	see above
mvert_dist	Horizontal distance from the farthest point of the catchment to the belonging gauge (length axis)	km	HAO (2007), HydroATLAS (Linke et al., 2019)
mvert_ang	Angle between North direction and connection from farthest point of catchment to belonging gauge (length axis), e.g. direction from north (farthest catchment point) to south (gauge): 180°, direction from east to west: 270°	degree	see above

elon_ratio	Elongation ratio R_e after Schumm (1956), ratio of the diameter D of an equivalent circle with the area of the catchment area to its length L (mvert_dist), $R_e = \frac{1}{L} \times \sqrt{\frac{4 \times A}{\pi}} = \frac{D}{L}$	-	see above
strm_dens	Stream density D_F, ratio of lengths of streams L_F to the catchment area A (area_calc), $D_F = \frac{\sum L_F}{A}$	km km^{-2}	EU-Hydro - River Network Database (EEA, 2019), HAO (2007), HydroATLAS (Linke et al., 2019)

[a] Upscaling approach: 1.

Table A4: Climatic indices.

Attribute	Description	Unit	Data source
p_mean	Mean daily precipitation [e]	mm day^{-1}	ERA5L (Muñoz Sabater, 2019a) [a]
et0_mean	Mean daily reference evapotranspiration ET0 [d]	mm day^{-1}	Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2 (Trabucco and Zomer, 2019) [b]
eta_mean	Mean daily total evapotranspiration [e]	mm day^{-1}	ERA5L (Muñoz Sabater, 2019a) [a]
aridity	Aridity, computed as the ratio of mean ET0 to mean precipitation [d,e]	-	ERA5L (Muñoz Sabater, 2019a) [a], Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2 (Trabucco and Zomer, 2019) [b]
p_season	Seasonality and timing of precipitation (estimated using sine curves) to represent the annual precipitation cycles, positive (negative) values indicate that precipitation sums are higher during summer (winter) months, values close to 0 indicate uniform precipitation throughout the year; Eq. (14) in Woods et al. (2009) [e]	-	see above
frac_snow	Fraction of precipitation falling as snow, i.e. falling on days with mean temperature below 0 °C [e]	-	see above

	separation is performed using the Ladson et al. (2013) digital filter		
baseflow_index _lfstat	equal than above, expect hydrograph separation is performed using a package from Tallaksen and Van Lanen (2004)	-	see above
hfd_mean	Mean half-flow date, i.e. number of days since the beginning of the hydrological year (1. October) on which the accumulated runoff reaches half of the annual volume; Court (1962)	days since 01. Oct.	see above
Q5	5% flow quantile (low flow)	mm day^{-1}	see above
Q95	95% flow quantile (high flow)	mm day^{-1}	see above
high_q_freq	Frequency of high-flow days (> 9 times median daily flow); Clausen and Biggs (2000), Table 2 in Westerberg and McMillan (2015)	days yr^{-1}	see above
high_q_dur	Mean duration of high-flow events (number of consecutive days > 9 times median daily flow); Clausen and Biggs (2000), Table 2 in Westerberg and McMillan (2015)	days	see above
low_q_freq	Frequency of low-flow days (< 0.2 times mean daily flow); Olden and Poff (2003), Table 2 in Westerberg and McMillan (2015)	days yr^{-1}	see above
low_q_dur	Mean duration of low-flow events (number of consecutive days < 0.2 times mean daily flow); Olden and Poff (2003), Table 2 in Westerberg and McMillan (2015)	days	see above
zero_q_freq	Percentage of days with no discharge	%	see above

975 [a] Period 1. October 1989 to 30. September 2009. [b] Upscaling approach: 2.

Table A6: Land cover characteristics.

Attribute	Description	Unit	Data source
lc_dom	3-digit short code of dominant land cover class [a,b]	-	CORINE (2012)
agr_fra	Fraction of agricultural areas (all CLC classes starting with number 2) [a,b]	-	see above

bare_fra	Fraction of bare areas, (CLC classes 332, 333) [a,b]	-	see above
forest_fra	Fraction of forest areas, (CLC classes 311, 312, 313) [a,b]	-	see above
glac_fra	Fraction of glaciers, (CLC class 335) [a,b]	-	see above
lake_fra	Fraction of natural or artificial water bodies with all-season water filling, (CLC class 512) [a,b]	-	see above
urban_fra	Fraction of areas mainly occupied by buildings including their connected areas, (CLC classes 111, 112, 121, 122, 123, 124) [a,b]	-	see above

[a] Upscaling approach: 1. [b] Land class nomenclature is listed in the folder "F_appendix".

980   **Table A7: Vegetation indices.**

Attribute	Description	Unit	Data source
lai_max	Maximum monthly mean of one-sided leaf area index (based on 12 monthly means) [a]	-	MODIS MCD15A3H (Myneni et al., 2015)
lai_diff	Difference between maximum and minimum monthly mean of one-sided leaf area index (based on 12 monthly means) [a]	-	see above
ndvi_max	Maximum monthly mean of NDVI (based on 12 monthly means) [a]	-	MODIS MOD09Q1 (Vermote, 2015)
ndvi_min	Minimum monthly mean of NDVI (based on 12 monthly means) [a]	-	see above
gvf_max	Maximum monthly mean of the green vegetation fraction (based on 12 monthly means) [a]	-	MODIS MOD09Q1 (Vermote, 2015), MODIS MCD12Q1 (Friedl and Sulla-Menashe, 2019)
gvf_diff	Difference between the maximum and minimum monthly mean of the green vegetation fraction (based on 12 monthly means) [a]	-	see above

[a] Upscaling approach: 1.

Table A8: Soil characteristics.

Attribute	Description	Unit	Data source
bedrk_dep	Depth to bedrock, maximum is 50 m [d]	m	Pelletier et al. (2016)

root_dep	Depth available for roots, maximum is 1.5 m [a,d]	m	European Soil Database Derived data (Hiederer, 2013a/b)
soil_poros	Total soil porosity [a,b,d]	-	see above
soil_condu	Saturated hydraulic conductivity, maximum is 2 m [a,b,d]	cm hr^{-1}	3D Soil Hydraulic Database of Europe (Toth et al., 2017)
soil_tawc	Total available water content (between field capacity and permanent wilting point) [a,b,d]	m	European Soil Database Derived data (Hiederer, 2013a/b)
sand_fra	Fraction of sand [a,b,c,d]	-	see above
silt_fra	Fraction of silt [a,b,c,d]	-	see above
clay_fra	Fraction of clay [a,b,c,d]	-	see above
grav_fra	Fraction of gravel [a,b,d]	-	see above
oc_fra	Fraction of organic material [a,b,d]	-	see above

[a] Areas marked as water or bedrock were excluded from calculation. [b] Aggregation weighted by depth over the different soil layers. [c] of the soil material smaller than 2 mm. [d] Upscaling approach: 1.

Table A9: Geologic attributes.

Attribute	Description	Unit	Data source
gc_dom	Dominant geologic class [b]	-	GLiM (Hartmann and Moosdorf, 2012)
gc_ig_fra	Fraction of "ice and glacier" (ig) [b]	-	see above
gc_mt_fra	Fraction of "metamorphites" (mt) [b]	-	see above
gc_pa_fra	Fraction of "acid plutonic rocks" (pa) [b]	-	see above
gc_pb_fra	Fraction of "basic plutonic rocks" (pa) [b]	-	see above
gc_pi_fra	Fraction of "intermediate plutonic rocks" (pi) [b]	-	see above
gc_py_fra	Fraction of "pyroclastics" (py) [b]	-	see above
gc_sc_fra	Fraction of "carbonate sedimentary rocks" (sc) [b]	-	see above
gc_sm_fra	Fraction of "mixed sedimentary rocks" (sm) [b]	-	see above
gc_ss_fra	Fraction of "siliciclastic sedimentary rocks" (ss) [b]	-	see above
gc_su_fra	Fraction of "unconsolidated sediments" (su) [b]	-	see above
gc_va_fra	Fraction of "acid volcanic rocks" (va) [b]	-	see above
gc_vb_fra	Fraction of "basic volcanic rocks" (vb) [b]	-	see above
gc_wb_fra	Fraction of "water bodies" (wb) [b]	-	see above

geol_perme	Subsurface permeability (log10) [a]	-	GLHYMPS (Gleeson et al., 2014)
peol_poros	Subsurface porosity [a]	-	see above

[a] Upscaling approach: 1. [b] Upscaling approach: 2.