Earth Syst. Sci. Data, 8, 279–295, 2016
www.earth-syst-sci-data.net/8/279/2016/
doi:10.5194/essd-8-279-2016

Open Access Earth System Science Data

Observation-based gridded runoff estimates for Europe (E-RUN version 1.1)

Lukas Gudmundsson and Sonia I. Seneviratne

Institute for Atmospheric and Climate Science, ETH Zurich, Universitaetstrasse 16, 8092 Zurich, Switzerland

Correspondence to: Lukas Gudmundsson (lukas.gudmundsson@env.ethz.ch)

Received: 19 November 2015 – Published in Earth Syst. Sci. Data Discuss.: 18 January 2016
Revised: 31 May 2016 – Accepted: 5 June 2016 – Published: 7 July 2016

Abstract. River runoff is an essential climate variable as it is directly linked to the terrestrial water balance and controls a wide range of climatological and ecological processes. Despite its scientific and societal importance, there are to date no pan-European observation-based runoff estimates available. Here we employ a recently developed methodology to estimate monthly runoff rates on regular spatial grid in Europe. For this we first assemble an unprecedented collection of river flow observations, combining information from three distinct databases. Observed monthly runoff rates are subsequently tested for homogeneity and then related to gridded atmospheric variables (E-OBS version 12) using machine learning. The resulting statistical model is then used to estimate monthly runoff rates (December 1950–December 2015) on a $0.5° \times 0.5°$ grid. The performance of the newly derived runoff estimates is assessed in terms of cross validation. The paper closes with example applications, illustrating the potential of the new runoff estimates for climatological assessments and drought monitoring. The newly derived data are made publicly available at doi:10.1594/PANGAEA.861371.

1 Introduction

River flow is one of the best monitored components of the terrestrial water cycle (Hannah et al., 2011; Fekete et al., 2012, 2015) and has therefore been included in the collection of essential climate variables that is featured by the World Meteorological Organization (Bojinski et al., 2014). However, despite its societal relevance (e.g. Vörösmarty et al., 2010) and key role in the earth system, there is to date no publicly available dataset that provides observation-based estimates of this variable at the pan-European scale. This situation stands in contrast to that of atmospheric variables, for which gridded estimates of, for example, precipitation and temperature (e.g Haylock et al., 2008) have been developed in the last decades. Despite the fact that gridded observations are usually limited in terms of their spatiotemporal resolution, they have the distinct advantage that they provide consistent estimates of relevant variables at every location within a spatial domain. As a consequence, gridded estimates of atmospheric variables have proven to be of great value for both scientists and practitioners in several fields (e.g. Hirschi et al., 2011; Gottfried et al., 2012).

In this paper we present a new monthly estimate of the amount of water draining from $0.5° \times 0.5°$ grid cells in Europe over the time period December 1950–December 2015. This quantity is referred to as gridded runoff estimate and eventually contributes to the discharge of large rivers (Gudmundsson and Seneviratne, 2015, referred to as GS15 from here onwards). To achieve this we employ a recently developed methodology (GS15) that combines observed river flow with gridded estimates of precipitation and temperature using machine learning. Consequently, the presented gridded runoff dataset is solely derived from observations and does not rely on strong modelling assumptions. Similar techniques have been proven successful for producing global estimates of land–atmosphere fluxes, such as evapotranspiration and gross primary production (Jung et al., 2011) and long-term streamflow characteristics, such as mean annual flow and the base flow coefficient (Beck et al., 2015).

In contrast to GS15, in which we developed and tested the methodology, we focus here on expanding the observational basis. More specifically, we assemble an unprecedented collection of observed river flow data which is subject to au-

tomated quality control and statistical homogeneity assessment. In addition we rely on the latest generation of station-based precipitation and temperature grids to estimate gridded runoff time series for Europe. Finally, the accuracy of the derived runoff estimates is assessed in terms of cross validation and its potential limitations are discussed in the context of example applications.

2 Note on terminology

This paper presents a dataset that estimates the monthly amount of water draining from $0.5° \times 0.5°$ grid cells. This quantity is referred to as "monthly runoff" and equates to the amount to water contributing to the discharge of large (continental-scale) river basins (GS15). Note that this definition is also consistent with the total grid cell runoff computed by continental- to global-scale models.

To estimate this quantity we rely on river- and streamflow observations from relatively small catchments (catchment area $\leq 500\,km^2$), which are converted to runoff rates per unit area and aggregated to monthly mean values. We note that daily streamflow is subject to processes like channel routing and therefore somewhat different from the above mentioned runoff rates. However, as the spatial and temporal scales of the associated processes are well below the resolution of the presented data product, these are not expected to impair the reliability of the presented monthly runoff estimates (see GS15 for details).

3 Data sources

3.1 Streamflow data

The presented dataset is developed using a collection of streamflow observations that is assembled from three major databases. Two of these are international collections which contain observations from many European countries (Sects. 3.1.1 and 3.1.2). As data from Spain are not up to date in these international collections, we additionally acquired the digital hydrological year book from this country (Sect. 3.1.3).

Prior to further computations daily and monthly river flow time series were converted into daily runoff rates, expressed in millimetres per day, using catchment areas provided by the respective databases.

3.1.1 The Global Runoff Data Base (GRDB)

The Global Runoff Data Centre (GRDC; http://grdc.bafg.de, last access: 9 May 2016) hosts the Global Runoff Data Base (GRDB), which is the largest international collection of river- and streamflow data. Although the GRDB is freely accessible, the GRDC is not permitted to distribute the complete database at once. Therefore we restricted our order to stations fulfilling the following set of criteria:

Stations should

1. be located in the WMO region 6 (Europe);
2. be within the following geographical domain: 25° W–70° E and 25–75° N;
3. not be located in Spain (see Sects. 3.1.3 and 4.4.1 for details on Spanish data);
4. have a minimum of 10 years of observations.

In February 2016 this resulted in a total of 1722 stations with daily values and 2047 stations with monthly values which were ordered from the GRDC. In many cases monthly data are computed by the GRDC on the basis of available daily values. There are, however, instances were only monthly data that were not computed by the GRDC are available (referred to as *originally monthly*). After filtering out monthly series that were computed on the basis of daily observations, the number of originally monthly series was found to be 860 and retained for further analysis. Monthly values calculated by the GRDC were discarded. Finally, one daily entry with missing information in catchment area was removed, resulting in 2046 daily time series.

3.1.2 The European Water Archive (EWA)

The EWA has been assembled by the European Flow Regimes from International Experimental and Network Data (Euro-FRIEND) project (http://ne-friend.bafg.de/servlet/is/7413/, last access: 9 May 2016) and is also held by the GRDC. A subset of the EWA was selected using the same criteria as for the GRDB (Sect. 3.1.1), resulting in a total of 3492 stations with daily and 3527 stations with monthly values. Only 56 originally monthly series were found and retained for further analysis. Removing entries with missing information on catchment area resulted in 3481 daily and 55 monthly records.

3.1.3 Anuario de aforos digital 2010–2011 (AFD)

Spanish streamflow data were retrieved from the digital hydrological year book (Anuario de aforos digital 2010–2011, AFD), which provides observations until 2010–2011 and is freely accessible online (http://ceh-flumen64.cedex.es/anuarioaforos/default.asp, last access: 25 May 2016). As this online platform does not allow for access to the full collection at once, we contacted the Spanish authorities and obtained a DVD containing the full database (Ministerio de Agricultura, Alimentación y Medio Ambiente, 2013). This database contains, among other information, streamflow data from 1197 gauging stations. Removing entries with missing information on catchment area reduced the number of time series to 1187.

3.2 Atmospheric data

Gridded observations of precipitation and temperature were obtained from the E-OBS (version 12) dataset (Haylock et al., 2008). E-OBS version 12 ranges from January 1950 to June 2015 and was extended to December 2015 using monthly data files that are provided for the remaining months at the time of the analysis. The E-OBS dataset provides interpolated station observations on regular spatial grids in different geographical projections. Here we chose data with a $0.5° \times 0.5°$ resolution on a regular latitude–longitude grid, which is consistent with GS15. Prior to further assessment, the daily E-OBS data were averaged to monthly mean values.

4 Streamflow data selection and preprocessing

4.1 Quality control of daily values

As the considered data stem from heterogeneous data sources, it is likely that individual daily observations differ in quality. To get first-order estimates of their credibility, all daily river flow observations were flagged according to a set of rules. As we are not aware of quality control (QC) procedures for runoff that are applicable to a large number of time series and are documented in the scientific literature, we adapt QC techniques that were developed for climatological records. More specifically, the set of rules described below is based on criteria mentioned by Reek et al. (1992) and (Project Team ECA&D and Royal Netherlands Meteorological Institute KNMI, 2013, referred to as EAC&D13 from here onward), which were adapted to the special characteristics of streamflow. In the following Q is used to denote daily runoff rates:

1. Days for which $Q < 0$ are flagged as *suspect*. The rationale behind this rule is that negative values are not physical.

2. Days for which

   $$\log(Q) - \text{mean}(\log(Q)) > 5 \times \text{SD}(\log(Q))$$

   are flagged as *suspect*. The aim of this rule is to catch extreme outliers that might be caused by instrument malfunction or processing errors, while not flagging extreme floods. Under the assumption that $\log(Q)$ is approximately normal distributed, this rule excludes outliers with a $\approx 2.8 \times 10^{-7}$ occurrence probability.

3. Values with ≥ 10 consecutive equal days for which $Q > 0$ are flagged as *suspect*. The rationale underlying this criterion is a trade-off between the fact that consecutive equal values can be caused by artifacts (e.g. instrument failures, flow regulation, ice jams) but can also reflect the true observation (e.g. related to low sensor sensitivity in the case of small day-to-day fluctuations).

4.2 Computing monthly means from daily values

As the presented data product is derived on the basis of monthly values, daily time series were aggregated to monthly means. Prior to the computation of monthly mean runoff rates daily values flagged as *suspect* are set as missing. Monthly mean runoff rates are only calculated if at least 25 days of the month are available, following the recommendations of EAC&D13. Imposing this restriction reduced the number of time series for which at least one monthly value could be computed (number of monthly time series calculated from daily values with at leat one monthly value: GRDB, 1707; EWA, 3296; AFD, 1184).

4.3 Combining daily and monthly river flow time series

Both GRDB and EWA provide data in daily as well as monthly resolution. In order to increase the spatial and temporal coverage of the observations underlying the presented data product, we aim at using originally monthly data to fill in missing values in monthly time series that were computed on the basis of quality controlled daily values (Sect. 4.2, referred to as *originally daily*). Unfortunately, the rules underlying the processing of the originally monthly series are not documented, which can lead to inhomogeneities if originally daily and originally monthly data are combined. To reduce the risk of such inhomogeneities the following set of rules is applied if merging originally daily and originally monthly series:

1. Include the unmodified originally daily values in the final collection if only these are available.

2. Include the unmodified originally monthly values in the final collection if only these are available.

3. If originally monthly data are available at time steps without originally daily data:

   a. Determine the number of overlapping time steps (n_{over}) and the squared Pearson correlation coefficient (R^2_{over}) between both the originally monthly and the originally daily time series.

   b. If $n_{\text{over}} \geq 24$ (sufficient data) and $R^2_{\text{over}} \geq 0.99$ (sufficiently similar): assume that time series can be merged reliably and merge them as follows:

      i. Use cumulative distribution function (CDF) matching (Leroux et al., 2014) to transform the distribution of the originally monthly series to match the distribution of the originally daily series. This is motivated by the common practice in remote sensing where CDF matching is used to combine time series stemming from different satellite-borne sensors (Leroux et al., 2014). The CDF matching is fitted only at locations where both originally monthly and originally

daily data are available but is used to transform all originally monthly data that were used to infill missing values in the originally daily series.

c. Use the transformed originally monthly data to infill missing values of the originally daily series.

(a) If $n_{over} < 24$ and $R^2_{over} < 0.99$: assume that time series cannot be merged reliably and keep only the series with the larger number of non-missing monthly values.

This procedure resulted in a total of 1892 monthly time series for GRDB and 3320 monthly time series for EWA that combine information from the originally daily and the originally monthly data.

4.4 Combining the river flow databases

4.4.1 Data from Spain

Spanish data are available directly from the Spanish authorities (see Sect. 3.1.3). Therefore, Spanish data contained in the GRDB and EWA were not considered, and the data stemming from AFD were directly entered into the final collection of European streamflow records.

4.4.2 Linking GRDB and EWA data

The GRDB and the EWA are to some extent populated with data from the same gauging stations. Therefore both databases need to be linked in order to avoid duplicated information. Unfortunately, linking the two databases is not straightforward, as there is no common database identifier. In addition, differences in naming conventions, inconsistent spelling of river and station names, round-off errors in station coordinates and typographical errors hamper the unambiguous linkage of the EWA and the GRDB. Further, both the GRDB and the EWA exhibit duplicated entries, which is likely related to their complex history, including irregular manual updates.

To overcome these issues we employ deduplication and record linkage techniques (Christen, 2012; Herzog et al., 2007) which are based on analysing the statistical similarity between the records. Although deduplication and record linkage techniques are quantitative methods, they usually depend on choices made by the analyst (Christen, 2012; Herzog et al., 2007). Such choices include, for example, (i) the data fields that are evaluated, (ii) the metrics used to quantify similarity, and (iii) quantitative thresholds that are used to make decisions. These choices have been identified experimentally by applying different combinations and evaluating the results carefully, which is common practice in deduplication and record linkage (Christen, 2012). In the following the final procedure for deduplication and record linkage is documented.

4.4.3 Procedure for deduplication and record linkage

Almost the same procedure is used for deduplication and record linkage. For convenience the following description is formulated for the deduplication task, in which the entries of a single database are compared to each other (for record linkage, the entries of two different databases are compared; differences for the deduplication and record linkage will be highlighted in Step 3):

Step 1. Meta-data similarity: the first step of deduplication is based on analysing the similarity of the *river names*, the similarity of the *station names* and the *geographical proximity* of all station pairs from the same country. Stations located in different countries are assumed to be different. These similarities are quantified using following distance measures:

a. The similarity between the river names and the station names is measured using the Jaro–Winkler distance, d_{JW} (Christen, 2012; van der Loo, 2014). The Jaro–Winkler distance is a popular measure for evaluating the similarity of character strings and ranges between $d_{JW} = 0$ (identical) to $d_{JW} = 1$ (no matching characters). In the following, $d_{JW,river}$ refers to the similarity of river names and $d_{JW,station}$ refers to the similarity in the station names.

b. The geographical proximity was quantified using

$$d_G = \begin{cases} 1 & \text{if } d_{GCD} > 5 \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where d_{GDC} is the great circle distance in kilometres calculated from the geographical coordinates of the station pairs. If the stations are not more than 5 km apart, d_G takes a value of 0, indicating similarity. The rationale for this threshold is that small geographical differences between stations can be related to roundoff errors in the coordinate values (e.g. 39.49214° N vs. 39.49° N).

To get an overall evaluation of the similarity of station pairs we finally compute the mean distance,

$$d_m = \frac{1}{3}(d_{JW,river} + d_{JW,station} + d_G). \quad (2)$$

Candidate duplicates are then defined as those pairs for which $d_m \leq 0.25$. In the case of multiple assignments, only the pair with the minimum d_m value is retained. The threshold value was identified experimentally, aiming at minimising false assignments, while not missing too many duplicates.

Step 2. Time series similarity: in a second step, the monthly river runoff series of the candidate duplicates that were identified in Step 1 are analysed in terms of their temporal overlap and their coefficient of determination (squared correlation coefficient), R^2. Based on the following set of criteria, database entries were classified as either "very likely identical" and "very likely different":

a. *Time series do not overlap* $\rightarrow$ *very likely different*. The rationale behind this choice is that both time series are independent and may, for example, represent time series before and after repairing or upgrading of a gauging station.

b. $R^2 > 0.99 \rightarrow$ *very likely identical*. Correlations close to one, indicate identical time series. Minor departures from $R^2 = 1$ may occur, for example, due to rounding errors in the data files.

c. $R^2 < 0.90 \rightarrow$ *very likely different*. This value has been identified experimentally.

d. $d_{\mathrm{JW,river}} + d_{\mathrm{JW,station}} \leq 0.01 \rightarrow$ *very likely identical.* Small positive values of $d_{\mathrm{JW,river}}$ and $d_{\mathrm{JW,station}}$ usually stem from minor typographical differences.

Finally, the remaining candidate duplicates were evaluated in a clerical review (Christen, 2012; Herzog et al., 2007) and manually classified into *very likely identical* and *very likely different*.

Step 3. Merging the records : different merging procedures were applied for deduplication and record linkage:

Deduplication: if duplicated entries were identified, the entry with more data points in the streamflow time series was kept. The other entry was discarded. No attempts to merge the time series have been made, as this was found to only affect a small number of stations with similar record length.

Record linkage: if two entries of GRDB and EWA were found to be very likely identical the time series were merged as follows:

i. n_{over}, the number of overlapping months, was identified.

ii. If $n_{\mathrm{over}} \geq 24$, the shorter of both time series was used to fill in missing values of the time series with more data points. The meta-data of the time series with more data points were kept. To reduce the risk of inhomogeneities, CDF matching was used to transform the series that was used to fill in missing values. Note that this procedure was also used for combining originally monthly and originally daily time series (see Sect. 4.3).

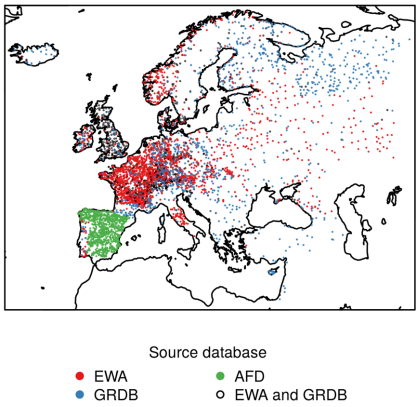

Figure 1. Locations of streamflow stations, stemming from the three considered data collections. Records from the EWA and the GRDB that were identified as *very likely identical* are indicated by black circles

iii. If $n_{\mathrm{over}} < 24$, the entry with more data points in the monthly runoff time series was kept.

4.4.4 Deduplication and record linkage results for GRDB and EWA

The deduplication procedure identified 18 very likely duplicates in the EWA and 16 very likely duplicates in the GRDB collection. Linking the deduplicated records from GRDB and EWA resulted in the identification of 4384 unique stations.

4.4.5 A combined European monthly runoff database (ERDB)

The 4384 linked records from the EWA and the GRDB were combined with the 1184 stations from AFD (Fig. 1). The total number of available stations contributing to this European runoff database (ERDB) is 5568. Figure 2 shows the spatial and temporal coverage of the available streamflow observations. Generally, observations are most abundant throughout the second half of the twentieth century. The month with the largest number of available streamflow observations (4336) is May 1980. Figure 3 provides information on the fraction of missing months in the combined dataset together with information on the seasonal distribution of missing values. Overall the fraction of missing months increases for more arid conditions. In addition, it is interesting to note that there is a general tendency for most missing values to occur in winter in cold regions (e.g. Scandinavia and the Alps), whereas late summer months are more likely to have the highest fraction of missing values in other regions. This is also reflected in the seasonal cycle of the total number of missing months, which has two distinct peaks: one in winter and one in late summer. Table 1 lists further summary statistics on the fraction of missing values, time series length and the catchment areas of the final collection of monthly runoff series.

Table 1. Percentiles of selected statistics of the monthly runoff database. Shown are the fraction of missing months (Fraction missing), the time series length in months (Length) as well as the catchment area in square kilometres (Area).

Percentile	0 %	10 %	25 %	50 %	75 %	90 %	100 %
Fraction missing	0.00	0.00	0.01	0.05	0.20	0.41	0.96
Length	1	132	240	432	648	951	2496
Area	0.07	42	112	333	1421	7211	1.36×10^6

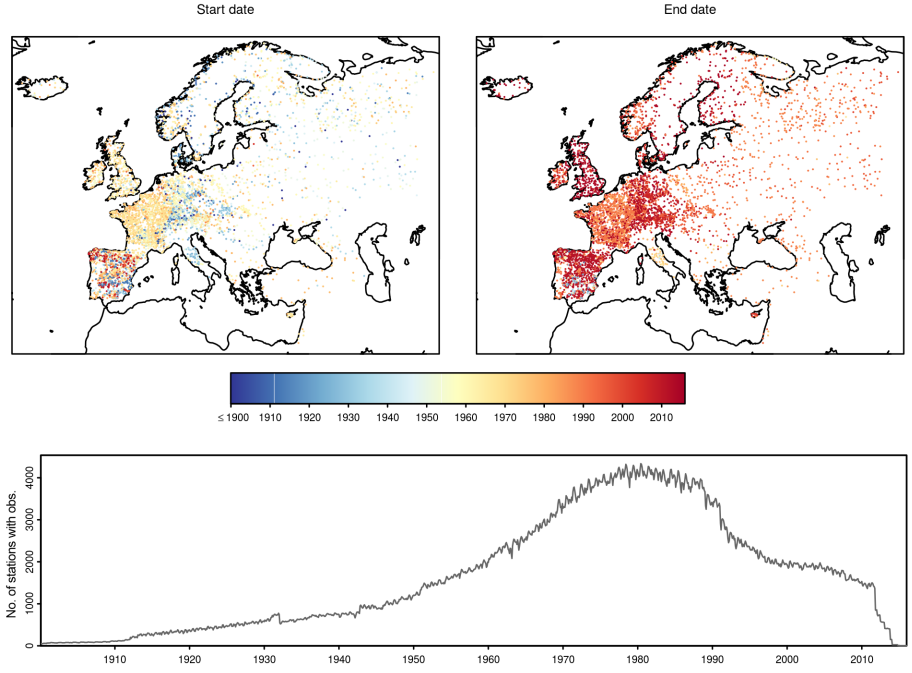

Figure 2. Spatial and temporal coverage of available streamflow observations. The top row shows the date of the first and the date of the last available observation at each station. The bottom panel shows the total number of stations with observations for each month.

4.5 Homogeneity testing

Climate records can exhibit changes which do not reflect real climatic or environmental change. In the context of river flow, such breakpoints could, for example, be related to changes in instrumentation, gauge resaturation, recalibration of rating curves, flow regulation or channel engineering. In the climatological literature such effects are commonly referred to as inhomogeneities. While a substantial body of literature is devoted to the treatment of inhomogeneities in atmospheric variables (e.g. Buishand, 1982; Alexandersson, 1986; Peterson et al., 1998; Wijngaard et al., 2003; Reeves et al., 2007; Costa and Soares, 2009; Vicente-Serrano et al., 2010; Domonkos, 2013), there is only limited literature concerned with the homogeneity testing of streamflow time series using automated methods (Buishand, 1984; Chu et al., 2013).

Identification of inhomogeneities in large data collections is usually based on tests that aim at identifying breakpoints in the considered time series. Such breakpoints can, for example, be a sudden shift in the mean, variance or higher-order moments. For the presented data product the test battery for inhomogeneity detection that is used by EAC&D13 is employed:

1. Standard normal homogeneity test (Alexandersson, 1986),
2. Buishand range test (Buishand, 1982),
3. Pettitt test (Pettitt, 1979),
4. Von Neumann ratio test (von Neumann, 1941).

The power of this test battery has been evaluated for temperature and precipitation series in Europe (Wijngaard et al., 2003), which increases the confidence in the reliability of these methods.

The considered tests are based on the assumption that the data points of the time series are independent and identically distributed (iid). To approximate this assumption, the monthly mean time series (Sect. 4.2) were preprocessed as follows, aiming at de-trending, de-seasonalising and pre-whitening the data:

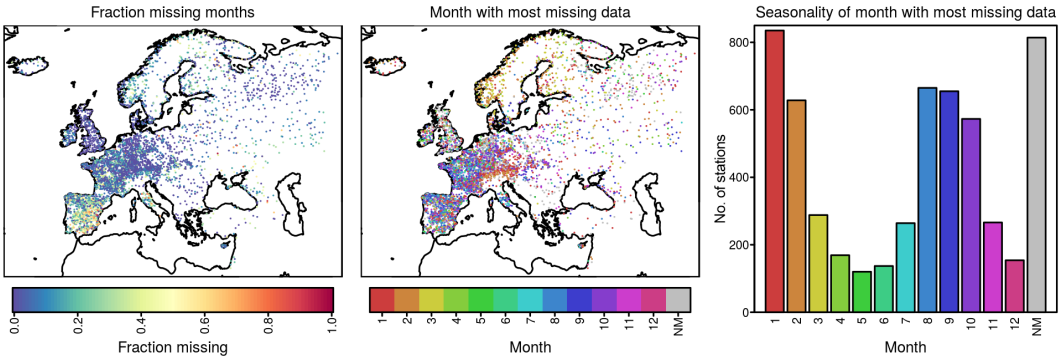

Figure 3. Overview on the spatial and seasonal distribution of missing months. Shown are the fraction of missing months at each station (left), the month which has on average most missing values at each station (centre) and the regional frequency distribution of the months with the most missing values (right). NM indicates no missing values.

1. As runoff usually has a skewed distribution, the monthly time series were log-transformed. As the logarithm is not defined for zero values, 0.01 was added before transformation.

2. To remove the seasonal cycle and to reduce the influence of monotonic trends, the log-transformed monthly time series were detrended for each month separately. For this, a linear least-squares trend was fitted to all Januaries, Februaries, etc. and subsequently subtracted from the corresponding months.

3. The detrended runoff residuals can still exhibit a high degree of serial correlation, violating the iid assumption. Therefore the residuals were further pre-whitened. For this we followed previous studies (Chu et al., 2013; Burn and Hag Elnur, 2002) and considered the residuals of a lag-1 autocorrelation model fitted to the data.

The four tests were subsequently applied to the preprocessed time series. Following EAC&D13, the credibility of time series is classified based on the number of tests that reject the null hypothesis of no breakpoint:

1. *useful*: 0 or 1 test rejects the null hypothesis at the 1 % level.

2. *doubtful*: 2 tests reject the null hypothesis at the 1 % level.

3. *suspect*: 3 or 4 tests reject the null hypothesis at the 1 % level.

The test battery was applied to monthly runoff series that had at least 24 monthly values from 1950 onwards, corresponding to the time window of the presented data product. Figure 4 shows the number of rejected null hypothesis for each station. Table 2 shows the total number of rejections.

Table 2. Number of stations for which 0, 1, . . . , 4 of the considered tests reject the null hypothesis of no breakpoint (1 % level) at monthly resolution. Stations with more than one rejection are marked as *suspect*.

No. rejections	0	1	2	3	4
No. stations	1049	3780	618	121	0

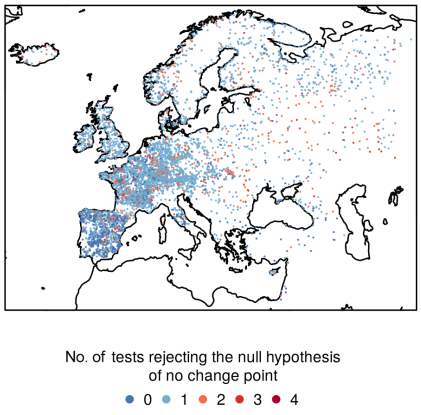

Figure 4. Homogeneity testing: number of tests that reject the null hypothesis of no breakpoint at each station considered at the 1% level. Stations marked blue (zero or one rejection) are considered *useful*. Stations marked red (more rejections) are considered *suspect*.

4.6 Assigning monthly runoff rates to the 0.5° × 0.5° grid of the E-OBS data

The methodology for estimating runoff at ungauged locations proposed by GS15 relies on assigning gauging stations with relatively small catchments to regular spatial grids. Here the monthly mean runoff rates of the selected stations were assigned to the 0.5° × 0.5° grid defined by the E-OBS data using the following steps:

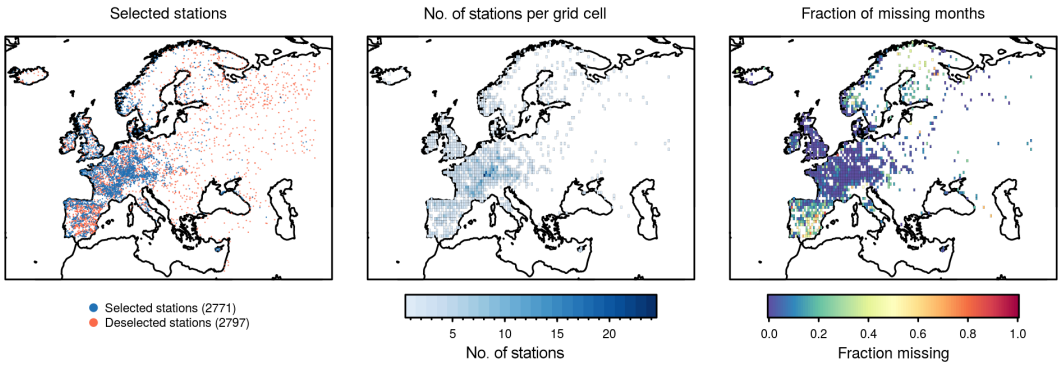

Figure 5. Assigning stations to the 0.5° × 0.5° grid cells defined by the E-OBS data. Left: selected stations fulfilling all selection criteria (see Sect. 4.6). Centre: number of stations per grid cell. Right: fraction of months with no or insufficient data.

1. Select stations:

   a. Only stations with catchment areas $\leq 500\,\mathrm{km}^2$ are selected. This threshold roughly corresponds to half the area of a grid cell at 71° N and aims at reducing the catchment area that is not located within the grid cell.

   b. Only stations with at least 24 non-missing months from 1950 onwards are selected

   c. Only stations that are labelled *useful* in the homogeneity analysis (Sect. 4.5) are selected.

   d. Only stations with a long-term mean runoff less than $10\,000\,\mathrm{mm\,year}^{-1}$ are selected as larger values are deemed to be physically very unlikely.

2. Assign stations to the grid cells which include the station coordinates.

3. Compute the weighted mean runoff rate of all stations within a grid cell, using the catchment areas of the available stations as weights. The weights are calculated for each time step separately to account for irregular temporal coverage of the stations.

This procedure resulted in a total of 2771 selected stations which were assigned to 1073 grid cells, implying that there are on average 2.5 stations assigned to each grid cell. Figure 6 shows the frequency distribution of the number of stations that were assigned to grid cells. The selected stations are shown in Fig. 5. Figure 5 also shows the number of stations in each grid cell as well as the fraction of non-missing months. Figure 7 provides a general overview on the spatial and temporal coverage of the gridded station data.

As the above-described procedure can assign data from several stations with different temporal coverage to one grid cell, it can happen that the resulting time series exhibits sudden jumps or other inhomogeneities. To reduce the influence of such artifacts the homogeneity testing that was applied to the station data (Sect. 4.5) was also applied to the gridded observations.

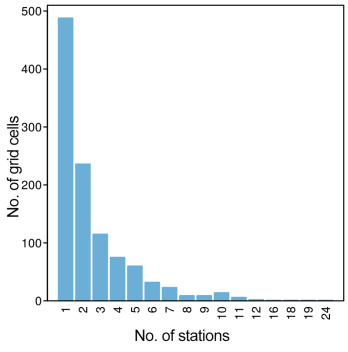

Figure 6. Frequency distribution of grid cells with 1, 2, ..., 24 stations.

Table 3. Number of grid cells for which 0, 1, ..., 4 of the considered tests reject the null hypothesis of no breakpoint (1 % level) at monthly resolution. Grid cells with more than one rejection are excluded from the analysis.

No. rejections	0	1	2	3	4
No. stations	90	871	82	30	0

Table 3 shows the total number of rejections of the test battery. Grid cells for which more than one test rejected the null hypothesis at the 1 % level were excluded from further analysis. Figure 8 shows the shows the final selection of 961 grid cells.

5 Observational gridded runoff estimates for Europe

5.1 Estimating runoff on a regular spatial grid

The technique used to estimate gridded runoff time series is identical to the approach introduced by GS15. For convenience we provide here a brief overview of this method. For

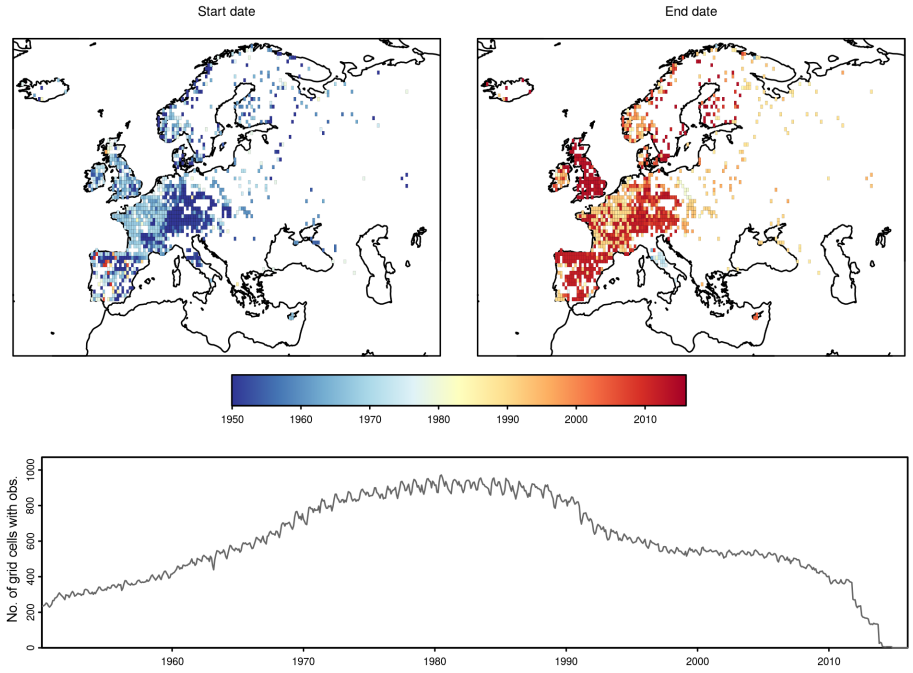

Figure 7. Spatial and temporal coverage after assigning the monthly runoff series to the 0.5° × 0.5° defined by the E-OBS data. The top row shows the date of the first and the date of the last available observation at each station. The bottom panel shows the total number of stations with observations for each month.

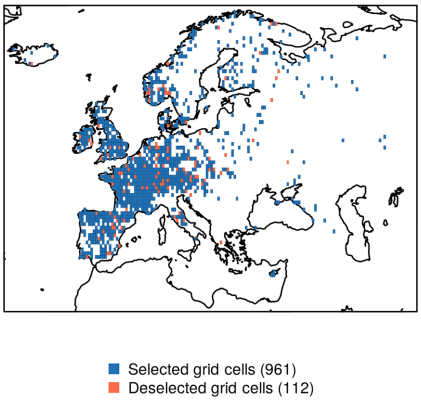

Figure 8. Final selection of grid cells with observations. Only grid cells with homogeneous time series were selected. See text for details.

a full description of the employed methods we refer to GS15. Following GS15 we aim at modelling the monthly runoff rate $Q_{x,t}$ at the grid location x and at time step t as a function of gridded precipitation, $P_{x,t}$, and temperature, $T_{x,t}$. For this we assume that

$$Q_{x,t} = h(\tau_n(P_{x,t}), \tau_n(T_{x,t}), \quad (3)$$

where $\tau_n(X_{x,t}) = [X_{x,t}, X_{x,t-1}, \ldots, X_{x,t-n}]$ is a time lag operator that gives access to the past n time steps. As in GS15, we chose $n = 11$, implying that monthly runoff rates are estimated on the basis of the precipitation and temperature evolution of the preceding year. The function h represents a random forest (RF; Breiman, 2001). RFs are flexible machine learning tools that are based on classification and regression trees that are grown on bootstrap samples of the data. For estimating monthly runoff on the 0.5° × 0.5° grid of the E-OBS data the model (Eq. 3) was trained using the selected grid cells with observed monthly runoff rates and E-OBS precipitation and temperature. The fitted model was subsequently applied to all grid cells of the E-OBS data to derive a pan-European estimate of monthly runoff.

5.2 Model selection and validation

5.2.1 Cross-validation experiments

As in GS15 the model selection and validation is conducted using two independent cross-validation experiments. For the first experiment, the grid cells with observations were randomly split into 10 equally sized subsamples. The model was then trained using 9 of the 10 subsamples and subsequently used to predict the remaining subsample. This procedure was repeated until each subsample has been left out once and is referred to as cross validation in space. This focuses on the accuracy of estimates at locations that were not used for model training. The second experiment focuses on the accuracy at time steps that were not used for model training. For this the available data where split into 10 consecutive time blocks. The model was then trained using 9 of the 10 time

blocks and subsequently used to predict the time block that has been left out. This procedure was repeated until each time block has been left out once.

5.2.2 Model selection

As any other machine learning tools, RFs have a number of parameters that control the trade-off between the flexibility and the reliability of the resulting model. While GS15 used the default parameters recommended by Hastie et al. (2009), we found that this led to a in slight overfitting of the model for the extended observational basis used in this study. One of the control parameters is the minimum node size, n, which determines the number of observations retained in the final branches of the individual regression and classification trees contributing to the RF (see GS15 Sect. 3.2 for an overview of the algorithm and Breiman, 2001, or Hastie et al., 2009, for further details). Generally speaking, RFs are more flexible for smaller n, implying the possibility of achieving better fits. This, however, also means that the model is more prone to overfitting the data, i.e. an increased risk of fitting the model to noise instead of to the true signal. An additional feature of RFs is that they rely on an ensemble of B classification and regression trees that are grown on bootstrap samples of the data. Generally, RFs become more stable as B increases. However, depending on the size of the training problem, very large B may become prohibitive.

To investigate the effect of different values of n on the model accuracy we performed the above-described cross-validation experiments for $n = 10, 20, \ldots, 50$. Note that $n = 10$ was used in GS15. In addition, we also assess the effect of B on the stability of the estimate, aiming at determining whether a reduced B also yields stable results. More specifically we assess $B = 1000, 500, 250$, where $B = 1000$ was used in GS15.

As in GS15 model selection is based on the global root mean square error (RMSE), computed over all time steps and grid cells. Uncertainty in the RMSE is quantified in terms of 95 % bootstrap confidence intervals (2000 replications). The optimal values of n and B is then selected as follows, aiming at identifying the least flexible model (larger n) of which the performance is close to the performance of the most flexible model (smallest n) while reducing the computational requirements (smaller B):

Step 1. Choose optimal *n* value. For all B values:

a. Identify $\mathrm{RMSE}_{n=10}$ the RMSE for the smallest $n = 10$ value.

b. Choose any larger n value for which the RMSE is within the 95 % confidence bound of $\mathrm{RMSE}_{n=10}$.

c. If the results between cross validation in space and cross validation in time differ, choose the smaller n value.

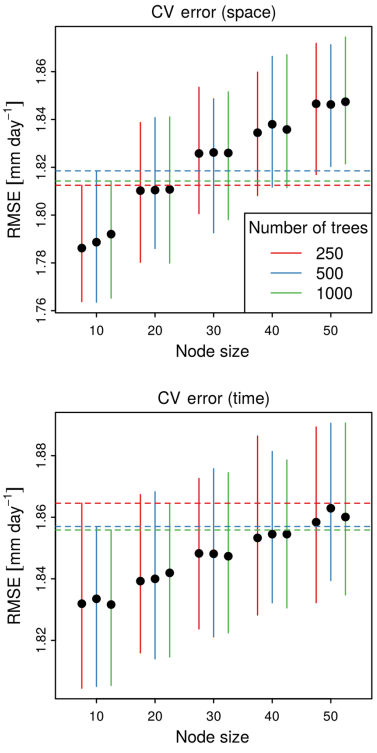

Figure 9. Cross-validation error different values of the nose size parameter (n) and different number of trees (B). The two panels show results for cross validation in space (top) and cross validation in time (bottom). Shown is the root mean square error (RMSE) together with 95 % confidence intervals. Vertical lines indicate 95 % confidence interval of the RMSE. Dashed horizontal lines indicate the upper confidence interval for $n = 10$.

Step 2. Choose *B* optimal value. Choose the smallest B value for which the RMSE lies within the 95 % confidence value of the RMSE for $B = 1000$

Figure 9 shows the RMSE for both cross validation in time and cross validation in space, as well as for all considered values of n and B. Based on the criteria described above, $n = 20$ and $B = 250$ were selected for the final data product. In the remainder of the article, only results for these parameters are shown.

5.2.3 Accuracy of the runoff estimates

We employ here the same performance metrics that have been used by GS15 to quantify the accuracy of the gridded runoff estimate. For convenience we reproduce here the definition of the considered metrics, where o_t refers to a time series of observed runoff rates at a grid cell and m_t represents the corresponding model estimate. For a detailed discussion of the different measures we refer the reader to GS15.

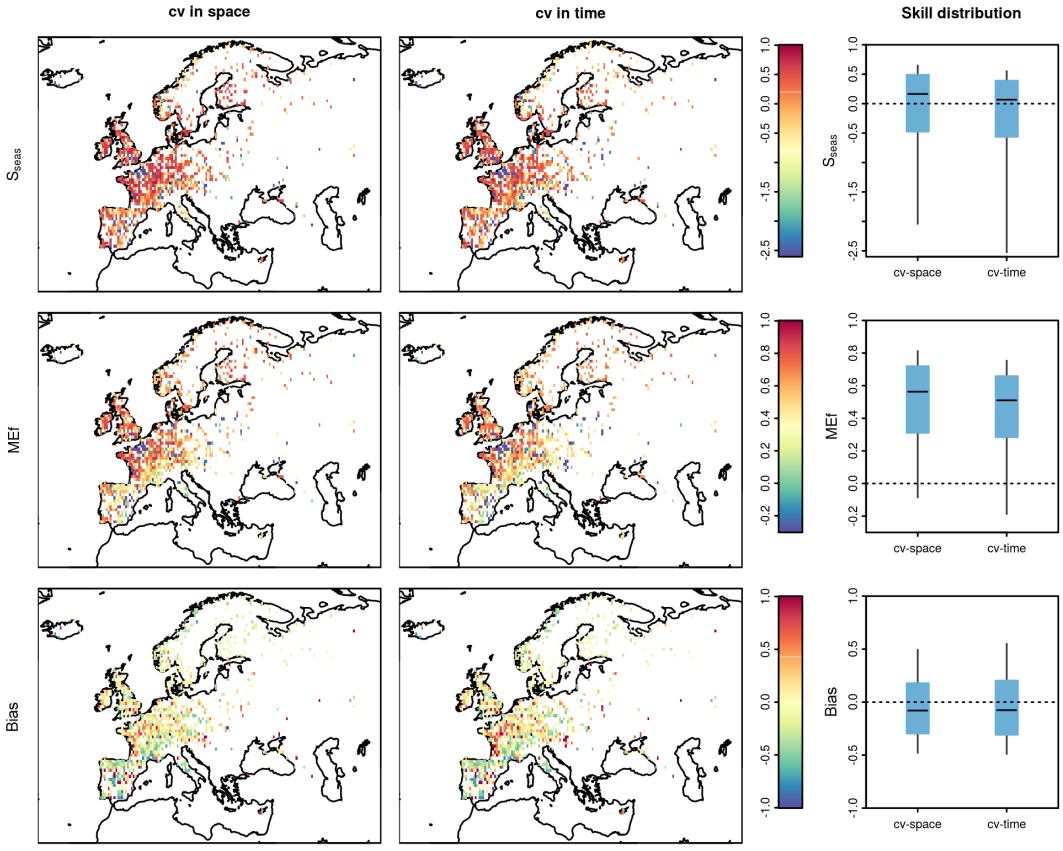

Figure 10. Spatial distributions and box plots (whiskers: 10th and 90th percentiles; box: interquartile range; bar: median) of S_{seas}, MEf and BIAS for both cross-validation experiments.

1. The seasonal cycle skill score (Wilks, 2011)

$$S_{\text{seas}} = 1 - \frac{\sum_t (m_t - o_t)^2}{\sum_t (m_t - \text{seas}(o_t))^2}, \tag{4}$$

where $\text{seas}(o_t)$ refers to the long-term mean runoff for each month. S_{seas} ranges from $-\infty$ to one (best value) and positive values indicate that m_t is on average closer to the observations than the long-term mean seasonal cycle.

2. The model efficiency (Wilks, 2011; Nash and Sutcliffe, 1970)

$$\text{MEf} = 1 - \frac{\sum_t (m_t - o_t)^2}{\sum_t (m_t - \text{mean}(o_t))^2}, \tag{5}$$

where $\text{mean}(o_t)$ refers to the long-term mean of the observation. MEf ranges between $-\infty$ and one (best value). Positive values indicate that m_t is closer to the observations than the observed long-term mean.

3. The relative model bias

$$\text{BIAS} = \frac{\text{mean}(m_t - o_t)}{\text{mean}(o_t)}, \tag{6}$$

which has an optimal value of zero. Positive and negative values indicate overestimation and underestimation respectively.

4. The coefficient of determination (squared correlation coefficient), R^2. R^2 ranges from zero to one (best value).

5. The coefficient of determination between the observed and the modelled mean annual cycle, R^2_{CLIM}. R^2_{CLIM} ranges from zero to one (best value).

6. The coefficient of determination between the monthly anomalies (i.e. monthly time series with the long-term mean of each month removed), R^2_{ANO}. R^2_{ANO} ranges from zero to one (best value).

Figures 10 and 11 display the results of both cross-validation experiments. Shown are the spatial patterns as well

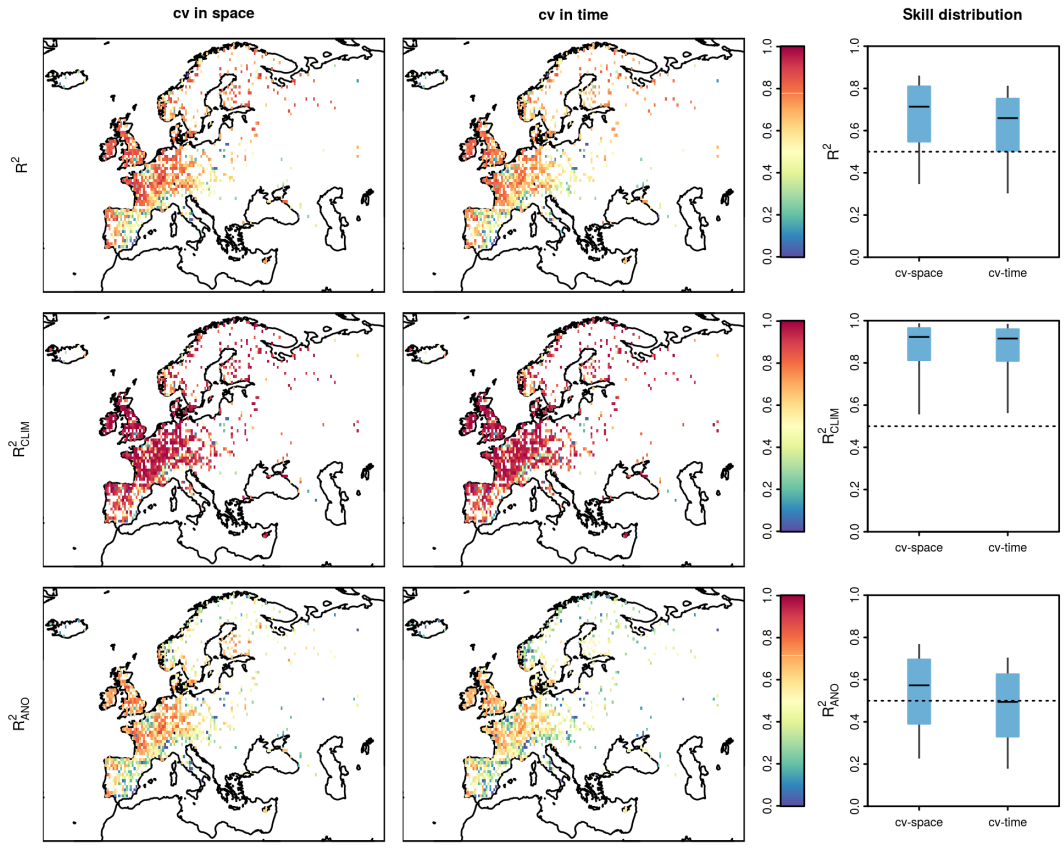

Figure 11. Spatial distributions and box plots (whiskers: 10th and 90th percentiles; box: interquartile range; bar: median) of R^2, R^2_{CLIM} and R^2_{ANO} for both cross-validation experiments.

as the overall distribution of all considered performance metrics. Generally, the accuracy of the presented dataset is in line with GS15, including the fact that the performance for the cross validation in space is somewhat higher than the performance for the cross validation in time. For both cross-validation experiments there is no clear spatial pattern of S_{seas}. This shows that the overall performance of the estimate does not depend on the region. The fact that the median of S_{seas} is well above zero shows that the runoff estimates are closer to the observations than mere repetitions of the mean annual cycle at most considered locations. The situation is similar for MEf, highlighting the consistency between both measures. The relative bias also exhibits some spatial patterns, with a tendency for increased underestimation toward the south. However, the median of this measure is approximately zero, showing that the runoff estimates developed are approximately unbiased. This is a slight improvement over GS15 and may be related to the increased number of considered stations or to the different atmospheric data used. The coefficient of determination, R^2, is generally highest in the centre of the spatial domain, which coincides with the region with the highest station density. The median R^2 values are relatively high, highlighting the ability of the estimate to capture the temporal dynamics of the observations. In general there is little spatial variability in the coefficient of determination between the observed and the estimated climatologies, R^2_{CLIM}. This, together with the fact that median R^2_{CLIM} is very high, highlights that the gridded runoff estimate is capable of capturing the mean seasonal cycle with a high degree of accuracy. Also, the anomaly correlation, R^2_{ANO}, has a weak spatial pattern, with a tendency towards increased correlation in the centre of the spatial domain. Overall the anomaly correlation is somewhat lower than R^2, owing to the fact that the regular mean annual cycle has been removed. Nevertheless, median R^2_{ANO} is larger than 0.5 for cross validation in space and close to 0.5 for cross validation in time, highlighting that the estimates can capture more than half of the variance of the anomalies.

5.3 Properties and limitations of the observation-based gridded runoff estimates

The final observation-based gridded runoff dataset is created by first training the model using all available stations and E-OBS precipitation and temperature. Subsequently the model is used to estimate monthly runoff rates [mm day^{-1}] at all

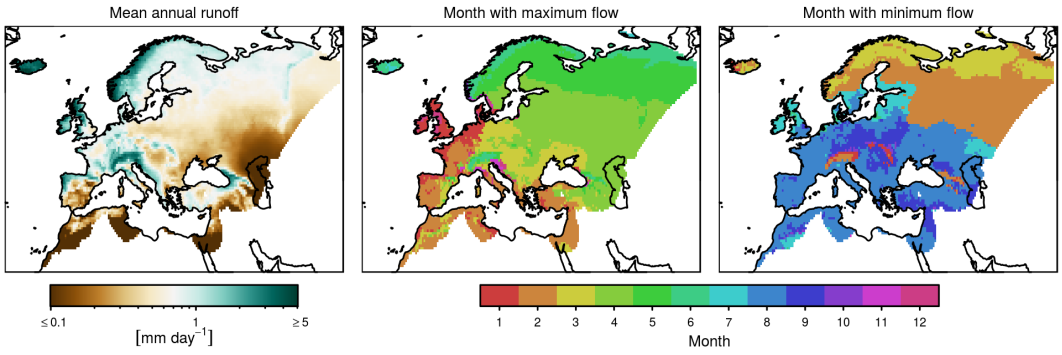

Figure 12. Long-term mean of the presented gridded runoff field as well as the month of the maximum and minimum of the mean annual cycle.

grid cells of the E-OBS forcing. This procedure results in the following features and limitations which should be considered for any application of the newly developed gridded runoff estimates for Europe:

1. The spatial and temporal extent of the data is determined by the coverage of the forcing data.

2. A consequence of the time-lag operator in Eq. (3) is that any missing month in the forcing data will result in 12 missing months in the runoff estimates.

3. Most station data are located in central and western Europe, suggesting that the data will have the highest degree of accuracy in these regions. In other regions the reliability of the data is expected to decrease gradually. Therefore special care should be taken if analysing the data in regions with low station coverage.

4. The E-OBS dataset also covers parts of the Caspian Sea and other large inland water bodies. Although it might not be physically meaningful to provide runoff estimates for these locations, we opted not to remove the corresponding grid cells from the dataset. The rationale underlying this decision is that the definition of shorelines in gridded data products depends on several assumptions and we want to allow the users to make such choices corresponding to their needs.

5.4 Example applications

In the following we present two example applications of the newly developed dataset. These applications closely follow the ones presented in GS15.

5.4.1 Long-term mean runoff statistics

Figure 12 shows the long-term mean of the gridded runoff estimates as well as the month with the maximum and the minimum of the mean annual cycle. The map of the long-term

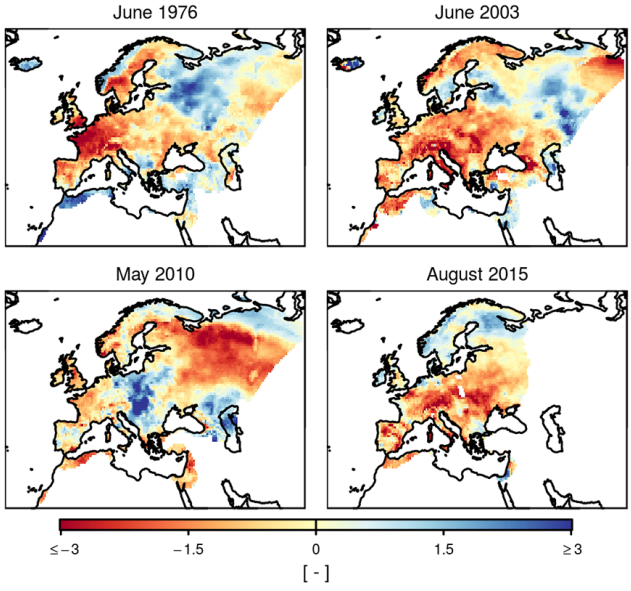

Figure 13. Standardised runoff anomalies for selected drought events in Europe.

mean highlights that central and northern Europe have highest mean annual runoff rates, whereas the south and the east are generally drier. The maps displaying the months with the maximum and the months with the minimum of the mean annual cycle show distinct regional differences. In western and southern Europe, the peak of the seasonal cycle occurs in the winter months, followed by a summer minimum. In northern Europe, the minimum runoff occurs in the winter months, followed by a peak in spring. In eastern Europe, maximum runoff rates occur in spring and are followed by a minimum in summer.

5.4.2 Drought monitoring

As runoff reflects the excess water that is available to ecosystems, it is an interesting candidate for drought monitoring. To

assess droughts, we follow previous studies (Zaidman et al., 2002; GS15) and use standardised runoff anomalies as a drought index. These are computed by first log-transforming the runoff time series at each grid cell. Subsequently the 30-year long-term mean of each month at each grid cell is subtracted from the log-transformed time series (base period: 1961–1990). Finally, the time series is divided by the 30-year standard deviation of each month.

Figure 13 shows the standardised runoff anomalies for four well-documented events with exceptionally dry conditions. Drought conditions in 1976 were among the most severe in Europe throughout the course of the 20th century (Tallaksen and Stahl, 2014). Summer 2003 is well known for its exceptionally hot and dry conditions (Schär et al., 2004; Andersen et al., 2005; Seneviratne et al., 2012). Spring 2010 shows dry conditions in the advent of the intense heatwave that struck Russia a few months later (Barriopedro et al., 2011; Orth and Seneviratne, 2015; Hauser et al., 2016). In summer 2015 large parts of Europe exhibited extremely hot and dry conditions (Hoy et al., 2016), which is reflected, for example, in reported extreme low-flow return periods ($>$ 20 years) for a large number of catchments in central Europe (Van Lanen et al., 2016).

6 Conclusions

In conclusion, we presented an observational dataset that provides monthly pan-European runoff estimates and ranges from December 1950 to December 2015. The data are a significant update of our previous assessment (GS15), which only included data ranging to 2001. The dataset is based on an unique collection of streamflow observations from small catchments which were upscaled on a $0.5° \times 0.5°$ grid on the basis of gridded precipitation and temperature data using machine learning. Two cross-validation experiments document the overall performance of the newly developed estimates. These experiments show that the accuracy of the data is in line with previous results (GS15), highlighting the robustness of the estimation technique used. The two example applications highlight the utility of the newly developed pan-European runoff estimates, for both climatological assessments and drought monitoring. These examples show that the presented gridded dataset allows for an unprecedented observational view on large-scale features of runoff variability in Europe, especially in regions with limited observational coverage.

7 Data availability

The data are publicly available in NetCDF format (Gudmundsson and Seneviratne, 2016) and can be downloaded from http://dx.doi.org/10.1594/PANGAEA.861371. A table documenting the considered stations is available as a supplement and described in Appendix A.

Appendix A: Meta-data of the considered stations

The streamflow observations collected in the ERDB (Sect. 4.4.5) provide an unprecedented opportunity for observation-based freshwater research in Europe. As the data are protected by copyright, we cannot make this collection publicly available. Instead, we include a meta-data table of all considered stations, which should allow other researchers to reproduce the collection if they have access to the original databases (Sect. 3.1).

In the following the different fields of this meta-data table are briefly described. For convenience, we partition the description of the meta-data into three blocks, labelled Part A to Part C:

Part A. Basic station information: summarises information on names, spatial location and temporal coverage:

ERDB.id The database identifier used to organise ERDB. This identifier is structured as `AA_XXXXXXX`, where `AA` is the country code and `XXXXXXX` a running number.

country Country code.

river Name of the river or stream.

station Name of the station.

longitude Longitude of the station in decimal degrees.

latitude Latitude of the station in decimal degrees.

altitude Altitude of the station in metres above sea level.

area Catchment area in square kilometres.

start.date Date of the first entry in the time series.

end.date Date of the last entry in the time series.

length Time series length in number of months.

number.months Number of months with non missing data.

frac.missing The fraction of missing months.

Part B. Record linkage results: summarises the results of the record linkage procedure described in Section 4.4.3. Note: if both the fields **EWA.no** and **GRDB.no** contain values, this indicates that the records of EWA and GRDB have been linked.

EWA.no Database identifier of EWA, if any EWA record is assigned to the entry.

GRDB.no Database identifier of GRDB, if any GRDB record is assigned to the entry.

AFD.no Database identifier of AFD, if any AFD record is assigned to the entry.

river.dist The value of $d_{\mathrm{JW,river}}$, if more than one database was used to generate the record.

station.dist The value of $d_{\mathrm{JW,station}}$, if more than one database was used to generate the record.

latlon.dist The value of d_{GCD} in kilometres, if more than one database was used to generate the record.

latlon.bin.dist The value of d_{G} in kilometres, if more than one database was used to generate the record.

mean.dist The value of d_{m} (Eq. 2), if more than one database was used to generate the record.

Part C. Homogeneity testing: summarises the results of the homogeneity assessment (Sect. 4.5). Note that the homogeneity assessments have only been conducted from 1950 onwards.

SNHtest The results of the standard normal homogeneity test. Following values are possible: `"NS"`, the test does not reject the null hypothesis of no break point; `"p5"`, the test rejects the null hypothesis, $p < 0.05$; `"p1"`, the test rejects the null hypothesis, $p < 0.01$; `"NSD"`, insufficient data (fewer than 24 months).

BHRtest The results of the Buishand range test. See **SNHtest** for possible values.

PETtest The results of the Pettitt test. See **SNHtest** for possible values.

VONtest The results of the Von Neumann ratio test. See **SNHtest** for possible values.

The Supplement related to this article is available online at doi:10.5194/essd-8-279-2016-supplement.

Acknowledgements. The support of the ERC DROUGHT-HEAT (contract no. 617518) and DROUGHT-R&SPI projects (contract no. 282769) is acknowledged. We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://www.ecad.eu). The efforts to assemble the European Water Archive (EWA) by the UNESCO IHP VII FRIEND programme, the data collection and management by the GRDC, and the provision of data by Spanish authorities are gratefully acknowledged.

Edited by: A. Gelfan
Reviewed by: C. Prudhomme, H. Müller Schmied, G. V. Ayzel, and V. Moreydo

References

Alexandersson, H.: A homogeneity test applied to precipitation data, J. Climatol., 6, 661–675, doi:10.1002/joc.3370060607, 1986.

Andersen, O. B., Seneviratne, S. I., Hinderer, J., and Viterbo, P.: GRACE-derived terrestrial water storage depletion associated with the 2003 European heat wave, Geophys. Res. Lett., 32, L18405, doi:10.1029/2005GL023574, 2005.

Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M., and García-Herrera, R.: The Hot Summer of 2010: Redrawing the Temperature Record Map of Europe, Science, 332, 220–224, doi:10.1126/science.1201224, 2011.

Beck, H. E., de Roo, A., and van Dijk, A. I.: Global maps of streamflow characteristics based on observations from several thousand catchments, J. Hydrometeor., 16, 1478–1501, doi:10.1175/JHM-D-14-0155.1, 2015.

Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., and Zemp, M.: The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy, B. Am. Meteorol. Soc., 95, 1431–1443, doi:10.1175/BAMS-D-13-00047.1, 2014.

Breiman, L.: Random Forests, Mach. Learn., 45, 5–32, doi:10.1023/A:1010933404324, 2001.

Buishand, T.: Some methods for testing the homogeneity of rainfall records, J. Hydrol., 58, 11–27, doi:10.1016/0022-1694(82)90066-X, 1982.

Buishand, T.: Tests for detecting a shift in the mean of hydrological time series, J. Hydrol., 73, 51–69, doi:10.1016/0022-1694(84)90032-5, 1984.

Burn, D. H. and Hag Elnur, M. A.: Detection of hydrologic trends and variability, J. Hydrol., 255, 107–122, doi:10.1016/S0022-1694(01)00514-5, 2002.

Christen, P.: Data Matching, Springer, doi:10.1007/978-3-642-31164-2, 2012.

Chu, M. L., Ghulam, A., Knouft, J. H., and Pan, Z.: A Hydrologic Data Screening Procedure for Exploring Monotonic Trends and Shifts in Rainfall and Runoff Patterns, J. Am. Water Resour. As., 50, 928–942, doi:10.1111/jawr.12149, 2013.

Costa, A. and Soares, A.: Homogenization of Climate Data: Review and New Perspectives Using Geostatistics, Math. Geosci., 41, 291–305, doi:10.1007/s11004-008-9203-3, 2009.

Domonkos, P.: Efficiencies of Inhomogeneity-Detection Algorithms: Comparison of Different Detection Methods and Efficiency Measures, J. Climatol., 2013, 15 pp., doi:10.1155/2013/390945, 2013.

Fekete, B. M., Looser, U., Pietroniro, A., and Robarts, R. D.: Rationale for Monitoring Discharge on the Ground, J. Hydrometeorol., 13, 1977–1986, doi:10.1175/JHM-D-11-0126.1, 2012.

Fekete, B. M., Robarts, R. D., Kumagai, M., Nachtnebel, H.-P., Odada, E., and Zhulidov, A. V.: Time for in situ renaissance, Science, 349, 685–686, doi:10.1126/science.aac7358, 2015.

Gottfried, M., Pauli, H., Futschik, A., Akhalkatsi, M., Barancok, P., Benito Alonso, J. L., Coldea, G., Dick, J., Erschbamer, B., Fernandez Calzado, M. R., Kazakis, G., Krajci, J., Larsson, P., Mallaun, M., Michelsen, O., Moiseev, D., Moiseev, P., Molau, U., Merzouki, A., Nagy, L., Nakhutsrishvili, G., Pedersen, B., Pelino, G., Puscas, M., Rossi, G., Stanisci, A., Theurillat, J.-P., Tomaselli, M., Villar, L., Vittoz, P., Vogiatzakis, I., and Grabherr, G.: Continent-wide response of mountain vegetation to climate change, Nature Clim. Change, 2, 111–115, doi:10.1038/nclimate1329, 2012.

Gudmundsson, L. and Seneviratne, S. I.: Towards observation-based gridded runoff estimates for Europe, Hydrol. Earth Syst. Sci., 19, 2859–2879, doi:10.5194/hess-19-2859-2015, 2015.

Gudmundsson, L. and Seneviratne, S. I.: E-RUN version 1.1: Observational gridded runoff estimates for Europe, link to data in NetCDF format (69 MB), doi:10.1594/PANGAEA.861371, 2016.

Hannah, D. M., Demuth, S., van Lanen, H. A. J., Looser, U., Prudhomme, C., Rees, G., Stahl, K., and Tallaksen, L. M.: Large-scale river flow archives: importance, current status and future needs, Hydrol. Process., 25, 1191–1200, doi:10.1002/hyp.7794, 2011.

Hastie, T., Tibshirani, R., and Friedman, J. H.: The Elements of Statistical Learning – Data Mining, Inference, and Prediction, Second Edition, Springer Series in Statistics, Springer, New York, 2nd Edn., available at: http://www-stat.stanford.edu/~tibs/ElemStatLearn/ (last access: 5 July 2016), 2009.

Hauser, M., Orth, R., and Seneviratne, S. I.: Role of soil moisture versus recent climate change for the 2010 heat wave in western Russia, Geophys. Res. Lett., 43, 2819–2826, doi:10.1002/2016GL068036, 2016.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, J. Geophys. Res., 113, D20119, doi:10.1029/2008JD010201, 2008.

Herzog, T. N., Scheuren, F. J., and Winkler, W. E.: Data Quality and Record Linkage Techniques, Springer, New York, USA, 2007.

Hirschi, M., Seneviratne, S. I., Alexandrov, V., Boberg, F., Boroneant, C., Christensen, O. B., Formayer, H., Orlowsky, B., and Stepanek, P.: Observational evidence for soil-moisture impact on hot extremes in southeastern Europe, Nat. Geosci., 4, 17–21, doi:10.1038/ngeo1032, 2011.

Hoy, A., Hänsel, S., Skalak, P., Ustrnul, Z., and Bochníček, O.: The extreme European summer of 2015 in a long-term perspective, Int. J. Climatol., doi:10.1002/joc.4751, online first, 2016.

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E. J., Papale, D., Sottocornola, M., Vaccari, F., and Williams, C.: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, J. Geophys. Res., 116, G00J07, doi:10.1029/2010JG001566, 2011.

Leroux, D. J., Kerr, Y. H., Wood, E. F., Sahoo, A. K., Bindlish, R., and Jackson, T. J.: An Approach to Constructing a Homogeneous Time Series of Soil Moisture Using SMOS, IEEE T. Geosci. Remote Sens., 52, 393–405, doi:10.1109/TGRS.2013.2240691, 2014.

Ministerio de Agricultura, Alimentación y Medio Ambiente: Anuario de Aforos Digital 2010–2011, DVD, available at: http://publicacionesoficiales.boe.es/detail.php?id=573028013-0001 (last access: 5 July 2016), 2013.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – A discussion of principles, J. Hydrol., 10, 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

Orth, R. and Seneviratne, S. I.: Introduction of a simple-model-based land surface dataset for Europe, Environ. Res. Lett., 10, 044012, doi:10.1088/1748-9326/10/4/044012, 2015.

Peterson, T. C., Easterling, D. R., Karl, T. R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D., Vincent, L., Heino, R., Tuomenvirta, H., Mestre, O., Szentimrey, T., Salinger, J., Førland, E. J., Hanssen-Bauer, I., Alexandersson, H., Jones, P., and Parker, D.: Homogeneity adjustments of in situ atmospheric climate data: a review, Int. J. Climatol., 18, 1493–1517, doi:10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T, 1998.

Pettitt, A. N.: A Non-Parametric Approach to the Change-Point Problem, J. R. Stat. Soc. C-App., 28, 126–135, doi:10.2307/2346729, 1979.

Project Team ECA&D and Royal Netherlands Meteorological Institute KNMI: Algorithm Theoretical Basis Document (ATBD), Tech. Rep. 10.7, Royal Netherlands Meteorological Institute KNM, available at: http://eca.knmi.nl/documents/atbd.pdf (last access: 5 July 2016), 2013.

Reek, T., Doty, S. R., and Owen, T. W.: A Deterministic Approach to the Validation of Historical Daily Temperature and Precipitation Data from the Cooperative Network, B. Am. Meteorol. Soc., 73, 753–762, doi:10.1175/1520-0477(1992)073<0753:ADATTV>2.0.CO;2, 1992.

Reeves, J., Chen, J., Wang, X. L., Lund, R., and Lu, Q. Q.: A Review and Comparison of Changepoint Detection Techniques for Climate Data, J. Appl. Meteor. Clim., 46, 900–915, doi:10.1175/JAM2493.1, 2007.

Schär, C., Vidale, P. L., Luthi, D., Frei, C., Haberli, C., Liniger, M. A., and Appenzeller, C.: The role of increasing temperature variability in European summer heatwaves, Nature, 427, 332–336, doi:10.1038/nature02300, 2004.

Seneviratne, S. I., Lehner, I., Gurtz, J., Teuling, A. J., Lang, H., Moser, U., Grebner, D., Menzel, L., Schroff, K., Vitvar, T., and Zappa, M.: Swiss prealpine Rietholzbach research catchment and lysimeter: 32 year time series and 2003 drought event, Water Resour. Res., 48, W06526, doi:10.1029/2011WR011749, 2012.

Tallaksen, L. M. and Stahl, K.: Spatial and temporal patterns of large-scale droughts in Europe: Model dispersion and performance, Geophys. Res. Lett., 41, 429–434, doi:10.1002/2013GL058573, 2014.

van der Loo, M.: stringdist: an R Package for Approximate String Matching, R Journal, 6, 111–122, 2014.

Van Lanen, H., Laaha, G., Kingston, D. G., Gauster, T., Ionita, M., Vidal, J.-P., Vlnas, R., Tallaksen, L. M., Stahl, K., Hannaford, J., Delus, C., Fendekova, M., Mediero, L., Prudhomme, C., Rets, E., Romanowicz, R. J., Gailliez, S., Wong, W. K., Adler, M.-J., Blauhut, V., Caillouet, L., Chelcea, S., Frolova, N., Gudmundsson, L., Hanel, M., Haslinger, K., Kireeva, M., Osuch, M., Sauquet, E., Stagge, J. H., and Van Loon, A. F.: Hydrology needed to manage droughts: the 2015 European case, Hydrol. Process., doi:10.1002/hyp.10838, online first, 2016.

Vicente-Serrano, S. M., Beguería, S., López-Moreno, J. I., García-Vera, M. A., and Stepanek, P.: A complete daily precipitation database for northeast Spain: reconstruction, quality control, and homogeneity, Int. J. Climatol., 30, 1146–1163, doi:10.1002/joc.1850, 2010.

von Neumann, J.: Distribution of the Ratio of the Mean Square Successive Difference to the Variance, Ann. Math. Stat., 12, 367–395, 1941.

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S. E., Sullivan, C. A., Liermann, C. R., and Davies, P. M.: Global threats to human water security and river biodiversity, Nature, 467, 555–561, doi:10.1038/nature09440, 2010.

Wijngaard, J. B., Klein Tank, A. M. G., and Können, G. P.: Homogeneity of 20th century European daily temperature and precipitation series, Int. J. Climatol., 23, 679–692, doi:10.1002/joc.906, 2003.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Vol. 100 of International Geophysics Series, Academic Press, Oxford, UK, 3rd Edn., 2011.

Zaidman, M. D., Rees, H. G., and Young, A. R.: Spatio-temporal development of streamflow droughts in north-west Europe, Hydrol. Earth Syst. Sci., 6, 733–751, doi:10.5194/hess-6-733-2002, 2002.