Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

Open Access Earth System Science Data Discussions

Observational gridded runoff estimates for Europe (E-RUN version 1.0)

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)

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 collect an unprecedented collection of river flow observations, combining information from three distinct data bases. Observed monthly runoff rates are first tested for homogeneity and then related to gridded atmospheric variables (E-OBS version 11) using machine learning. The resulting statistical model is then used to estimate monthly runoff rates (December 1950 - December 2014) 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 http://dx.doi.org/10.1594/PANGAEA.845725.

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 Organisation (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 observational estimates of this variable at the pan-European scale. This situation stands in contrast with that of atmospheric variables, for which gridded estimates of e.g. 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 2014. This quantity is referred to as gridded runoff estimate, and eventually contributes to the discharge of large rivers (Gudmundsson and Seneviratne, 2015b, 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

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

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 longterm 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 automated 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 is 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, which are first converted to runoff rates per unit area and subsequently 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 data bases. Two of these are international collections, which contain observations from many European countries (Sections 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 (Section 3.1.3).

3.1.1 The Global Runoff Data Base (GRDB)

The Global Runoff Data Centre (GRDC; http://grdc.bafg.de, last access: 27 Aug 2015) 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 data base at once. Therefore we restricted our order to stations with fulfilling the following set of criteria:

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

Stations should...

1. provide daily observations.
2. be located in the WMO region 6 (Europe).
3. be within the following geographical domain: 25°W - 70°E and 25°N - 75°N.
4. have a minimum of 10 years of observations.

This resulted in a total of 1431 stations which were ordered from the GRDC.

3.1.2 The European Water Archive (EWA)

The European Water Archive (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: 27 Aug 2015) and is also held by the GRDC. A subset of the EWA was selected using the same criteria as for the GRDB (Section 3.1.1), resulting in a total of 3553 stations.

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 is freely accessible online (http://ceh-flumen64.cedex.es/anuarioaforos/default.asp, last access: 27 Aug 2015). As this online platform does not allow to access the full collection at once, we contacted the Spanish authorities and obtained a DVD containing the full data base (Ministerio de Agricultura, Alimentación y Medio Ambiente, 2013). This data base contains among other information streamflow data from 1197 gauging stations.

3.2 Atmospheric Data

Gridded observations of precipitation and temperature were obtained from the E-OBS (version 11) dataset, ranging from 1950 - 2014 (Haylock et al., 2008). 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.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

Open Access Earth System Science Data Discussions

4 Streamflow data selection and pre-processing

4.1 Combining the river flow data bases

4.1.1 Data from Spain

Spanish data are available directly from the Spanish authorities (see Section 3.1.3). Therefore data from Spain were removed from the GRDB and the EWA, and the data stemming from AFD were directly entered into the final collection of European streamflow records.

4.1.2 Linking GRDB and EWA data

The GRDB and the EWA are to some extent populated with data from the same gauging stations. Therefore both data bases need to be linked, in order to avoid duplicated information. Unfortunately, linking the two data bases 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 e.g.: (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 are documented.

4.1.3 Procedure for deduplication and record linkage

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 data base are compared to each other (for record linkage, the entries of two different data bases are compared):

Step 1. Meta data similarity: The first step of deduplication is based on analysing the similarity of the *river name*, the similarity of the *station name* 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:

1. The similarity between the river names and the station names is measured using the Jaro-Winker distance, d_{JW} (Christen, 2012; van der Loo, 2014). The Jaro-Winker distance is a popular measure for evaluating the similarity of

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

Open Access Earth System Science Data Discussions

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.

2. 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 km 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 finally defined as those pairs for which $d_m \leq 0.25$. In 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 river flow time 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, data base entries were classified as either "very likely identical" and "very likely different":

1. *Time series do not overlap* → *very likely different*. The rationale behind this choice is that both time series are independent and may e.g. represent time series before and after renovation of a gauging station.
2. $R^2 > 0.99$ → *very likely identical*. Correlations close to one, indicate identical time series. Minor departures from $R^2 = 1$ may occur e.g. due to rounding errors in the data files.
3. $R^2 < 0.90$ → *very likely different*. This value has been identified experimentally.
4. $d_{JW,river} + d_{JW,station} \leq 0.01$ → *very likely identical.* Small positive values of $d_{JW,river}$ and $d_{JW,station}$ 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 missing*.

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 it was found that small differences in the available records can result in inhomogeneities.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

Open Access Earth System Science Data Discussions

4.1.4 Deduplication and Record linkage results for GRDB and EWA

The deduplication procedure identified 21 very likely duplicates in the EWA and 11 very likely duplicates in the GRDB collection. Linking the deduplicated records from GRDB and EWA resulted in the identification of 4003 unique stations.

4.1.5 A combined European runoff data base (ERDB)

The 4003 linked records from the EWA and the GRDB were combined with the 1197 stations from AFD (Figure 1). The total number of available stations contributing to this European runoff data base (ERDB) is 5200. 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 date with the largest number of available streamflow observations (4178) is Oct 28 1980.

The river flow time series of the combined collection were finally converted into daily runoff rates, expressed in mm day^{-1}.

4.2 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 normally distributed, this rule excludes outliers with a $\approx 2.8 \times 10^7$ occurrence probability.

3. Values with ≥ 10 consecutive equal days are flagged as *suspect*. The rationale underlying this criterion is a tradeoff 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 case of small day-to-day fluctuations).

Figure 3 maps the fraction of days that are flagged as suspect for each station. Note that the fraction of suspect days increases for more arid conditions.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

4.3 Computing 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.

4.4 Homogenity testing

Climate records can exhibit changes which do not reflect real climatic or environmental change. In the context of river flow, such breakpoints could e.g. be related to changes in instrumentation, gauge resaturation, re-calibration 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 e.g. be a sudden shift in the mean, the variance or in 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 (section 4.3) were pre-processed as follows, aiming at de-trending, de-seasonalising and pre-whitening the data:

1. As runoff has usually 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, ..., Decembers 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.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

Open Access Earth System Science Data Discussions

The four tests were subsequently applied to the pre-processed 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.

Figure 4 shows the number of rejected null-hypothesis for each station. Table 1 shows the total number of rejections.

4.5 Assigning monthly runoff rates to the $0.5° \times 0.5°$ 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 $0.5° \times 0.5°$ grid defined by the E-OBS data using the following steps:

1. Select stations:
   (a) Only stations with catchment areas ≤ 500 km^2 are selected. This threshold roughly corresponds to halve 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 that are labelled *useful* in the homogeneity analysis (section 4.4) are selected.
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 2759 selected stations which were assigned to 1036 grid cells, implying that there are on average 2.7 stations assigned to each grid cell. The selected stations are shown in Figure 5. Figure 5 also shows the number of stations in each grid cell as well as the fraction of non-missing months. Figure 6 provides a general overview on the spatial and temporal coverage of the gridded data.

In a last step we selected only those grid cells for the final analysis that covered at least 10 years of observations and had less than 30% of missing values at the monthly resolution. Figure 7 shows the final selection of grid cells.

5 Observational gridded runoff estimates for Europe

5.1 Estimating runoff on a regular spatial grid

The technique 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 a full description of the employed methods we refer to GS15. Following

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

Open Access Earth System Science Data Discussions

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

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

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 (Breiman, 2001), which is a flexible statistical learning tool. For estimating monthly runoff on the $0.5° \times 0.5°$ grid of the E-OBS data the model (Equation (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 Validation

5.2.1 Cross validation

As in GS15 the accuracy of the estimated gridded runoff rates is assessed using two independent cross-validation experiments. For the first experiment, the grid cells with observations were randomly split into ten equally sized sub-samples. 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 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 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 to GS15.

1. The seasonal cycle skill score

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

where $\text{seas}(o_t)$ refers to the long-term mean runoff for each month.

2. The model efficiency

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

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

Open Access Earth System Science Data Discussions

where mean(o_t) refers to the long-term mean of the observation.

3. The relative model bias

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

4. The coefficient of determination (squared correlation coefficient), R^2.

5. The coefficient of determination between the observed and the modelled mean annual cycle, R^2_{CLIM}.

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}

Figure 8 displays the results of both cross-validation experiments. Shown are the spatial patterns as well 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 is 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. Also the relative bias 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 developed runoff estimates 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 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 some 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 both cross validation experiments, highlighting that the estimates can capture more than half of the variance of the anomalies.

5.3 Example applications

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

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

Open Access Earth System Science Data Discussions

5.3.1 Longterm mean runoff statistics

Figure 9 shows the longterm 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 longterm 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.

5.3.2 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 longterm 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 through the 30 year standard deviation of each month.

Figure 10 shows the standardised runoff anomalies for three well documented events with exceptionally dry conditions. Drought conditions in 1976 were among the most severe in Europe throughout the course of the 20^{th} 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) and 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).

6 Data availability

The data are publicly available in NetCDF format (Gudmundsson and Seneviratne, 2015a) and can be downloaded from http://www.pangaea.de/. A table documenting the considered stations is available as supplementary information and described in Appendix A.

7 Conclusions

In conclusion we presented an observational dataset that provides monthly pan-European runoff estimates and ranges from December 1950 to December 2014. The data is 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 up-scaled 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 and show that the accuracy of the data is in line with previous results (GS15), highlighting the robustness of the used estimation technique. The two example

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

applications highlight the utility of the newly developed pan-European runoff estimates, both for climatological assessments, as well as for drought monitoring. These examples highlight 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.

Appendix A: Meta data of the considered stations

The streamflow observations collected in the ERDB (Section 4.1.5) 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 data bases (Section 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 data base 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 meters a.s.l.

area Catchment area in km^2.

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

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

Part B: Record linkage results summarises the results of the record linkage procedure described in Section 4.1.3. Note: If both the fields **EWA.no** and **GRDB.no** contain values, this indicates that the records of EWA and GRDB have been linked. In this case only the data base specified in the field **source.data.base** was used to generate the value:

source.data.base Acronym for one of the data bases listed in Section 3.1. This field corresponds to the data base that was used to generate the entry.

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

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

Open Access Earth System Science Data Discussions

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

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

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

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

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

latlon.bin.dist The value of d_G in km, if more than one data base was used to generate the record.

mean.dist The value of d_m (equation 2), if more than one data base was used to generate the record.

Part C: QC and homogeneity testing summarises the results of the QC (Section 4.2) and the homogeneity assessment (Section 4.4).

fraction.suspect The fraction of suspect days.

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"`: Not sufficient data (less than 5 years).

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.

Acknowledgements. The support of the ERC DROUGHT-HEAT project (contract no 617518) and the DROUGHT-R&SPI (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 effort to assemble the European Water Archive (EWA) by the UNESCO IHP VII FRIEND programme the data collection and management by the GRDC are gratefully acknowledged.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

Open Access Earth System Science Data Discussions

References

Alexandersson, H.: A homogeneity test applied to precipitation data, Journal of Climatology, 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, L18 405, doi:10.1029/2005GL023574, http://dx.doi.org/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, Bull. Amer. Meteor. Soc., 95, 1431–1443, doi:10.1175/BAMS-D-13-00047.1, 2014.

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

Buishand, T.: Some methods for testing the homogeneity of rainfall records, Journal of Hydrology, 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, Journal of Hydrology, 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, Journal of Hydrology, 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, JAWRA Journal of the American Water Resources Association, doi:10.1111/jawr.12149, 2013.

Costa, A. and Soares, A.: Homogenization of Climate Data: Review and New Perspectives Using Geostatistics, Mathematical Geosciences, 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, Journal of Climatology, 2013, 15, 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. Hydrometeor, 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, 10.1038/nclimate1329, 2012.

Gudmundsson, L. and Seneviratne, S. I.: E-RUN version 1.0: Observational gridded runoff estimates for Europe, link to data in NetCDF format (68 MB), doi:10.1594/PANGAEA.845725, 2015a.

© Author(s) 2016. CC-BY 3.0 License.

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

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, Hydrological Processes, 25, 1191–1200, doi:10.1002/hyp.7794, 2011.

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, D20 119, 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, Nature Geosci, 4, 17–21, doi:10.1038/ngeo1032, 2011.

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.

Ministerio de Agricultura, Alimentación y Medio Ambiente: Anuario de Aforos Digital 2010 - 2011, DVD, http://publicacionesoficiales.boe.es/detail.php?id=573028013-0001, 2013.

Orth, R. and Seneviratne, S. I.: Introduction of a simple-model-based land surface dataset for Europe, Environmental Research Letters, 10, 044 012, http://stacks.iop.org/1748-9326/10/i=4/a=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, International Journal of Climatology, 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, Journal of the Royal Statistical Society. Series C (Applied Statistics), 28, pp. 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, http://eca.knmi.nl/documents/atbd.pdf, 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, Bull. Amer. Meteor. 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. Climatol., 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, W06 526, doi:10.1029/2011WR011749, http://dx.doi.org/10.1029/2011WR011749, 2012.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

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

van der Loo, M.: stringdist: an R Package for Approximate String Matching, The R Journal, 6, 111–122, http://journal.r-project.org/archive/2014-1/loo.pdf, 2014.

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, International Journal of Climatology, 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, The Annals of Mathematical Statistics, 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, International Journal of Climatology, 23, 679–692, doi:10.1002/joc.906, 2003.

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

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

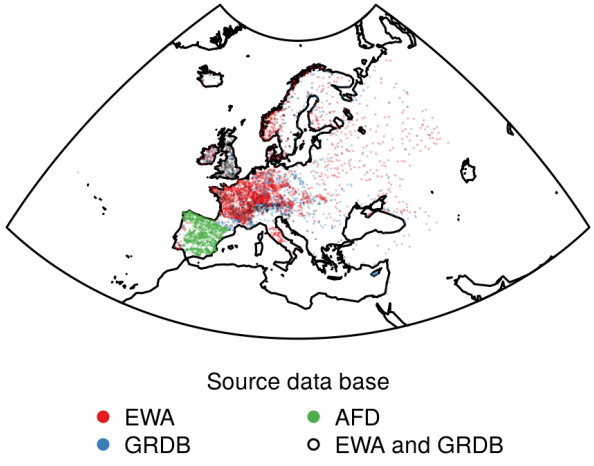

Figure 1. Locations of streamflow stations, stemming from the three considered data collections. Records from the EWA and the GRDB that where identified as *vex likely identical* are highlighted by black circles

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

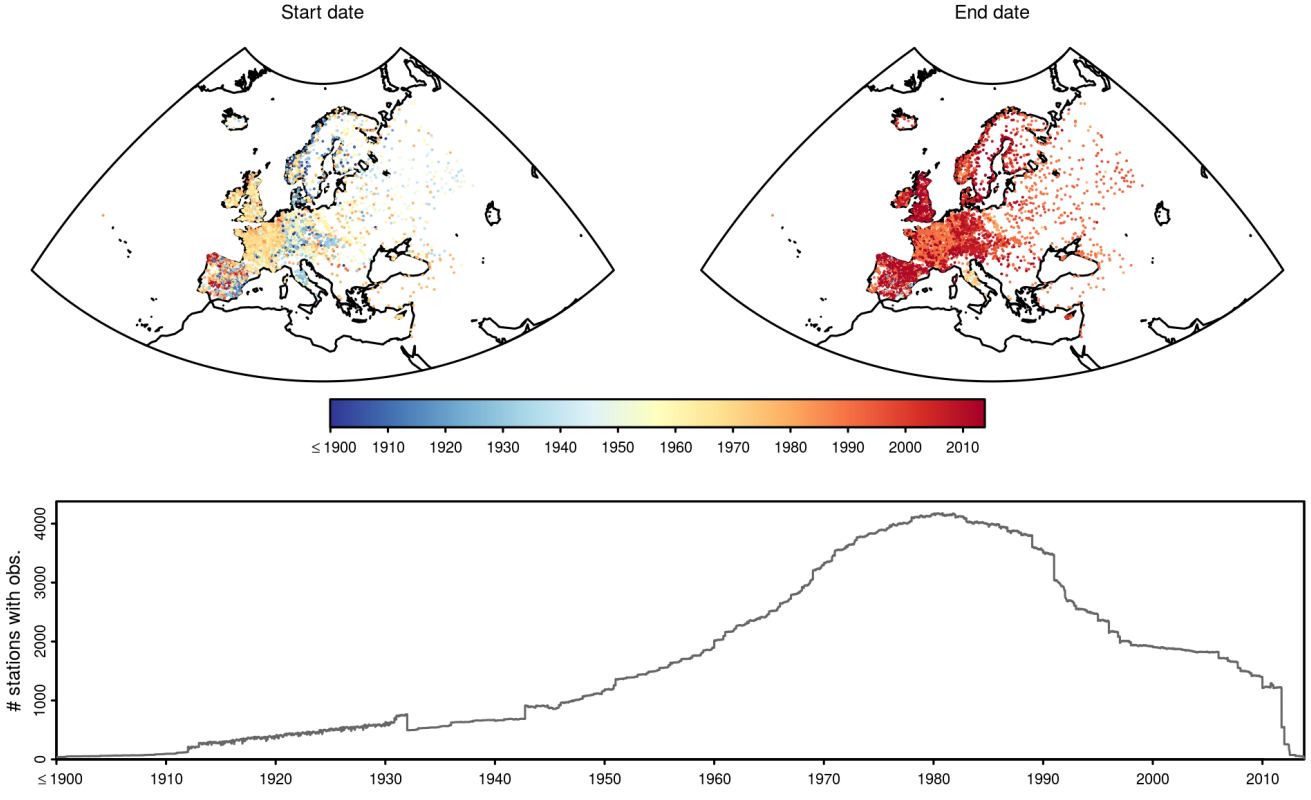

Figure 2. Spatial and Temporal coverage of available streamflow observations. The top row shows the date of the first and the date of the last available observation at each station. The bottom panel shows the total number of stations with observations for each day.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

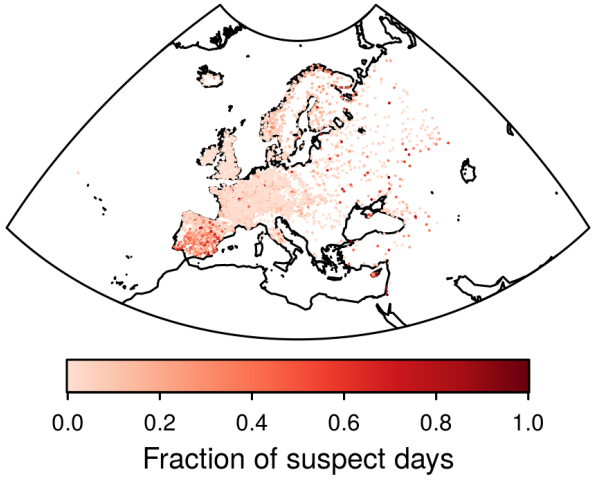

Figure 3. Fraction of days that are flagged as *suspect*.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

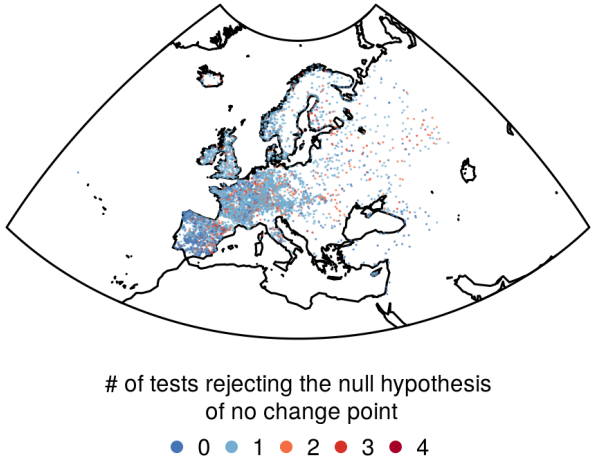

Figure 4. Homogeneity testing: Number of tests that reject the null-hypothesis of no breakpoint at each station considered at the 1% level. Stations marked blue (zero or one rejection) are considered *useful*. Stations marked red (more rejections) are considered *suspect*.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

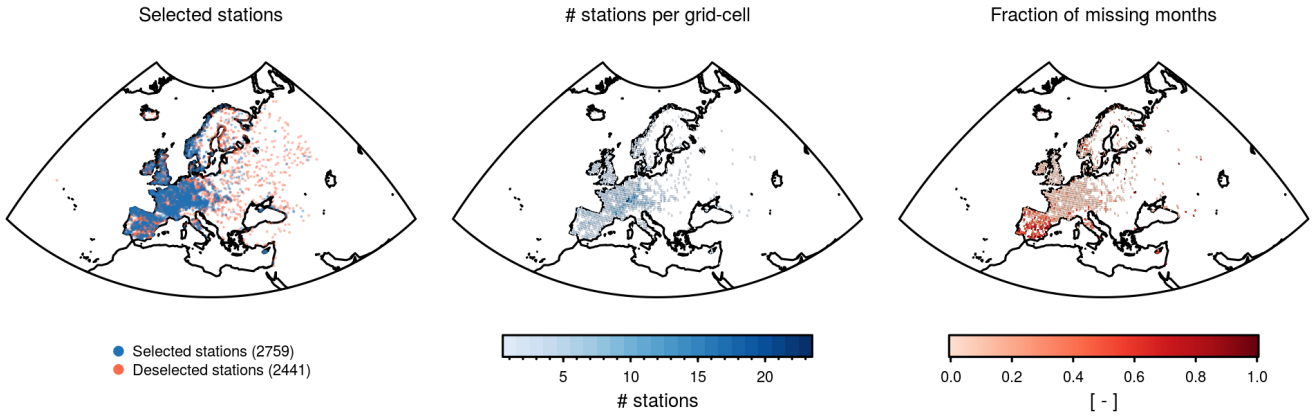

Figure 5. Assigning stations to the $0.5^\circ \times 0.5^\circ$ grid cells defined by the E-OBS data: Left: Selected stations, fulfilling all selection criteria (see Section 4.5). Centre: Number of stations per grid cell. Right: Fraction of months with no or not sufficient data.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

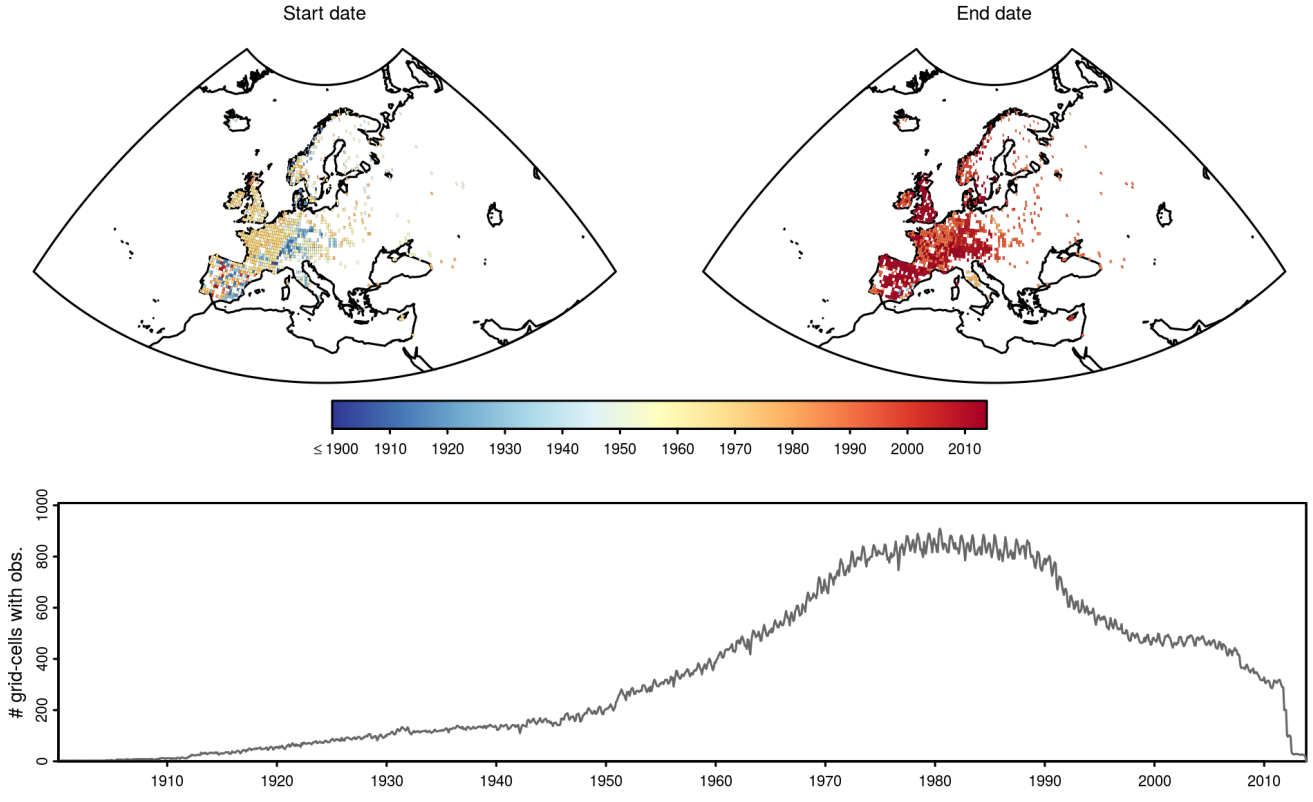

Figure 6. Spatial and Temporal coverage after assigning the monthly runoff series to the $0.5^\circ \times 0.5^\circ$ defined by the E-OBS data. The top row shows the date of the first and the date of the last available observation at each station. The bottom panel shows the total number of stations with observations for each month.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

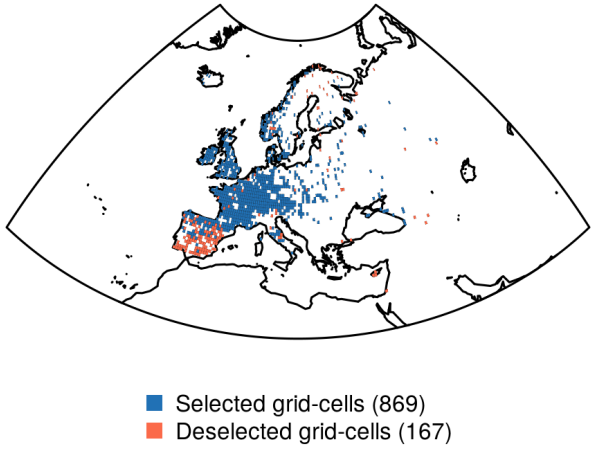

Figure 7. Final selection of grid cells with observations. Only grid cells that have at least 10 years of observations and have less than 30% missing values are selected.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

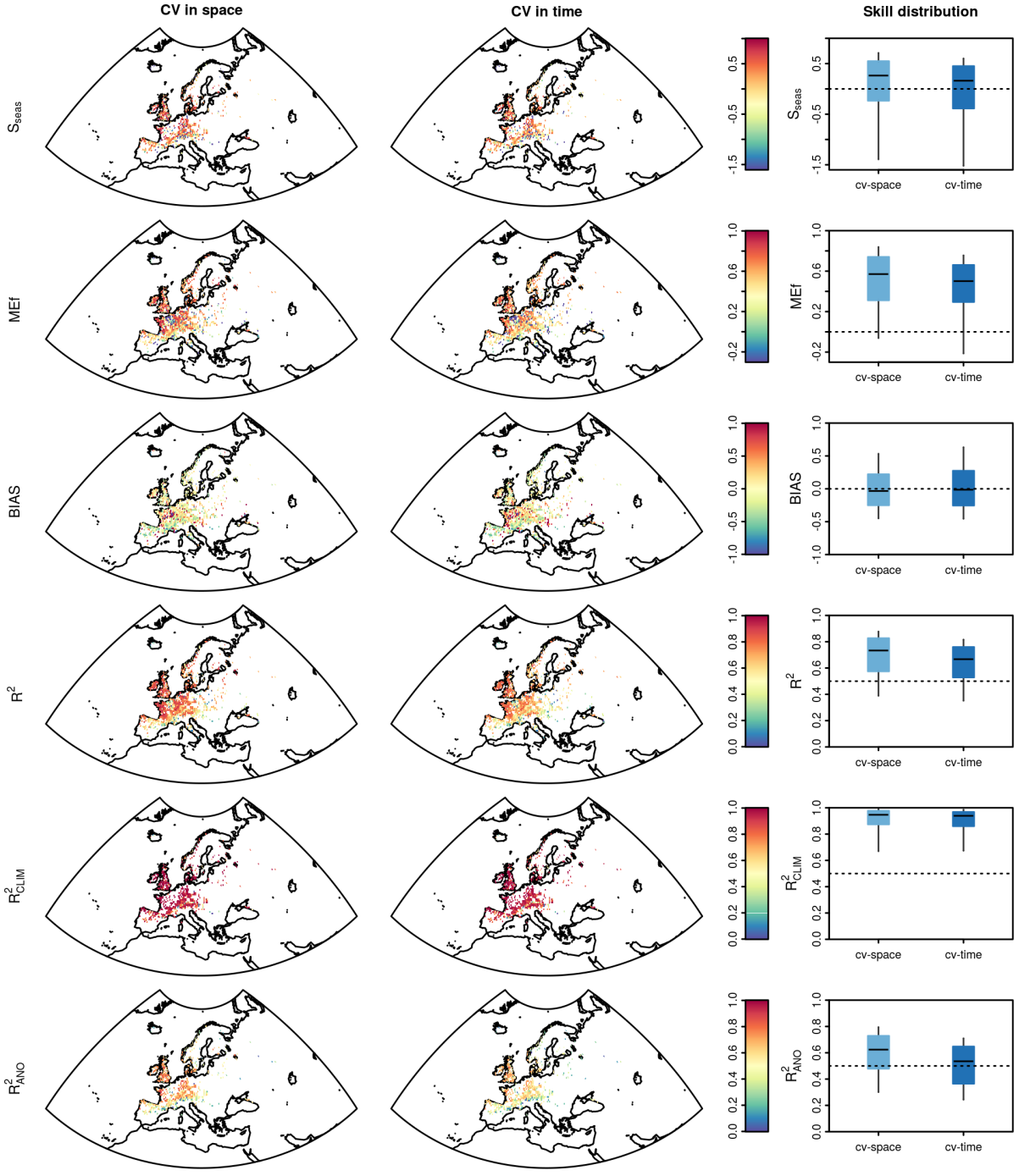

Figure 8. Spatial distributions and boxplots (whiskers: 0.1 and 0.9 percentiles, box: inter quartile range, bar: median) of all considered performance metrics and for both cross validation experiments.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

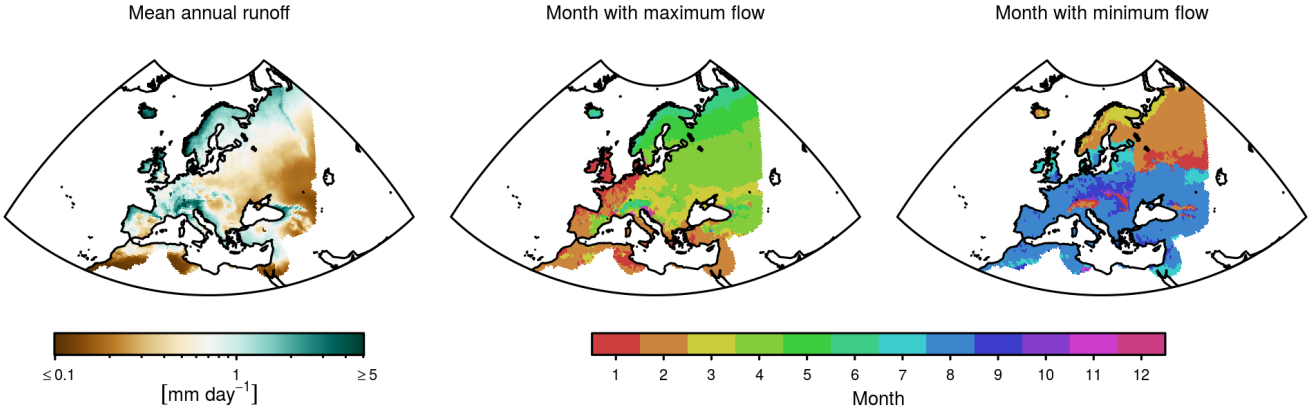

Figure 9. Longterm mean of the presented gridded runoff field as well as the month of the maximum and the minimum of the mean annual cycle.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

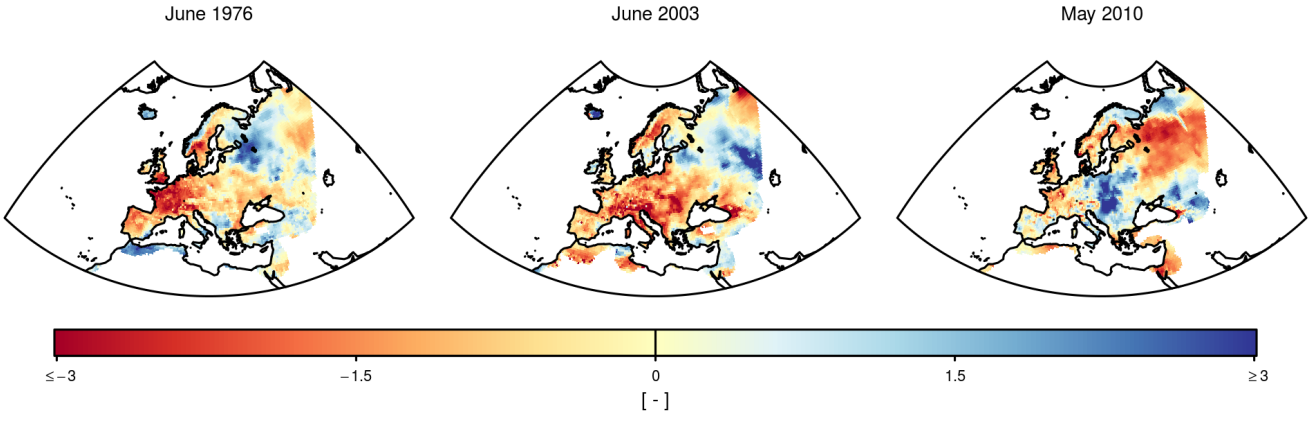

Figure 10. Standardised runoff anomalies for three selected drought events in Europe.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 18 January 2016
© Author(s) 2016. CC-BY 3.0 License.

Open Access Earth System Science Data Discussions

Table 1. 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*.

# rejections	0	1	2	3	4
# stations	1045	3401	614	140	0