A 500-year annual runoff reconstruction for 14 selected European catchments

Sadaf Nasreen[1], Markéta Součková[1], Mijael Rodrigo Vargas Godoy[1], Ujjwal Singh[1], Yannis Markonis[1], Rohini Kumar[2], Oldrich Rakovec[1,2], and Martin Hanel[1]

[1]Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Praha-Suchdol 16500, Czech Republic
[2]UFZ-Helmholtz Centre for Environmental Research, Leipzig 04318, Germany

Correspondence: Martin Hanel (hanel@fzp.czu.cz)

Abstract. Since the beginning of this century, Europe has been experiencing severe drought events (2003, 2007, 2010, 2018 and 2019) which have had ~~an~~ adverse impacts on various sectors, such as agriculture, forestry, water management, health, and ecosystems. During the last few decades, projections of the impact of climate change on hydroclimatic extremes were often capable of reproducing changes in the characteristics of these extremes. Recently, the research interest has been extended to include reconstructions of hydro-climatic conditions ~~, so as~~ to provide historical context for present and future extremes. While there are available reconstructions of temperature, precipitation, drought indicators, or the 20^{th} century runoff for Europe, ~~long-term~~ multi-century annual runoff reconstructions are still lacking~~(e. g, monthly or daily runoff series for short periods are commonly available). Therefore, we considered~~ . In this study, we have used reconstructed precipitation and temperature ~~fields for the period between 1500 and 2000 together with reconstructed scPDSI, natural proxy data, and observed runoff over 14 European catchments to calibrate and validate the semi-empirical hydrological model GR1A and~~ data, Palmer Drought Severity Index and available observed runoff across fourteen European catchments in order to develop annual runoff reconstructions for the period 1500–2000 using two data-driven ~~models (Bayesian recurrent and long short-term memory neural network). The validation of input precipitation fields revealed an underestimation of the variance across most of Europe. On the other hand, the data-driven models have been proven to correct this bias in many cases, unlike the semi-empirical hydrological modelGR1A~~and one conceptual lumped hydrological model. The comparison to observed ~~historical~~ runoff data has shown a good match between the reconstructed and observed runoff and ~~between the runoff~~ their characteristics, particularly deficit volumes. On the other hand, the validation of input precipitation fields revealed an underestimation of the variance across most of Europe, which is propagated into the reconstructed runoff series. The reconstructed runoff is available via figshare, an open source scientific data repository, under the DOI https://doi.org/10.6084/m9.figshare.15178107, (Sadaf et al., 2021).

1 Introduction

Global warming has impacted numerous land surface processes (Reinecke et al., 2021) over the last few decades, resulting in more severe droughts, heat waves, floods, and other extreme ~~weather~~events. Droughts, in particular, pose a serious threat to Europe's ~~hydrology~~water resources. The flow of many rivers is greatly hampered by prolonged droughts, which restrain the availability of fresh water for agriculture and domestic use. For example, the 2003 drought significantly reduced European river flows by approximately 60 to 80% relative to the average. Likewise, the annual flow levels ~~of~~ at several river gauges have decreased by 9 to 22% over the last decade ~~(Krysanova et al., 2008; Middelkoop et al., 2001; Uehlinger et al., 2009; Su et al., 2020)~~ (Middelkoop et al., 2001; Krysanova et al., 2008; Uehlinger et al., 2009; Su et al., 2020) due to a lack of rainfall and a warmer climate. ~~For~~ In the last 40 years, low river flows ~~have rendered~~ complicated water-power generation ~~impossible~~ in the UK, resulting in a 45£ million loss each year ~~. However, there has been less focus on the water deficit in streams, rivers and other reservoir's, the so-called hydrological droughts (Van Loon, 2015). Most importantly, runoff , which supplies rivers with a significant amount of water , is potentially valuable for water security management. The challenging element is that the nonlinear behavior of hydro-climate fluctuations cannot be explicitly interpreted using data from the most recent centuries (Markonis and Koutsoyiannis, 2016). Continuous records of runoff /discharge series are no longer available , including various~~ (Anonymous, 2020).

While runoff is a key element related to water security, it is challenging to interpret recent hydroclimate fluctuations (multi-year droughts in particular) considering observed runoff records (Markonis and Koutsoyiannis, 2016; Hanel et al., 2018), which are in general seldom available for years prior to 1900. In this way, the community does not have runoff information on various severe multi-year droughts and pluvial periods~~. On the other hand, proxy-based~~ , which can be assessed only indirectly using (typically seasonal or annual) reconstructions ~~are alternatively used, considering~~ based on various proxy data, such as past tree-rings ~~(Cook et al., 2015; Casas-Gómez et al., 2020; Tejedor et al., 2016; Kress et al., 2010; Nicault et al., 2008)~~(Nicault et al., 2008; Kress
, speleothem (Vansteenberge et al., 2016), ice cores, sediments (Luoto and Nevalainen, 2017) and documentary and instrumental evidence ~~(Pfister et al., 1999; Dobrovolný et al., 2010; Wetter et al., 2011; Brázdil and Dobrovolný, 2009) to pinpoint extreme events and the detection of climate change~~(Pfister et al., 1999; Brázdil and Dobrovolný, 2009; Dobrovolný et al., 2010; Wetter et a
.

The majority of existing reconstructions focus on temperature ~~(Dobrovolný et al., 2010; Luterbacher et al., 2004; Emile-Geay et al., 201~~
~~, precipitation (Boch and Spötl, 2011; Wilson et al., 2005; Murphy et al., 2018; Wilhelm et al., 2012) or drought (Büntgen et al., 2010; Brá~~
~~and flood reconstructions (Swierczynski et al., 2012; Wetter et al., 2011)~~(Luterbacher et al., 2004; Xoplaki et al., 2005; Casty et al., 2005;
, precipitation (Wilson et al., 2005; Boch and Spötl, 2011; Wilhelm et al., 2012; Murphy et al., 2018) or droughts (Büntgen et al., 2010; Kr
and floods (Wetter et al., 2011; Swierczynski et al., 2012). A few studies have been conducted for the reconstruction of runoff-drought deficit series ~~(Hanel et al., 2018; Moravec et al., 2019; Hansson et al., 2011; Kress et al., 2014; Martínez-Sifuentes et al., 2020) . However~~(Hansson et al., 2011; Kress et al., 2014; Hanel et al., 2018; Moravec et al., 2019; Martínez-Sifuentes et al., 2020). However, these studies are either local or regional, or cover a relatively short ~~time~~ period. As an example~~of Hansson et al. (2011), which~~ , Hansson et al. (2011) introduced a runoff series for the Baltic Sea ~~only, between~~ from 1550 ~~and~~ to 1995 years using temperature

and atmospheric circulation indices. Similarly, Sun et al. (2013) has used tree-ring proxies to reconstruct runoff in the Fenhe River Basin in China's Shanxi region over the last 211 years. As another example, Caillouet et al. (2017) provides a 140-year dataset of reconstructed streamflow over 662 natural catchments in France since 1871 using the GR6J hydrological model, highlighting several well-known extreme low flow events. A multi ensemble modeling approach using GR4J has been applied by Smith et al. (2019) to develop UK-based historical river flows and examine the potential of reconstruction for capturing peak and low flow events from 1891 to 2015.

~~Conversely, the~~ The available reconstructed precipitation and temperature series (or fields) can be used to reconstruct runoff with ~~a hydrological model (Tshimanga et al., 2011; Armstrong et al., 2020). This can be achieved through a~~ hydrological (process-based~~model of varying complexity, with the advantage of following~~) models (Tshimanga et al., 2011; Armstrong et al., 2020) respecting general physical laws~~– e.g.,~~ , such as preserving mass balance ~~, etc.Physical based models: MIKE SHE(Im et al., 2009) and VELMA(Laaha et al., 2017)~~ (e.g. MIKE SHE; Im et al., 2009 or VELMA; Laaha et al., 2017) or data-driven methods ~~, such as~~ which are able to capture complex non-linear relationships (for instance support vector machines ~~(Ji et al., 2021; Zuo et al., 2020) ,~~ (Zuo et al., 2020; Ji et al., 2021); artificial neural networks ~~(ANNs; Kwak et al., 2020; Hu et al., 2018; Senthil Kumar et al., 2005) , random forests (Breiman, 2001; Contreras et al., 2021), and Shannon entropy (Thiesen et al., 2019) are able to capture complex non-linear relationships~~ANNs; Senthil Kumar et al., 2005; Hu et al., 2018; Kwak et al., 2020; random forests (Ghiggi et al., 2019; Li et al.). While the lack of physical constraints in the data-driven models limits their application under ~~contrasting (changing)~~ changing boundary conditions (in comparison with those of the model training period), their advantage is that they can often directly use biased reconstructed data as an input series.

The objective of the present study is to provide a ~~long-term, hydrological reconstruction for the Central~~ multi-century annual runoff reconstruction for fourteen European catchments, utilizing the available ~~gridded~~ precipitation (Pauling et al., 2006) and temperature (Luterbacher et al., 2004) reconstructions ~~, natural proxies (Ljungqvist et al., 2016) and other long-term historical data sources~~and Old World Drought Atlas Self-calibrated Palmer Drought Severity Index (scPDSI) reconstruction (Cook et al., 2015). Specifically, we ~~use a combination of a conceptual~~ assessed a conceptual lumped hydrological model (GR1A; Mouelhi et al., 2006) and two data-driven models ~~(Chen et al., 2020; Okut, 2016) to simulate the annual evolution of runoff~~ (Long Short Term Memory neural network (LSTM; Chen et al., 2020) and Bayesian Regularized Neural Network (BRNN; Okut, 2016) for annual runoff simulation over the period 1500–2000. ~~We pay particular attention to low flows during drought years. Using long-term data on climatic conditions and runoff may provide an efficient technique of visualizing droughts and low flow periods.~~

The structure of the paper is as follows: the considered hydroclimatic reconstructions, ~~natural proxies~~ drought indicator and observed data are ~~described in Section~~introduced in Sect. 2. In ~~Section~~Sect. 3, we ~~introduce the data selection and~~ describe the data pre-processing, hydrological and data-driven models~~and~~ , the drought identification ~~. The reconstructed input fields~~methodology and goodness-of-fit assessment. The accuracy of the employed precipitation and temperature reconstructions, as well as ~~our runoff simulations considering four input data combinations (precipitation, temperature, raw proxy and drought indicator) and two data-driven approaches, together with the hydrological model~~ the derived runoff simulations are evaluated

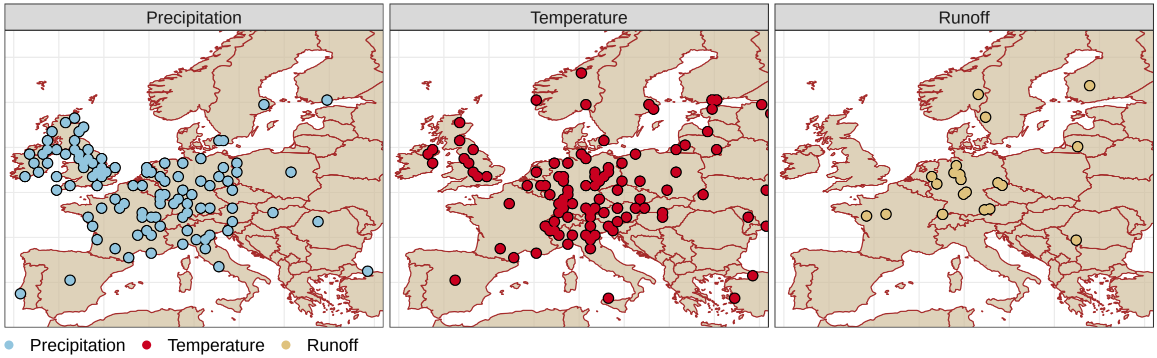


Figure 1. Spatial distribution of the observed GHCN precipitation and temperature stations ~~,~~ and GRDC ~~discharge~~ runoff gauges~~and proxies for precipitation and temperature~~.

in ~~Section~~Sect. 4. Finally, we ~~provide certain guidelines on~~ summarize the advantages and limitations of reconstructed datasets in the concluding section 5.

2 Data

Herein, we used precipitation (Pauling et al., 2006) and temperature (Luterbacher et al., 2004) reconstructions for the past half-millennium ~~,~~ and scPDSI drought indicator data from the Old World Drought Atlas (Cook et al., 2015)~~, and natural proxies (Ljungqvist et al., 2016). For benchmarking, we relied on~~ . For validation the reconstructed datasets, we considered the observational data records of precipitation and temperature (Menne et al., 2018), as well as runoff from the Global Runoff Data Center (GRDC; ~~Fekete et al. 1999). The data-sets~~ Fekete et al., 1999), which was also used for model calibration. The datasets are summarized in Table 1 and are described in more detail below.

2.1 ~~Hydroclimatic reconstructions~~Precipitation

We used reconstructed seasonal precipitation ~~and temperature gridded data~~ (0.5° x 0.5°) over Europe (30.25° ~~N-70~~N–70.75°N / 29.75°~~W-39~~W–39.75°E) from 1500 to ~~the present day. To this end Pauling et al. 2006, reconstructed~~ 2000 years. Reconstructed precipitation (P) was ~~done by applying~~ derived by Pauling et al. (2006) through principal component regression ~~to documented evidence~~ based on documented evidences (i.e., memoirs, annals, newspapers), speleothem proxy records (Proctor et al., 2000) and tree-ring chronologies from the International Tree-Ring Data Bank (ITRDB~~). ~~; Jeong et al., 2021) .

2.2 Temperature

Reconstructed temperature (T) is obtained from Luterbacher et al. (2004) which relies on historical records and seasonal natural proxies (i.e., ice cores from Greenland and tree-rings from Scandinavia and Siberia). ~~We refer to these data sets~~ Reconstructed

Table 1. Summary of considered ~~data sets~~datasets

Reference	Domain	Temporal coverage ~~(CE)~~*(CE)	Spatial resolution	Variables
Pauling et al. (2006)	Europe	1500–2000	0.5° x 0.5°	~~Seasonal Precipitation~~seasonal precipitation
Luterbacher et al. (2004)	Europe	1500–2000	0.5° x 0.5°	~~Seasonal Temperature~~seasonal temperature
Menne et al. (2018)	Global	1760–2010	26000 point stations	~~Mean Temperature~~monthly mean temperature
Menne et al. (2018)	Global	1760–2010	20590 point stations	~~Mean Precipitation~~monthly mean precipitation
~~Ljungqvist et al. (2016) Northern Hemisphere 800–2005 130 point stations Temperature~~~~Ljungqvist et al. (2016) Northern Hemisphere 800–2005 197 point stations Hydro-proxies~~Cook et al. (2015)	Europe	0–2012	0.5° x 0.5°	summer Palmer Drought Severity Index

*Common Era

temperature data was available at the same spatial and temporal resolution as precipitation (see Table 1). We refer both of these datasets as reconstructed forcings ~~. Additionally~~or reconstructed precipitation/temperature fields.

2.3 Self-calibrating Palmer Drought Severity Index (scPDSI)

In addition, we used data from the Old World Drought Atlas (OWDA; ~~Cook et al. 2015~~Cook et al., 2015) which contains information regarding moisture conditions across Europe, specifically the self-calibrated Palmer Drought Severity Index (scPDSI) using summer-related, tree-ring proxies ~~for the period from~~ over the period 0 to 2012 CE.

2.4 ~~Other hydro-climate proxy information~~

~~We also included a raw proxy series for hydroclimatic variables by Ljungqvist et al. 2016 in our analysis, as mentioned in Supplementary tables (S1 and S2). We considered 20 precipitation related proxies consisting of three tree-ring widths, eight lake sediments, five peat bogs, two speleothems and two peat humidifications. Similarly, there were 17 temperature-based proxies including six tree-rings, three ice cores, three lake sediments, two speleothems and three written records. These proxies are not evenly distributed across Europe (Fig. 1). The available series, typically spanning hundreds of years, were restricted to 1500 – 2000 in our study. Data standardization was conducting by subtracting the mean and dividing by standard deviation (both calculated considering the time-series after 1900). Missing values were calculated by linear approximation and, in this way, we obtained a consistent set of proxy information for each (annual) time step. It has been previously established that these proxies correlate well with climatic variables, such as precipitation and temperature (Riechelmann and Gouw-Bouman, 2019) .~~

2.4 The Global Historical Climatology Network (GHCN)

The GHCN dataset (GHCN; ~~(Peterson and Vose, 1997)~~Peterson and Vose, 1997) – one of the largest observational databases, collated by the National Oceanic and Atmospheric Administration (NOAA~~,~~ ; Quayle et al., 1999) – was used to verify the accuracy of the precipitation and temperature reconstructions. The GHCN-m (version 2) ~~data-set~~ dataset contains observed temperature, rainfall and pressure data from 1701 to 2010. Data for the majority of stations are, however, available after 1900. GHCN-m precipitation and temperature from GHCN V2, as well as from the new GHCN V4 version were included in the preliminary analysis (Menne et al., 2012). We found 113 precipitation and 144 temperature stations within the European domain (see Fig. 1) with records dating back earlier than 1875. Most stations are geographically concentrated in Central Europe, and few stations are located in the eastern and northern areas of Europe (see ~~Fig.~~ Table 2). These data, hereafter, are referred to as the GHCN data.

2.5 Observed runoff

The Global Runoff Data Center (GRDC; www.bafg.de/GRDC/EN/Home/homepage_node.html) provides data for more than 2780 gauging stations in Europe, with the oldest records starting from 1806. ~~The runoff series from the GRDC were selected~~

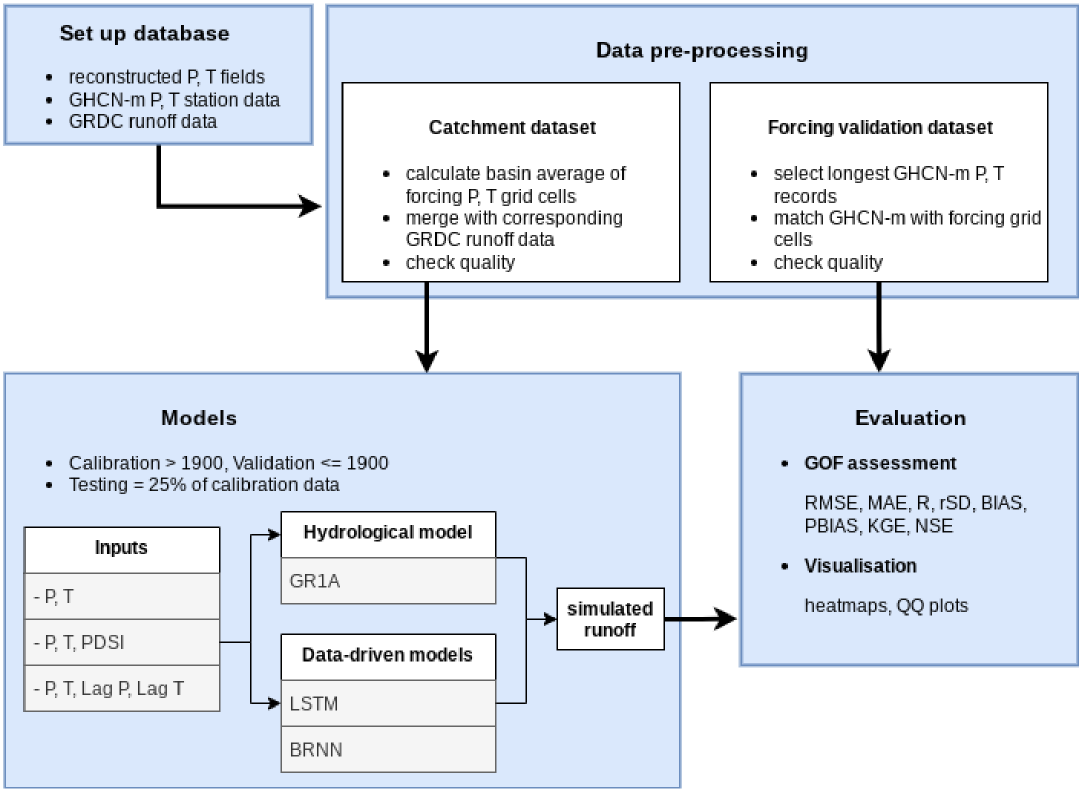


Figure 2. A schematic overview of the study work-flow.

~~based on the condition of data availability,~~ Only the GRDC runoff time series with at least 25 years ~~prior to 1900.~~ of data prior to 1900 were selected. In total, there were 21 such stations predominantly available in Central Europe: 11 in Germany, two in France, two in Switzerland, one in the Czech Republic, one in Sweden, one in Finland, one in Lithuania and one in Romania (see Fig. 1). These stations cover 12 European river basins (Rhine, Loire, Elbe, Danube, Wesser, Main, Glama, Slazach, Nemunas, Gota Alv, Inn and Kokemaenjoke), with areas ranging from nearly 6 100 km^2 (Kokemaenjoki, Muroleenkoski, Finland) to 576 000 km^2 (Danube, Orsova, Romania). The mean annual discharge (Q_{mean}) varies from 50 m^3s^{-1} to 5 600 m^3s^{-1} and spans different time periods for each catchment.

The most extensive records were available in ~~KRV~~ Sweden and Dresden, containing the longest discharge series of 212 and 208 years, respectively. The gauging station in Köln also provided 195 years of data for the Rhine River. Note that some of the gauging stations are located ~~in close proximity~~ nearby and therefore have a greater degree of similarity in ~~relation to the~~ their runoff time-series (e.g., two stations in Basel, Rhine). Detailed information relating to all ~~the selected stations and their silent characteristics are~~ selected stations is provided in Table 2.

Table 2. ~~Salient feature of selected~~ Selected study catchments.

~~station~~Station	~~river~~River	GRDCno	~~latitude~~ Latitude [°N]	~~longitude~~ Longitude [°E]	~~drainage~~ ~~Area~~ Drainage area [km^2]	~~mean~~ Mean annual discharge [m^3s^{-1}]	~~start~~ Start year	~~length~~ ~~(year)~~Length [year]
Orsova, RO	Danube	6742200	44.7	22.42	576232	5602	1840	151
Decin, CZ	Elbe	6140400	50.79	14.23	51123	309	1851	150
Dresden, DE	Elbe	6340120	51.05	13.73	53096	332	1806	208
Elverum, NO	Gloma	6731401	60.88	11.56	15426	251	1871	44
Vargoens KRV, SW	Gota Alv	6229500	58.35	12.37	46885.5	531	1807	212
Wasserburg, DE	Inn	6343100	48.05	12.23	11983	354	1827	177
Muroleenkoski,FI	Kokemaenjoki	6854104	61.85	23.910	6102	53.1	1863	155
Blois, FR	Loire	6123300	47.58	-0.86	38240	362	1863	117
Montjean, FR	Loire	6123100	47.58	1.33	110000	911	1863	117
Schweinfurt-Neuer Hafen	Main	6335301	50.03	10.22	12715	103	1845	156
Weurzburg, DE	Main	6335500	49.79	9.92	14031	108	1824	177
Smalininkai, LT	Nemunas	6574150	55.07	22.57	81200	531	1812	185
Basel Rheinhalle, CH	Rhine	6935051	47.55	7.61	35897	1043	1869	140
Basel Schifflaende, CH	Rhine	6935052	47.55	7.58	35905	1042	1869	127
Köln, DE	Rhine	6335060	50.93	6.96	144232	2085	1817	195
Rees, DE	Rhine	6335020	51.75	6.39	159300	2251	1815	183
Burgausen, DE	Salzach	6343500	48.15	12.83	6649	258	1827	174
Hann-Münden DE	Wesser	6337400	51.42	9.64	12442	109	1831	182
Bodenwerder, DE	Wesser	6337514	51.97	9.51	15924	145	1839	175
Vlotho DE	Wesser	6337100	52.17	8.86	17618	170	1820	194
Intschede, DE	Wesser	6337200	52.96	9.12	37720	320	1857	154

2.6 Study area

In the first ~~section~~ part of the study, the ~~analysis is performed~~ grid-based reconstruction of precipitation and temperature was verified against the available GHCN data across the European region bounded by (30.25° ~~N-70.75~~N – 70.75° N / 29.75° ~~W-39.75~~W – 39.75° E)~~, in which the grid-based reconstruction of precipitation and temperature was verified against the observation data. In the second section, we focus~~. The second part focuses on 21 specific Central European catchments, corresponding to the available long-term GRDC discharge records. The study area and the observational data of the hydroclimatic variables are shown in ~~Figure~~Fig. 2.

3 Methods

This section is divided into three parts. The first part describes the ~~selection and~~ pre-processing of the reconstructed ~~forcing~~ forcings (i.e., precipitation and temperature) for validation across Europe and the preparation of data for runoff simulation in ~~several catchments . The hydrologic~~ 21 catchments (Sect. 3.1). Used hydrologic (Sect. 3.2) and data-driven models ~~used,~~ (Sect. 3.3) for runoff simulation are introduced in the second part. Finally, ~~we~~ section 3.4 describe the methods for the evaluation of simulated runoff ~~(including drought identification)~~and reconstructed forcings and section 3.5 presents the methods to identify annual runoff droughts.

3.1 Data pre-processing

We prepared two datasets. The first ~~consists of reconstructed forcings and the corresponding~~ was used for forcings validation and consists of observed GHCN data for all available European stations with long records (see ~~Section 2.4) . We considered the selected GHCN stations and data from the~~ Sect. 2.4) and values of corresponding grid cells ~~of~~ from the reconstructed forcings ~~for this forcing validation exercise. To understand how the reconstructed forcings match the GHCN data across time scales, we aggregated both the reconstructed forcings and the GHCN data from seasons to 1, 2, …30 years and calculated various goodness-of-fit (GOF) metrics (see further in Appendix A1).~~ dataset.

The second dataset ~~represents the data of~~ was created as the basis for runoff reconstruction containing the observed runoff data for 21 selected catchments ~~and consists of reconstructed forcings and the proxy data and runoff for the calibration and validation of individual catchments (Fig. 2) .~~ (Table 2) and corresponding input variables into the models (the GR1A hydrologic model, the BRNN and LSTM data-driven models) that were used for 1500–2000 runoff simulation. We have considered several input variables (Table 3) – reconstructed precipitation and temperature and Old World Drought Atlas scPDSI. It is worth noting that other natural proxies have also been considered within the models, however, since the added value was negligible, we do not present these data and results here. The catchment average precipitation~~and temperature~~ , temperature and scPDSI were estimated from the ~~reconstructed forcings~~ corresponding (gridded) datasets by averaging the ~~grid cells covering the specific catchment boundary. Similarly, we calculated the average catchment PDSI from the OWDA, and also selected the raw proxy data from inside the catchment or within a 100 km buffer around the catchment.~~relevant grid cells over the catchments. Data were split into two parts: calibration (1900–2000) and validation (<=1900) to assess the model's accuracy and to select an appropriate model. The data pre-processing, model selection, and evaluation of the models are depicted in Fig. 2.

3.2 Hydrologic model (GR1A)

~~To simulate runoff in each catchment, we~~ We applied the annual time-scale hydrologic model, GR1A (Mouelhi et al., 2006) ~~. This model builds upon the work of~~ to simulate annual runoff in each catchment. GR1A is a conceptual lumped hydrologic model Manabe (1969), considering dynamic storage and antecedent precipitation conditions. The model consists of a simple mathematical equation with a single (optimized) parameter:

$$Q_i = P_i \left\{ 1 - \frac{1}{\left[1 + \left(\frac{0.8P_i + 0.2P_{i-1}}{XE_i} \right)^2 \right]^{0.5}} \right\} \tag{1}$$

where Q, E and P represent annual runoff, basin average potential evapotranspiration and basin average precipitation, respectively ~~;~~ and i denotes the year~~specific index~~. The parameter X is optimized ~~,~~ individually for each catchment by maximizing the Nash-Sutcliffe efficiency (NSE) between ~~the observed and modelled runoffdata~~observed and simulated runoff. The potential evapotranspiration was calculated using the temperature-based formula ~~, provided by Oudin et al. (2005)~~(Oudin et al., 2005). Compared to other conceptual models from the GR family (GR4J, GR5J), GR1A is simple to use and it allows for analyzing many variants, particularly defining best antecedent rainfall and potentially useful to predict wet and dry hydrologic conditions (Mouelhi et al., 2006).

3.3 Data-driven models

Data-driven methods, Artificial Neural Networks ~~(ANNs; Kwak et al., 2020; Hu et al., 2018; Senthil Kumar et al., 2005) in particular, have been~~ (ANNs; Senthil Kumar et al., 2005; Kwak et al., 2020) can describe nonlinear relationships and are widely used for rainfall-runoff prediction. ~~ANN algorithms are very flexible in describing non-linear relations.~~ The ANNs consist of artificial neurons organized in layers and connections that route the signal through the network. Each connection has an associated weight that is optimized within the calibration (in the context of ANNs, known as training). There are many ~~kinds of ANN~~ types of ANNs which differ in terms of structure and type of connections, as well as direction ~~,~~ and functional forms used for neuron activation or training.

In the present study, we considered two approaches: ~~long short term memory~~ Long Short Term Memory (LSTM) neural networks and Bayesian ~~regularized neural networks~~ Regularized Neural Networks (BRNN). These ~~techniques are commonly used to determine the relationship between rainfall and runoff (Hu et al., 2018; Xiang et al., 2020; Kratzert et al., 2018; Ye et al., 2021)~~ approaches have been commonly used in past rainfall-runoff modelling studies(Hu et al., 2018; Kratzert et al., 2018; Xiang et al., 2020; Ye . We considered combinations of ~~gridded~~ reconstructed forcing, OWDA-based scPDSI, ~~proxies and lagged gridded~~ and lagged forcing as an input into the network for both model types. Specifically, the network using only ~~gridded~~ reconstructed forcing is referred to as "~~Gridded", the network with a combination of gridded forcing and natural proxies is known as "Gridded~~P+~~Proxies~~T", the network with ~~gridded~~ reconstructed forcing and OWDA scPDSI is termed as "~~Gridded~~P+~~PDSI"~~ T+PDSI"; and finally the network which ~~include lagged gridded~~ includes 1-year lagged forcing is referred to as "~~Gridded~~P+T+Lag". We also considered and explored lag times longer than 1 year. However the correlation between precipitation and runoff drops significantly at lag times longer than 1 year, here-in we focus on results for the "P+T+Lag(=1)" model.

Figure A1 shows the architecture of LSTM, which is a modified version of the recurrent neural network, based on the back-propagation algorithm (Hochreiter and Schmidhuber, 1997). In this structure, LSTM allows to learn a long-term ~~data set~~ dataset and controls the overfitting problem (Chen et al., 2020). LSTM generally consists of two unit states (hidden and cell states) and three distinct gates (hidden, input and output). In this process, a given cell state saves the long-term memory at the previous unit, while hidden states act as a working memory to process information inside the gates. These gates can determine which information needs to be processed, remembered and transferred in the next state. With LSTM, different activation functions, such as hyperbolic tangent ~~$tanh$ and sigmoidσ~~and sigmoid, can be used to update unit states. The implementation of the LSTM is carried out by means of R packages: “keras” (Arnold, 2017) and “tensorflow” (Abadi et al., 2016).

The training process of the LSTM is time consuming due to its inherent complexity. Therefore, the BRNN ~~method was proposed because of its~~ models providing fast learning and ~~high convergence approximation~~convergence were considered as well. Moreover, the ~~BRNN helps~~ BRNNs help to tackle the complex relationship between rainfall and runoff ~~responses (Ye et al., 2021). This method implements~~ (Ye et al., 2021). BRNNs implement the initial values of the ~~ANN~~ network parameters, using Bayesian regularization (Okut, 2016). Initial weights are set up ~~as~~ based on a prior distribution function during model training~~, typically taken as a normal distribution~~. By applying Bayesian formulation, weight parameters keep updating prior probability distribution to the posterior probability distribution. We trained this model in R using the “brnn” function of the “caret” package (Kuhn, 2015). More details are available in Appendix A3.

~~In both cases, the model optimization runs were conducted several times, and the one with the best performance was considered for further evaluation. To~~ To set the optimal hyperparameters of the models (such as the number of neurons and activation functions) and to reduce the likelihood of overfitting during the calibration/training, ~~a fraction of the calibration data was used to check the performance of~~ the model performance was cross-checked considering an independent (or so-called ~~"testing"~~“testing”) set. ~~In addition, the network parameters (such as the number of neurons, activation functions, etc.) were iteratively tuned to yield fast convergence and good skill.~~ This latter was separated from the calibration data (1900–2000) as a (random) fraction (25%). This process of the model development was repeated several times, minimizing the Root Mean Square Error (for BRNN) and Mean Square Error (for LSTM) for each catchment individually. The model with the best performance was then chosen for further evaluation.

3.4 Goodness-of-fit assessment

We used a set of seven statistical metrics to assess the performance of simulated runoff, namely: Nash–Sutcliffe efficiency (NSE), ~~index of agreement (D), Pearson correlation~~ Pearson Correlation (R), ~~relative error in standard deviation~~ Standard Deviation Ratio (rSD), Kling-Gupta efficiency (KGE), ~~root mean square error~~ Root Mean Square Error (RMSE), ~~mean absolute error~~ Mean Absolute Error (MAE), Bias (BIAS) and Relative Bias (relBIAS). The mathematical formulations of these metrics are provided in Appendix A1.

3.5 Runoff drought identification

To check the utility of our reconstruction, we finally explore how well the annual runoff droughts are represented in the simulations. Our study considers hydrological droughts, defined ~~based on~~ as the streamflow deficit, following the threshold level approach ~~(Yevjevich, 1967; Rivera et al., 2017; Sung and Chung, 2014)~~(Yevjevich, 1967; Sung and Chung, 2014; Rivera et al., 2017) . This approach is typically used for daily or monthly time scales, considering 0.1 or 0.2 quantile threshold levels. To accom-

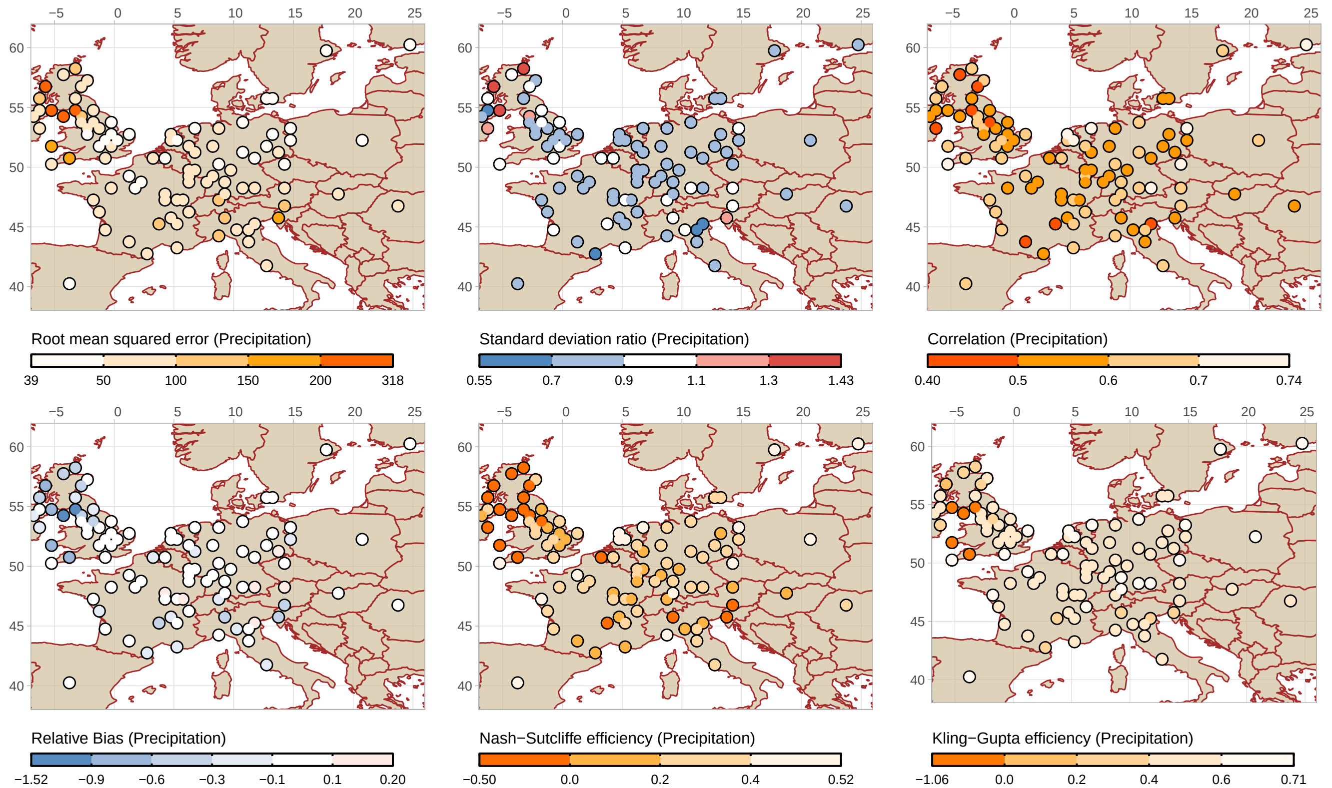


Figure 3. Validation of reconstructed precipitation (Pauling et al., 2006) against GHCN observations. The left- and right- most figures represent the minimum and maximum for the corresponding indicator.

modate the annual scale ~~as~~ used here, we defined the start of the drought, when the annual runoff anomaly falls below the 0.33 quantile (regular drought) and the 0.05 quantile (extreme drought). The drought persists until the runoff rises above the threshold again. ~~Drought length~~ Annual drought duration and severity (the cumulative difference of runoff and the threshold) were then calculated for each identified drought year. ~~Hydrologic~~ Hydrological drought series can be further assessed to understand the critical aspects of runoff (temporal) dynamics and to classify past droughts in Europe ~~(Cook et al., 2015; Wetter and Pfister, 2013)~~ (Wetter and Pfister, 2013; Cook et al., 2015).

4 Results and discussion

In this section, we analyze the ~~500-year-long~~ 500-year annual reconstruction over space and time across Europe. Firstly, we provide a comparison between the GHCN observed precipitation and temperature ~~,~~ and the corresponding grid cells from Pauling et al. (2006) and Luterbacher et al. (2004) reconstructions. Next, the reconstructed annual runoff series for the selected catchments are evaluated against the corresponding observed GRDC runoff data.

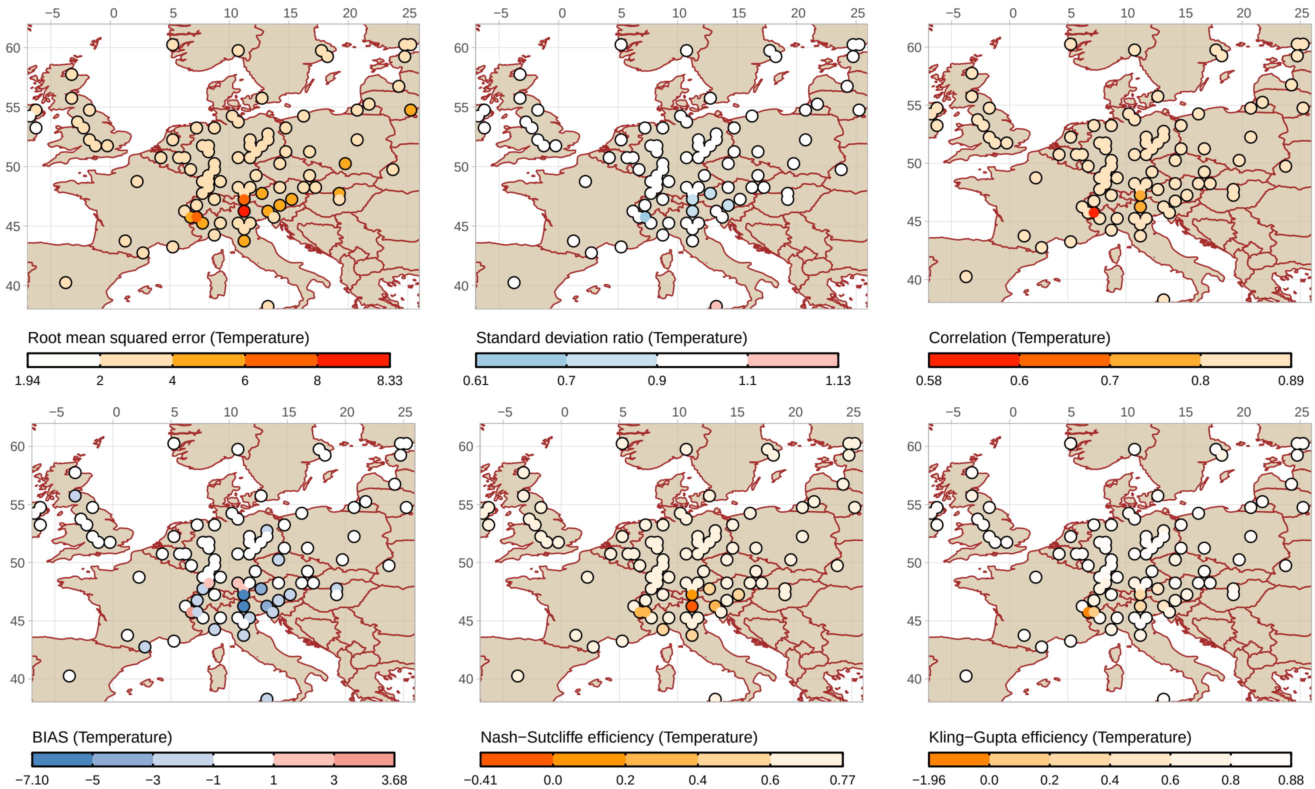


Figure 4. Validation of reconstructed temperature (Luterbacher et al., 2004) against GHCN observations. The left- and right- most figures represent the minimum and maximum for the corresponding indicator.

Two distinct model types were investigated, i.e., a process-based conceptual lumped hydrological model (GR1A) and two data-driven models (BRNN and LSTM). While the former takes ~~gridded~~ reconstructed forcing of precipitation and temperature as an input, in the case of the latter, we also considered PDSI ~~, natural proxies~~ and lagged reconstructed precipitation and temperature fields, as shown in ~~Tab.~~Table 3. Statistical metrics, such as NSE, KGE, RMSE, MAE, R~~and D~~, BIAS and relBIAS (Appendix A1) are used to quantify the predictive skills of the models examined.

4.1 Evaluation of reconstructed precipitation and temperature fields

The 500-year ~~long~~ annual paleoclimate reconstructions of precipitation (~~P~~P) and temperature (~~T~~T) were validated against the GHCN observation data. The ~~spatial map for~~ map showing the comparison is given in Figs. 3 and 4. The reconstructed data are ~~verified against observational P and T~~ evaluated against observational P and T across 99 and 94 European sites, respectively. Figure 3 shows that ~~the~~ for most of the sites the correlation coefficient (R) of ~~P~~ P reconstruction at most of the sites is above 0.5; the ~~index of agreement (D) is larger than 0.6~~Relative bias (relBIAS) is between -0.1 and 0.1; KGE and NSE are showing values below 0.5 ~~(NSE)~~ and 0.6 ~~(KGE)~~respectively; the rSD ~~measurement is greater than~~ is between 0.7 and 0.9 and RMSE varies between ~~50 and 100. We found relatively good performance values for temperature reconstruction~~ 0 and 150.

The performance of the temperature reconstruction was relatively better, as depicted in Fig. 4. In this case, RMSE ~~, estimated~~ between reconstructed and observational ~~T~~T, is around 0.2°C; rSD fluctuates between 0.95 and 1.05, while R is higher than 0.84 and ~~D is above 0.90~~Bias is less than 0.5°C, except for stations located in the Alps. The NSE and KGE values were above 0.5 at ~~many stations. Some stations indicated a worse performance and could not adequately capture the observed temperature variability.~~

~~Furthermore, we tested the skill of gridded reconstructed forcings to capture the multi-temporal characteristics of observed P and T dynamics, i.e., aggregated time-scale features ranging from seasonal to 30-year data. To this end, the seasonal values of the P and T data series were aggregated from 0.25 to a 30-year period (with annual increments) with no overlapping windows. The GOF statistics (Section **??**) between each GHCN station and the corresponding reconstruction grid cell were estimated. In Figure **??**, we present the median gof statistics (black line), the ranges between the 25th and 75th (light envelope) and the 10th and 90th quantiles (dark envelope) of the distribution of the gof statistics over the stationsfor each aggregated time-step. The RMSE for precipitation and temperature drops from initially high values for seasonal scales to relatively stable values for aggregations with a duration greater than 10 years. This is expected since the RMSE depends on the number of observations. With regards to other statistics, except for correlation which shows relatively stable values over aggregations, it is evident that the reconstruction skill decreases the greater the (aggregation) time-scale. In particular, the variance is underestimated and this underestimation is more substantial for long aggregations (see rSD panel in Fig. **??**). This may imply that the utility of multi-year (drought) assessment, utilizing the reconstructed forcing datasets can be limited (and should be interpreted with caution)~~the majority of the stations. Low skill observed at some locations can be explained by the unresolved variability of grid-cell average temperature, especially in regions with complex terrain.

It is worth noting that the large spread of ~~gof~~ GOF statistics is mainly due to the outlying values at the grid ~~cell,~~ cells located along the boundary of the domain (i.e., the interface between land and sea/ocean) and high elevations (cf. also Figs. 3 and 4). In general, reconstructed precipitation exhibits greater differences from observations than temperature. This may be because the proxies considered in the reconstruction rely on different seasons and climate conditions. Additionally, the shortest available instrumental data before the 20th century could encounter certain technical errors, such as problems with instrumental tools, station relocation and dating issues (Dobrovolný et al., 2010). Moreover, other studies (e.g., Ljungqvist et al., 2020) stated that the precipitation series employed for the reconstructions were relatively shorter and more erroneous than the temperature series before the 20th century (Pauling et al., 2006; Harris et al., 2014). Finally, the chosen statistical technique (principal component regression) could also possibly contribute to variance inflation with larger time-scales (Pauling et al., 2006).

~~30 year rolling window of the 500 years gridded data and GRDC observations across 30°N–45°N, 10°W–40°E, the envelope span of two quantiles between the (10th and 90th), (10,75) percentiles of grid cell values for each region. The median value is represented with black thick line, while vertical axis shows the corresponding metric scale separately, for precipitation and temperature~~

4.2 Assessment of the reconstructed runoff simulations

~~For runoff prediction, we have considered several input variables for the models (i.e., the GR1A hydrologic model, the BRNN and LSTM data-driven models), as detailed in Table 3. The available GRDC observed runoff time-series at each gauging location were split into two parts: calibration (1900–2000), used to identify model parameters and validation (prior to 1900), for independent verification using the GOF statistics.~~

The GR1A conceptual hydrological model was driven by the gridded reconstruction of ~~P and T~~ P and T to simulate the annual runoff for each catchment separately. The simulated annual runoff series were then compared to the corresponding GRDC observations and the results were summarized by means of GOF statistics. As can be seen in ~~Table~~ Figure 5, the correlation and NSE statistics for calibration achieve reasonable results at most of the catchments, with a few exceptions (i.e., Kokemenjoki, Goeta, Nemunas and Inn). These (relatively poorer catchment skills in northern Europe) are in line with the previous findings of Seiller et al. (2012) who noted that the lumped hydrological models often exhibit larger uncertainties and fail to capture the extreme catchment values (both high and low). Another study of Fathi et al. (2019) suggested that the performance of the GR1A model is less efficient than the new Budyko framework based SARIMA model in simulating the annual runoff across the Blue Nile and the Danube catchment. This ~~is because~~ may be due to the simplified nature of the model that does not easily capture the complex relationship between rainfall and runoff variability.

In general, statistical values in heat-maps indicate that the neural network algorithms are more skilled for runoff prediction than the GR1A model. The NSE and R statistics (~~Tab~~Fig. 5) for the BRNN and LSTM models indicate a significant improvement in runoff prediction, as compared to the results obtained through the GR1A model. ~~This is especially true with regard to the catchment in Switzerland (Basel Rheinhalle~~ For instance, for Basel Rheinhalle the NSE increases from 0.27 to 0.73 (BRNN) and 0.75 (LSTM) for calibration, and 0.2 to 0.54 ~~for validation). Inclusion of climate-related natural proxies in addition to the reconstructed forcings as an input to the model BRNN(Gridded+Proxies) did not make any significant contribution to the model skill. However, the combination of~~ (BRNN) and 0.52 (LSTM) for validation. Moreover, including scPDSI from OWDA with reconstructed forcing ~~BRNN(Gridded(~~P+~~PDSI), greatly increased the~~ T+PDSI) increases the performance slightly more (NSE 0.76 for both BRNN and LSTM for calibration and 0.57 and 0.59 for validation and BRNN and LSTM, respectively) and considering the lagged forcing results in the best performance (NSE ~~from 0.2 to 0.62).~~ 0.75/0.8 for calibration and 0.6/0.54 for validation, for BRNN/LSTM).

Similarly ~~at most of the~~ for all sites, the ~~simulation based on reconstructed forcings in combination with OWDA scPDSI, yielded a positive~~ data-driven exhibited a strong correlation with the observed runoff, with the GR1A simulations resulting most frequently in lower correlations. Other metrics (RMSE, MAE, KGE~~and D~~, rSD and relBIAS) are shown in ~~Tabs. S3 and S4~~ Figs. S1 – S5 in Supplementary material~~)~~. Across many study locations, the combination of reconstructed forcings ~~with their (one year) lagged version~~ and their 1-year lag performed the best in terms of rapid convergence (the number of iterations needed) and high accuracy from all input combinations ~~(Gridded, Gridded+PDSI, Gridded+Proxies and Gridded+Lag) for both~~ for both data-driven models (BRNN, LSTM). ~~In general, statistical values in heat-maps indicate that the neural network algorithms are more skilled for runoff prediction than the GR1A model.~~ For the validation period, the mean NSE (across all catchments) for the GR1A model is ~~0.1638~~0.16, for the BRNN(~~Gridded~~P+T+Lag) it is ~~0.6836~~ 0.68 and improves to ~~0.7347~~ 0.73 for the LSTM(~~Gridded~~P+T+Lag). In the case of the mean KGE, GR1A ~~is 0.617~~yields 0.62, BRNN(~~gridded~~P+~~lag) is 0.737~~ T+Lag) is 0.73 and LSTM(~~gridded~~P+~~lag) is 0.785~~T+Lag) is 0.78.

To further demonstrate the differences between the individual models, we show the simulated runoff series for all models for those catchments with the highest (~~Blois, Loire~~Blois-Loire) and lowest (~~Smalininkai, Nemunas~~Smalininkai-Nemunas) performance in ~~Figure~~Fig. 6. The performance of the models is comparable during the calibration period for the Loire River. Clearly, all data-driven models are capable of mimicking the observed runoff, while the GR1A model exhibited certain minor deviations, primarily until 1930. In the validation period, the differences between the models are more visible, in particular, for above-average flows. At the beginning of the validation period (1870–1880) all models failed to simulate the high annual flows.

In the case of Nemaunas catchment, the GR1A simulation deviates extremely from the observed data and cannot capture the mean flow level. However, the calibration is poor even for the data-driven models and, does not simulate the year-to-year variability appropriately. Interestingly, ~~the~~ for the validation period the error in the GR1A model ~~is less in relation to validation than calibration. The~~ decreases. The performance of the data-driven models ~~perform in a similar way to that of calibration , with only minor differences between the two~~ is similar in validation and calibration periods. Looking at the ~~gof~~ GOF statistics, the models considering OWDA-based scPDSI or lagged forcings (e.g., P_{t-1}) perform slightly better in terms of KGE than the other model configurations. ~~This improvement may be due to a better representation of temporal dependency structure, introduced either by scPDSI or a consideration of the forcing values from the previous year in the case of LSTM(Gridded+Lag) and BRNN(Gridded+Lag).~~

4.3 The annual runoff reconstruction datasets

As a first step, we excluded the catchments that exhibited poor performance in validation (see ~~Table~~ Fig. 5). As a threshold, we considered validation NSE of 0.5 for at least one model, following the approach used by Ayzel et al. (2020). In this step, we excluded seven catchments (Vlotho-Wesser, Decin-Elbe, Burghausen-Salzach, Smalininkai-Nemaunas, Vargoens ~~KRV- Goeta~~KRV-Goeta, Elverum-Glama, Muroleekoski-Kokemenjoki) out of 21, ending up with a set of simulations for 14 catchments (highlighted by the thick box in ~~Tab~~Fig. 5).

Secondly, we identified the candidate best models for each of the 14 selected catchment, considering the ~~gofs based on~~ GOFs based on the validation NSE and R greater than 0.5 and ~~0.70, respectively, for the validation period. In addition, the model performance with respect to the remaining measures (D,~~ 0.7, respectively. The best model for each catchment was finally subjectively selected from those models considering the remaining validation measures (BIAS, rSD, KGE, RMSE and MAE) ~~was also considered. Eventually, we decided to utilize that model since the metrics used (NSE, KGE, R, D, RMSE, MAE) to produce better results in one particular model.~~ as well. The resulting selected models are shown in Table 3. The combination of ~~gridded~~ reconstructed forcing with lagged values results in the best performance over nine catchments, of which seven are driving the BRNN and the remainder the LSTM. The LSTM with ~~gridded~~ reconstructed forcing and OWDA-scPDSI was best in one case, and the remaining four were most appropriately simulated with the BRNN~~and~~

~~BRNNGridded~~(P+~~PDSI~~T) and BRNN(P+T+PDSI). It should be noted that the differences between the models performing well are small, as noted in ~~Figure~~Fig. 6 and further demonstrated in ~~Figure~~Fig. 7. The latter figure compares the cumulative distribution functions of annual runoff for the periods 1500–1800, 1800–1900 and 1900–2000, as simulated by the BRNN(P+T+Lag) and LSTM(P+T+PDSI) – the two best performing models – and the GR1A (the most ~~distinctive simulations~~ deviating simulation from the best model) with the distribution of the observed annual runoff for the Basel-Rheinhalle Rhine catchment. For the calibration period (post-1900), the models perform well except the GR1A, which generally overestimated the observed maxima. The cumulative distribution of BRNN and LSTM ~~simulations~~ simulated runoff values are very similar for the validation period except for the top and bottom 5% in 1500–1800. The GR1A simulation showed significant differences for the entire ~~observed~~ distribution, thus overestimating/underestimating the maxima/minima. The difference from the best model can be expressed in terms of KGE – even here, it was evident that the GR1A model deviated considerably (KGE 0.6–0.7) while the LSTM is very similar to the BRNN (KGE 0.92–0.96). The most severe drought year identified by the models was the same in the periods 1500–1800 and 1900–2000 (Fig. 7 left and right panels), while for 1800–1900 the models identified either 1865 (GR1A, LSTM) or 1858 (BRNN, 2nd worse for LSTM). Please note that the 1858 low water mark is available at Laufenburg Pfister et al. (2006) near Basel and was regarded as one of the worst winter droughts in the last 200 years.

The resulting 14 annual runoff reconstructions are available at https://doi.org/10.6084/m9.figshare.15178107 and are shown in supplementary ~~figures~~ material (Figs. ~~S1, S2, and S3~~S6, S7, and S8). As an example, we present only two runoff reconstructions here (Fig. 8). As an additional validation for ~~these~~ the reconstructed series, we ~~present the~~ inspected the quantile-quantile plots of the observed ~~runoff versus the reconstructed runoff in~~ and reconstructed runoff (Fig. 9). The simulated series are generally consistent with the observed runoff, especially for the Montjean-Loire, Köln-Rhine, and Basel Schifflaende-Rhine catchments, which exhibit the best relationship between the observed and the simulated runoff.

Finally, to check the consistency of our reconstructed dataset, we compared the skill of our simulation with respect to the GRDC runoff observation and the GSWP3-forced GRUN monthly runoff (Ghiggi et al., 2019) datasets. The gridded GRUN datasets were averaged over the respective catchments for the comparability (Supplementary Fig. S9, S10). Our reconstruction outperforms GRUN data in terms of RMSE, MAE, relBIAS and NSE across the majority of the catchments, while the correlation to GRDC runoff is slightly higher for GRUN compared to our reconstruction. The variability, which our data-driven models underestimate (on average by 16.5%), is overestimated by GRUN (on average by 17.2%). Since the correlation compensates for the Bias, the KGE for our reconstruction and GRUN is comparable. This suggests that GRUN could be used for data-driven model training, provided at least some information on flow characteristics is available in the catchment.

4.4 Identification of low flows and significant hydrological drought events

In the final step of the analysis, we compared the droughts identified in the reconstructions with the GRDC observed series (Fig. 9). The ~~match~~ agreement between the simulated and observed runoff deficit is ~~less~~ lower compared to the annual runoff time series. For most of the stations, the simulated deficit is lower than the corresponding observed estimates. This suggests that the reconstructed precipitation and temperature fields do not represent the inter-annual variability correctly~~, which is in~~

Table 3. Selection of best model for runoff in individual catchments

Models	Catchments
BRNN(~~Gridded~~P+T)	Blois-Loire, Rees-Rhine
BRNN(~~Gridded~~P+T+PDSI)	Wuerzburg-Main and Orsova-Danube
BRNN(~~Gridded~~P+T+Lag)	Montjean-Loire, Köln-Rhine, Hann-Munden-Wesser, Dresden-Elbe, BaselRheinhalle-Rhine, Bodenwerder-Wesser, Wasserburg-Inn
LSTM(~~Gridded~~P+T+Lag)	NeuerHafen-Main, Intschede-Wesser
LSTM(~~Gridded~~P+T+PDSI)	Baselschifflaende-Rhine

~~line with findings from Fig. ??.~~ . Despite a widespread issue with the representation of inter-annual persistence, Fig. 10 shows that the runoff deficits are simulated reasonably well for the Rees-Rhine and Köln-Rhine catchments.

In the next step, we contrasted reconstructed drought patterns over the last 500 years ~~,~~ with data available from documentary evidence and other sources. In the case of extreme droughts, we considered the $q_{0.05}$ threshold before 2000 CE. Low flow analysis since 1500 and the maximum/minimum deficit values of catchments are shown in ~~Tab.~~ Table 4. In the 16^{th} century, the years 1536, 1540 and 1590 are associated with significant runoff deficits. The event of 1540, had already been reported ~~(Brázdil et al., 2019; Cook et al., 2015; Brázdil et al., 2013)~~ (Brázdil et al., 2013; Cook et al., 2015; Brázdil et al., 2019) as the worst event of the 16^{th} century and was also more severe in terms of hydrologic shifting. In 1540, almost 90% of the Rhine and Elbe River catchments (Basel and Cologne) experienced low yearly discharge, which ranked as the greatest low flows in the last five centuries (Leggewie and Mauelshagen, 2018). The seasonal precipitation was also deficient and was evident primarily in Central Europe and England (Dobrovolný et al., 2010). Wetter and Pfister 2013 stated that the spring and summer of 1540 was likely to have been warmer than the comparable period during the 2003 drought. The simulation shows that the drought during 1540 was evident in most study catchments, such as the Rhine, Main, Wesser, Loire and Danube, except ~~Wasserburg-INN.~~ Wasserburg-Inn.

In the 17^{th} century, the years 1603, 1616, 1631, 1666, 1669, 1676, 1681, 1684 and 1686 were simulated as exceptionally low-flow years. Furthermore, two events (~~1686 and~~ 1669 and 1686) were associated with the maximum water deficit across several study catchments. Baselschifflaende-Rhine catchment is a good example of this, which experienced a severe runoff deficit during 1669. Alternatively, 26 remarkable droughts have been captured in the Köln-Rhine catchment over the past 500 years, and the year 1686 reached the highest runoff deficit (156 mm/year). In addition, 1616 was the driest year of the 17^{th} century, the so-called “drought of the century~~”~~“ (Brázdil et al., 2013), which significantly impacted the major rivers in Europe (e.g., Rhine, Main and Wesser). Brázdil et al. (2018) identified three unusual drought periods (1540, 1616 and 1718–19) over the Czech lands, highlighting the 1616 drought, which caused widespread famine, dried up the Elbe river watershed and altered the climate of neighboring nations (Switzerland and Germany). The hunger stone of the Elbe River also revealed the exceptionally dry year of 1616 (Brázdil et al., 2013). During the 18^{th} century, a similar level of runoff deficit was simulated in the years 1706 and 1719.

During the 19th century, the years 1863, 1864, 1874, 1893 and 1899, were recognized as drought years in all catchments, while in the 20th century, the driest periods occurred in 1921, 1934, 1949 and 1976. The 1921 drought in the Blois-Loire, Rees-Rhine, Köln-Rhine, Orsova-Danube, BaselRheinhalle-Rhine and Baselschifflanede-Rhine catchments was ranked as the most exceptional drought in the 20th century. Three catchments (BaselRheinhalle-Rhine, Baselschifflanede-Rhine and Blois-Loire) ~~were simulated with a high deficit for~~ exhibited a high runoff deficit during the year 1921. A noticeable increase in temperature was experienced across Europe, and certain areas were notably affected by a heat wave in July of that year. The majority of Central Europe, southern England and Italy were affected by this drought, including London, where the rainfall was found to have decreased by 50 to 60% relative to the average (Cook et al., 2015). Similar to our results, certain photographs from the Dutch newspaper (De Telegraaf) show the lowest river flows in the Rhine (Switzerland), Molesey Weir (the Thames River, UK) and Loire River (France, van der Schrier et al., 2021). The precipitation totals were recorded as the lowest since 1774, and the year was also ranked top (in terms of deficit rainfall) in the Great Alpine region ~~(Haslinger and Blöschl (2017))~~(Haslinger and Blöschl, 2017), where the rainfall deficit began in winter 1920/21 and lasted until autumn 1921. Monthly runoff anomalies analyzed from the GRUN dataset (Ghiggi et al., 2019) show that August 1976 was the fifth driest month between 1900 and 2014, with some of our study catchment also signaling the 1976 yearly drought (e.g Köln-Rhine, Hann-Munden-Wesser, Bodenwerder-Wesser).

In summary, the reconstructed annual runoff corresponded well to the majority of extreme drought years (e.g., 1540, 1616, 1669, 1710, 1724, 1921, as highlighted in ~~Tab.~~ Table 4) and previously demonstrated in the OWDA-based PDSI tree-ring reconstructions or other references ~~(Wetter and Pfister, 2013; Cook et al., 2015; Dobrovolný et al., 2010; Brázdil et al., 2013; Markonis et al., 20~~ ~~.~~ (Dobrovolný et al., 2010; Brázdil et al., 2013; Wetter and Pfister, 2013; Cook et al., 2015; Markonis et al., 2018).

This might be the case as the tree-ring proxies involved in the developed reconstruction were the same, which could reveal the true nature of hydroclimatic shifts. Still, our reconstruction missed certain notable dry events, e.g., 1894 (Brodie, 1894) which was associated with unprecedented low levels of rainfall and excessive temperature rises in the south of England, the British Isles, and other European regions ~~(Cook et al., 2015; Hanel et al., 2018; Brodie, 1894).~~ (Brodie, 1894; Cook et al., 2015; Hanel et al., 2018 .

5 Conclusions

In this study, hydrological (GR1A) and two data-driven (BRNN, LSTM) models were used to ~~simulate~~ reconstruct the annual runoff during the period 1500–2000, considering various input fields. ~~Different input configurations were evaluated for runoff predictions.~~ Following validation of the simulated series, we provided ~~runoff reconstructions~~ annual runoff time-series for 14 catchments across Europe~~(Germany: Main, Rhine, Wesser, Inn, the Netherlands: Rhine and Romania: Danube)~~. The main findings can be summarized as follows:

1. Data-driven methods have proven to be helpful for annual runoff simulations even when there are deficiencies in the driving input fields. This contrasts with a conceptual lumped hydrological model, which would require bias correction before the simulation.

Table 4. Simulated runoff ~~deficiency of extreme cases over past 500 years~~ droughts since ~~2000 CE~~1500. Years in bold indicate extreme droughts.

~~**station Name**~~**Station name**	**No of events**	~~**simulated**~~ **Simulated low flow years**	~~**minimum**~~ **Minimum deficit (year)**	~~**maximum**~~ **Maximum deficit(year)**
Orsova-Danube	12	1536, **1540**, **1669**, **1686**, 1704, 1706, **1710**, 1746, 1834, 1943, 1947, 1990	2.19 (1704)	30.33 (1686)
~~Dresden-Elbe~~ Dresden-Elbe	1	**1669**		2.76 (1669)
Wasserburg-Inn	3	**1669**, **1686**, 1754	1.79 (1754)	27.8 (1669)
Blois-Loire	17	**1540**, 1603, 1631, 1634, **1669**, 1676, **1686**, 1706, **1710**, **1724**,1736, 1754, 1766, 1884, **1921**, 1945, 1949	0.07 (1766)	85.7 (1669)
Montjean-Loire	48	**1540**, 1603, 1607, **1616**, 1630, 1631, 1632, 1633, 1634, 1635, 1661, **1669**, 1670, 1676, 1680, 1681, 1684, 1685, **1686**, 1702, 1704, 1705, 1706, **1710**, 1715, 1717, 1718, 1723, **1724**, 1731, 1736, 1742, 1743, 1744, 1745, 1746, 1753, 1754, 1757, 1785, 1815, 1826, 1834, 1874, 1884, **1921**, 1945, 1949	1.95 (1874)	105.2 (1686)
NeurHafen-Main	~~6~~ 18	1590, ~~**1540**~~**1616**, **1669**, 1681, 1682, **1686**, 1704, 1706, **1710**, **1724**, 1746, 1754, 1755, 1814, 1865, 1934, 1943, 1964	~~0.26 (1681~~3.58 (1964)	~~83.84~~ 100.89 (1669)
Wuerzburg-Main	2	**1540**, **1669**	0.3 (1540)	17.0 (1669)
BaselRheinhalle-Rhine	21	1536, **1540**, 1590, 1603, **1616**, 1631, 1666, **1669**, 1676, 1681, **1686**, 1704, 1706, **1710**, **1724**, 1736, 1746, 1753, 1754, **1921**, 1949	1.78 (1704)	133.9 (1669)
Baselschifflaende-Rhine	~~22~~ 19	1536, **1540**, 1590, 1603, **1616**, 1666, **1669**, 1676, 1681, 1684, **1686**, 1706, **1710**, **1724**, ~~1728,~~ 1736, 1746, 1754, ~~1766, 1822, 1834, 1865,~~ **1921**, ~~1947,~~ 1949 ~~, 1976~~	~~3.60 (1766~~10.4 (1590)	~~370.8~~ 563 (1669)
Köln-Rhine	28	1536, **1540**, 1590, 1603, **1616**, 1631, 1634, **1669**, 1676, 1681, 1684, **1686**, 1704, 1706, **1710**, **1724**, 1736, 1744, 1745, 1746, 1753, 1754, 1858, 1865, 1874, **1921**, 1949, 1976	1.34 (1745)	157.6 (1686)
Rees-Rhine	18	1536, **1540**, 1603, 1631, 1666, **1669**, 1676, 1681, **1686**, 1704, 1706, **1710**, **1724**, 1736, 1746, 1754, **1921**, 1949	11.7 (1704)	96.0 (1669)
Hann-Munden-Wesser	11	**1540**, **1669**, 1681, **1686**, 1706, **1710**, **1724**, 1911, 1934, 1976, 1991	1.95 (1991)	46.6 (1669)
Bodenwerder-Wesser	15	**1540**, **1616**, 1631, **1669**, 1681, **1686**, 1706, **1710**, **1724**, 1754, 1858, 1874, 1911, 1934, 1976	0.029 (1858)	56.3 (1669)
Instchede- Wesser	18	**1540**, **1616**, 1631, **1669**, 1670, 1676, 1681, 1685, **1686**, 1706, **1710**, 1754, 1814, 1857, 1858, 1865, 1934, 1959	0.30 (1670)	134.4 (1669)

2. There is no significant difference between the BRNN and LSTM-simulated annual runoff neither in terms of the individual values nor in relation to the validation metrics.

3. Validation skill metrics suggest that for annual runoff prediction, it is beneficial to consider data-driven models that explicitly account for serial dependence either through input data (e.g., time-lagged input fields) or directly in the model structure (e.g., LSTM - networks).

4. The droughts identified in the reconstructed series correlated well with significant documented events (such as 1540, 1616, 1669, 1710, 1724 and 1921).

The reconstructed series relies heavily on the consistency of underlying reconstructed precipitation (Pauling et al., 2006) and temperature (Luterbacher et al., 2004) forcing fields. Unfortunately, ~~this~~ those cannot be fully verified directly, due to the lack of sufficient long-term observational ~~data sets~~datasets. With the limited information (GHCN), we identified several notable deficiencies in the reconstructed forcings, in particular, underestimated variance in precipitation reconstruction, leading to inconsistencies in observed runoff (e.g., demonstrated by the poor results of GR1A for some catchments). Moreover, proxy records ~~are spatially heterogeneous (also used in the development of gridded reconstructions). Due to the fact that some regions are~~ that were used for the derivation of precipitation and temperature input fields are spatially heterogeneous with some regions being better represented than others~~and inevitably this results~~. This inevitably leads to poor performance over the latter.

~~However, the~~ The skill of precipitation and temperature reconstructions across the selected catchments to ~~develop runoff is~~ derive annual runoff is still fairly good. In addition, the data-driven methods that were used in the paper are capable of removing systematic bias (as was proven in validation). We cannot be sure, though, that the link between reconstructed forcing and annual runoff is stationary when going back in time. Moreover, when the number of natural proxies included in the derivation of the forcing dataset decreases, the uncertainty increases. The reconstructed data should, therefore, always be considered with caution. ~~In addition, we showed that the skill of the reconstructed forcings decreases with time-scale. This may imply problems with the representation of multi-year droughts.~~

Future research could consider further improvements of the simulations, e.g., by training a meta-model combining the runoff simulations from several fitted models. ~~Since~~ In addition, since interest is not often focused on the runoff series, but on some other indicator (such as PDSI or deficit volume in the case of drought), it is also possible to simulate the drought indices directly, considering either the precipitation and temperature input fields or the simulated runoff. Finally, discrete classifiers (Kolachian and Saghafian, 2021) could also be used to simulate the drought (or water level) classes directly.

6 Data Availability

The annual runoff reconstruction were prepared using the ~~below data set~~ defined dataset and can be accessed at free, public repository Figshare (https://doi.org/10.6084/m9.figshare.15178107, Sadaf et al. 2021). The ~~gridded~~ reconstructed data of precipitation and temperature can be downloaded at website via link https://www.ncdc.noaa.gov/data-access/paleoclimatology-data. The monthly global historical climatological network (GHCN) provides revision and updated version (V4) for temperature and

(V2) precipitation which can be accessed via the link https://www1.ncdc.noaa.gov/pub/data/ghcn/. The data repositories of GRDC runoff is accessible for public at https://www.bafg.de/GRDC/EN/Home/homepage_node.html.

Appendix A

A1 Goodness-of-fit assessment

We used ~~a few statistical measures~~ several statistical indicators to assess the ~~skillfulness of runoff reconstruction using a gridded-based simulation and an observed data-set. These measurements are mathematically defined as follows:~~

$$rSD = \frac{SD_{g_i}}{SD_{o_i}}$$

~~The terms g_i and o_i~~ skill of annual runoff reconstruction. In following definitions, p and o refer to the ~~gridded and observed time seriesat point i, respectively . The standard deviations~~ predicted and observed series, respectively and i to year.

The Standard deviation ratio (rSD~~) returns the maximum value of 1. The observed~~ ; Ghiggi et al., 2021) is defined as

$$rSD = \frac{SD_p}{SD_o} \tag{A1}$$

with SD the standard deviation. The variability is underestimated when the value is less than one, ~~while the observed variability is~~ and overestimated when the value is greater than one.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(g_i - o_i)^2}{n}}$$

The Root Mean Square Error (RMSE; see e.g. Legates and McCabe Jr, 1999)

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(p_i - o_i)^2}{n}} \tag{A2}$$

and Mean Absolute Error (MAE; see e.g. Legates and McCabe Jr, 1999)

$$MAE = \sum_{i=1}^{n} \frac{|(g_i - o_i)|}{n}.$$

$$MAE = \sum_{i=1}^{n} \frac{|(p_i - o_i)|}{n} \tag{A3}$$

~~The RMSE and MAE~~

measure how well predictions fit the ~~measurements~~observations. MAE and RMSE values can range from 0 to infinity, with the former value indicating a perfect fit~~to a zero fit~~.

$$R = \frac{cor_{g_i}}{cor_{o_i}}$$

The Pearson's correlation coefficient (R) is defined as

$$R = \frac{\sum_{i=1}^{n}(p_i - \overline{p})(o_i - \overline{o})}{\sqrt{\sum_{i=1}^{n}(p_i - \overline{p})^2}\sqrt{\sum_{i=1}^{n}(o_i - \overline{o})^2}} \tag{A4}$$

~~Cor computed the correlation of observed and predicted data. The method can be specified as "kendall" or "spearman". Kendall's tau or Spearman's rho are used to estimate rank-based competence.~~ The Nash–Sutcliffe efficiency (NSE~~),~~ ; Nash and Sutcliffe, 197),

$$NSE = 1 - \frac{\sum_{i=1}^{n}(p_i - o_i)^2}{\sum_{i=1}^{n}(o_i - \overline{o})^2} \tag{A5}$$

is alternatively referred to as model efficiency~~(Nash and Sutcliffe, 1970), is a metric for the model's overall competence. It is defined as follows:~~

$$NSE = 1 - \frac{\sum_i (g_i - o_i)^2}{\sum (o_i - mean(o_i))^2}$$

. NSE = 1 corresponds to a perfect match between predicted and observed data, while a value less than 0 indicates that model predictions are on average less accurate than using the long-term mean of the observed time series ~~$mean(o_i)$~~$\overline{o}$.

~~Another coefficient of efficiency D, the index of agreement represents a decided improvement over the coefficient of determination but also is sensitive to extreme values, owing to the squared differences.~~ Systematic errors can be detected by using the absolute bias (BIAS)

$$d = 1 - \frac{\sum_i (g_i - o_i)^2}{\sum (|g_i - mean(o)||o_i - mean(o_i)|)^2}$$

$$BIAS = \overline{p} - \overline{o} \tag{A6}$$

~~The index of agreement ranges from 0.0 to 1.0, with higher values signifying a better agreement between the model and observations, similar to the interpretation of the coefficient of determination.~~ or relative bias (relBIAS)

$$KGE = 1 - ED$$

$$ED = \sqrt{s[1]*(r-1)^2 + s[2]*(\alpha-2)^2 + s[3]*(\beta-2)^2}$$

$$\alpha = \frac{\sigma_g}{\sigma_o}$$

$$\beta = \frac{\mu_g}{\mu_o}$$

$$relBIAS = \frac{\overline{p} - \overline{o}}{\overline{o}} \tag{A7}$$

which has an ideal value of 0. Positive bias values indicate that the model prediction overestimates observations, whereas negative values indicate underestimated model prediction.

The Kling-Gupta efficiency ~~(KGE) index~~ index (KGE; Gupta et al., 2009)

$$KGE = 1 - \sqrt{(R-1)^2 + (rSD-2)^2 + (\beta-2)^2} \tag{A8}$$

$$\beta = \frac{\overline{p}}{\overline{o}} \tag{A9}$$

is calculated using three primary components: ~~r, α~~R, rSD, and β~~. The symbol r denotes the Pearson product-moment correlation coefficient; α denotes the ratio of the standard deviations of the simulated and observed values;~~ , with R and rSD defined above and β ~~denotes~~ the ratio of the mean of the ~~simulated~~ predicted and observed values. ~~α, β, and r~~ rSD, β, and R have an ideal value of one. ~~s is a three-dimensional numeric representation of the scaling factors of length three that is used to adjust the relative importance of various components.~~

A2 ~~Data prepossessing of~~ Long short term memory (LSTM)

To build the LSTM model, we use the Keras environment (Arnold, 2017) with its high-level application programming interface (API) for neural networks and tensor ~~flows~~flow. Figure A1 represents the structure of the LSTM neural model for the rainfall runoff relationship in several catchments. We design our network by stacking one LSTM and two dense layers on top of one other. As shown in Fig. A1, the model configured four distinct input combinations, each of which was normalized to [0, 1] in the training and testing phases. The model parameters choose different batch shapes, units (similar as neurons) and epochs as described in Table A1. The model considers the Rectified Linear Unit (ReLU), using component wise multiplication and defining the dropout parameter as 0.1. According to Kingma and Ba (2014), the optimization algorithm plays a significant role in the algorithm's convergence and optimization. For this reason, Adam's optimizer is considered, as it performs stochastic gradient descent (SGD) more efficiently using the backpropagation algorithm. During compilation, the learning rate is set to ~~'~~0.001 ~~', or '~~or 0.002 ~~', and the model selects random batch sizes and epochs~~and the mean square error (MSE) is used to measure model accuracy. In addition, the mean absolute error (MAE) is a function used as an objective to minimize residues and achieve optimum value. The checkpoint algorithm is also applied to test the model's accuracy level. Finally, the best output of the model is saved, with minimum loss and better accuracy.

A3 Bayesian Regularized neural network (BRNN)

BRNN is a probabilistic technique for handling nonlinear problems. By using the caret package, the model 'brnn' was designed to work with a two-layer network as described by (MacKay, 1992; Foresee and Hagan, 1997). BRNN uses the Nguyen and Widrow algorithm to assign initial weights and the Gauss-Newton algorithm to optimise. Model is first trained on the training dataset, and its performance is checked by making a prediction on the testing dataset.

While selecting a model for train control, a simple boot resampling strategy was applied to evaluate performance. We tested the proposed model's predictive ability using a random bootstrap generator, with 75% of the observations in the training set and 25% in the testing set. RMSE was utilized as a loss function to compile and verify the model's accuracy. The model was fitted with 20 neurons, one hidden layer and implemented activation function $g_k(x) = \frac{exp(2x)-1}{exp(2x)+1}$. After compilation, the train function automatically selected the best model with the smallest RMSE as the final model. After getting the optimal model, the data is further evaluated the performance on testing data and predicted runoff values for the previous 500 years.

Table A1. Structure and hyperparameters of two data driven models (BRNN and LSTM) for Runoff predictions

Training algorithms	**Layer types**	**Activation functions**	**Hyperparameters**
BRNN	input, hidden, output	$g_{k(x)} = \frac{exp(2x)-1}{exp(2x)+1}$	Tunelength 20, neurons (1-20)
LSTM	input, hidden, output	Rectified Linear Activation (ReLU) $f(x) = \begin{cases} 0 & \text{when x<0} \\ x & \text{when x} \geq 0 \end{cases}$	Learning rate: 0.0001, epochs (30-200), units (5-150), batch input shapes: (1,1,2) for LSTM, (1,1,3) for LSTM(P+T+PDSI), (1,2,2) for LSTM(P+T+lag).

Author contributions. The study was initially designed by RK, MH and YM. Algorithms are coded with the assistance of YM, US and MH. Datasets were collected by VG and SN. The research was carried out by SN, MS, and MH, who also wrote the paper. OR and RK both helped to revise the manuscript.

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

Acknowledgements. This work was carried out within the bilateral project XEROS (eXtreme EuRopean drOughtS: multimodel synthesis of past, present and future events), funded by Czech Science Foundation (Grant No. 1924089J) ~~and~~ together with the Deutsche Forschungsgemeinschaft (Grant No. RA 3235/11) ~~+ IGA~~ and internal grant of the Czech University of Life Sciences (Project No.2020B0018). We thank the Global Runoff Data Centre (GRDC) for providing the observed runoff data. All analyses and visualisations were done using R.

-2.2cm

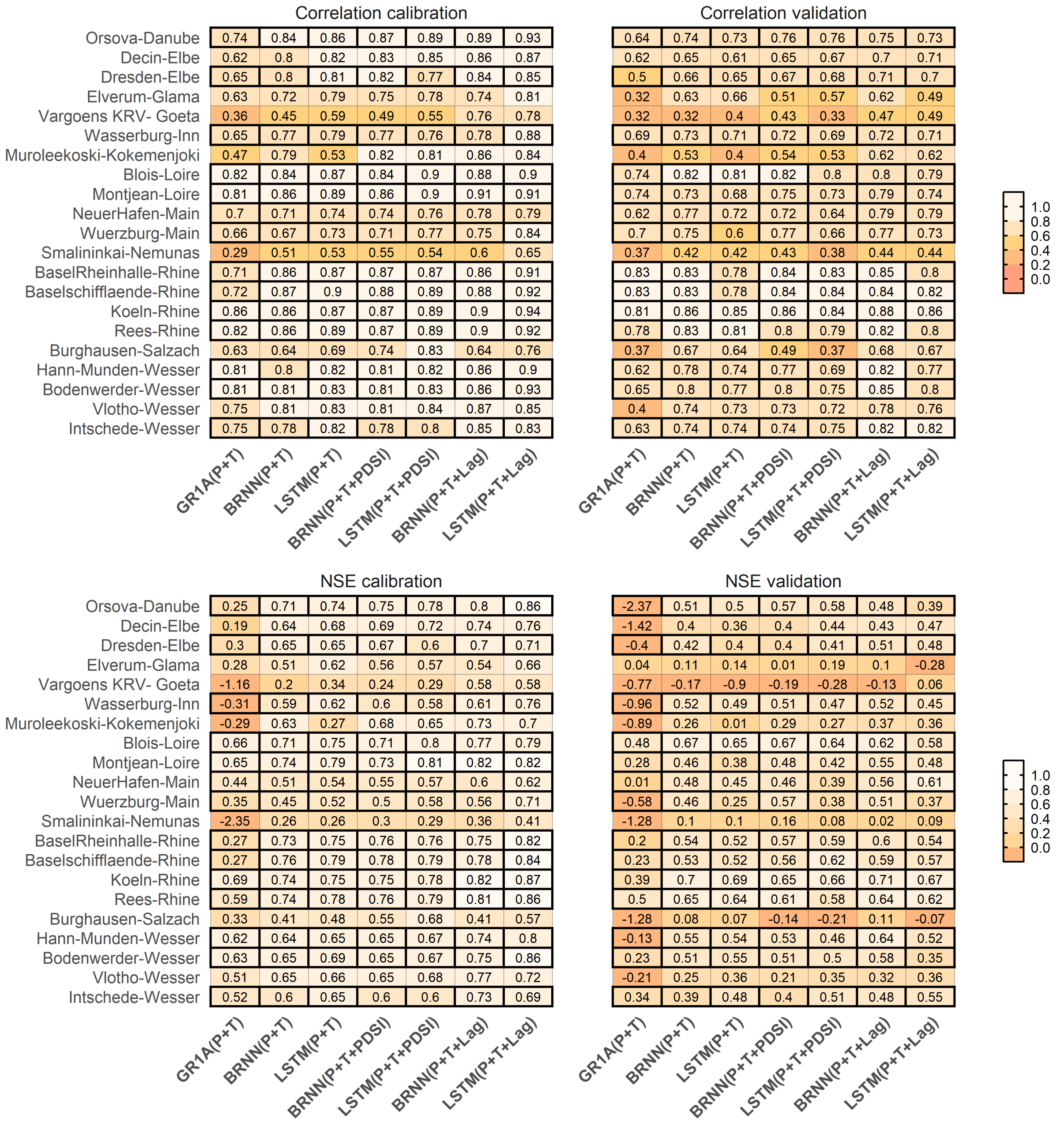


Figure 5. The correlation coefficient (top) and NSE (bottom) for calibration (left) and validation (right) of the considered models for 21 study catchments. The ~~y~~ vertical axis ~~consists of water gauge stations~~ represents the catchments (station name and ~~relevant~~ river) ~~in Central Europe, Alps~~ and ~~Lithuania~~the horizontal axis the considered models. The rectangular black frames represent the catchments with satisfactory validation.

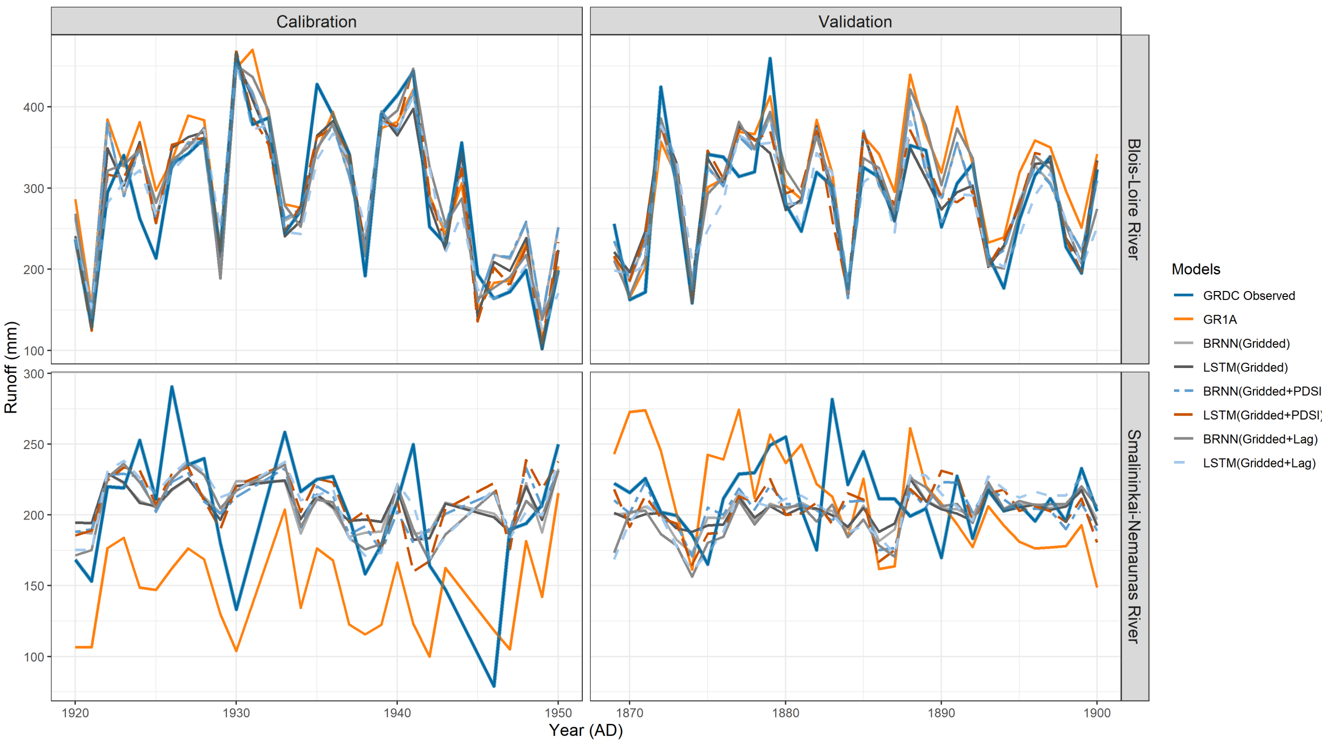


Figure 6. Comparison between the models for the station with the best (Bloise-Loire River, top) and the worst (Smalininkai-Nemaunas River, bottom) model fit.

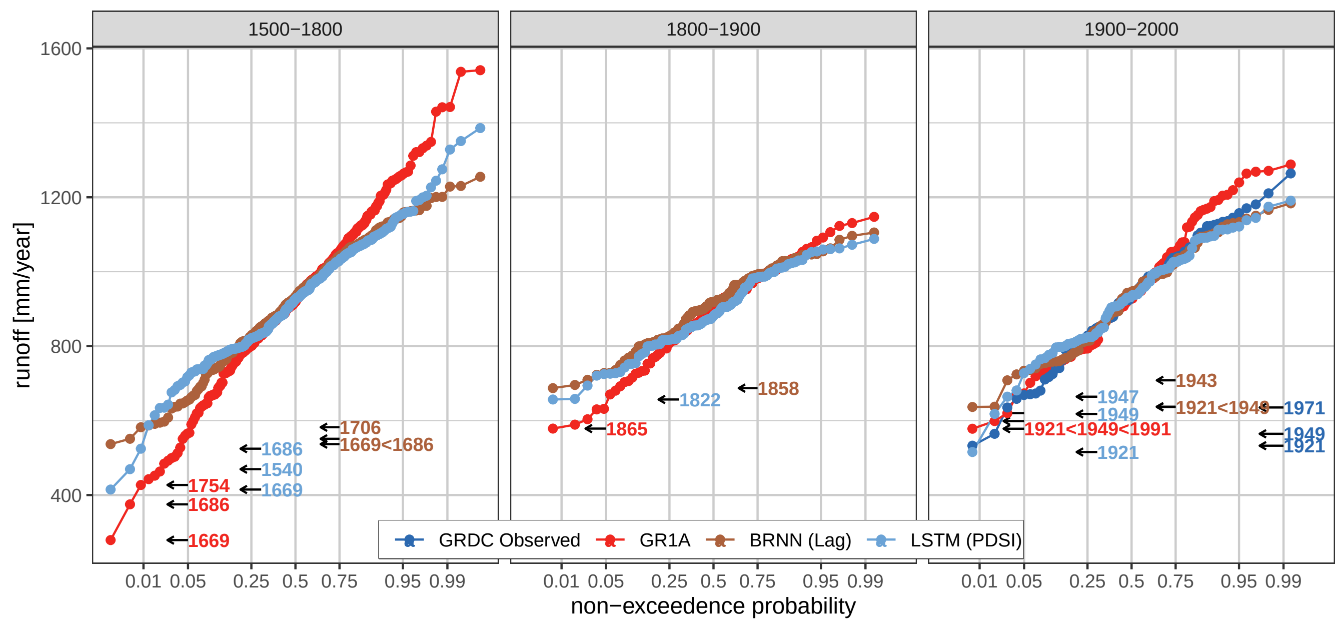
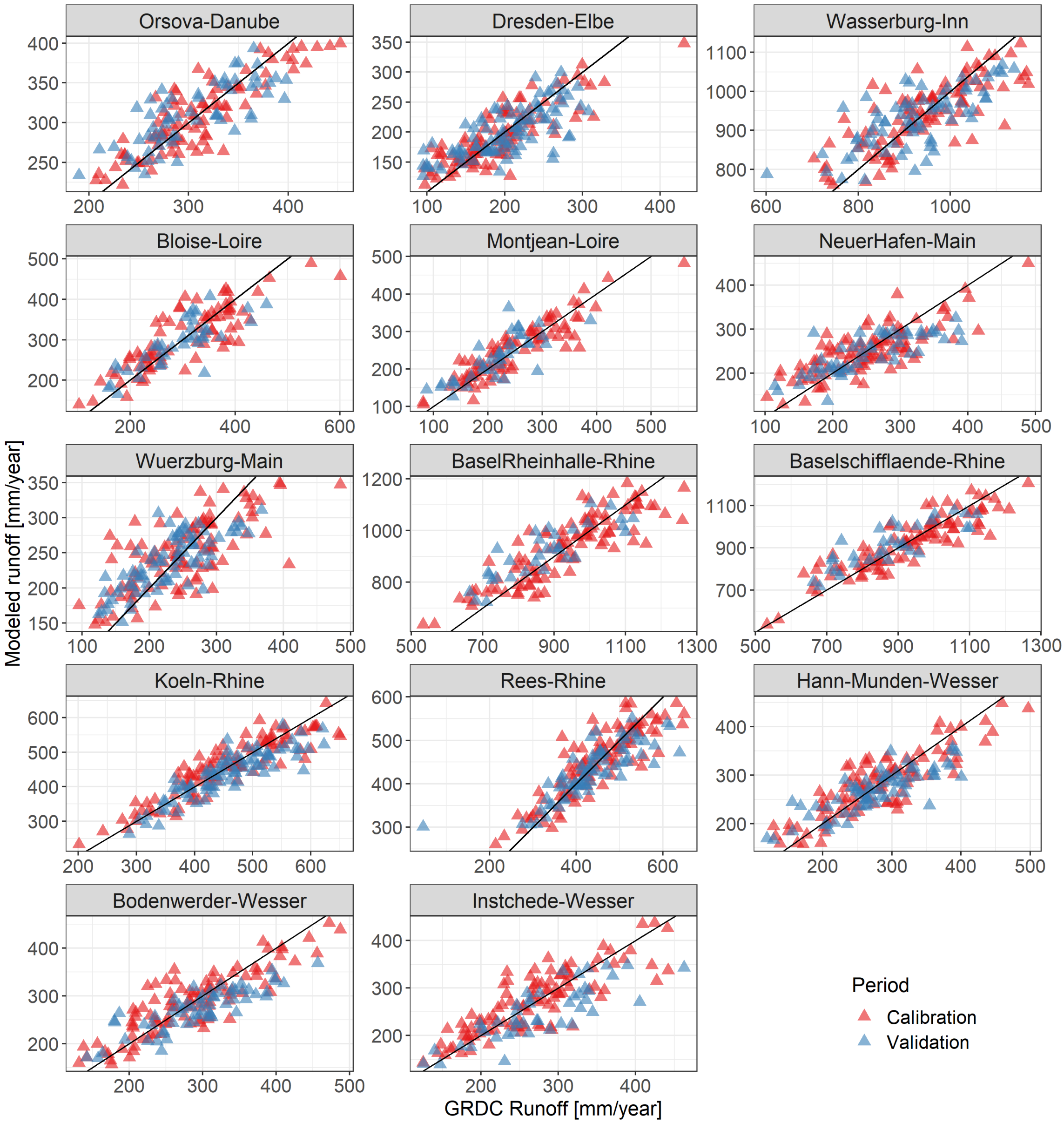


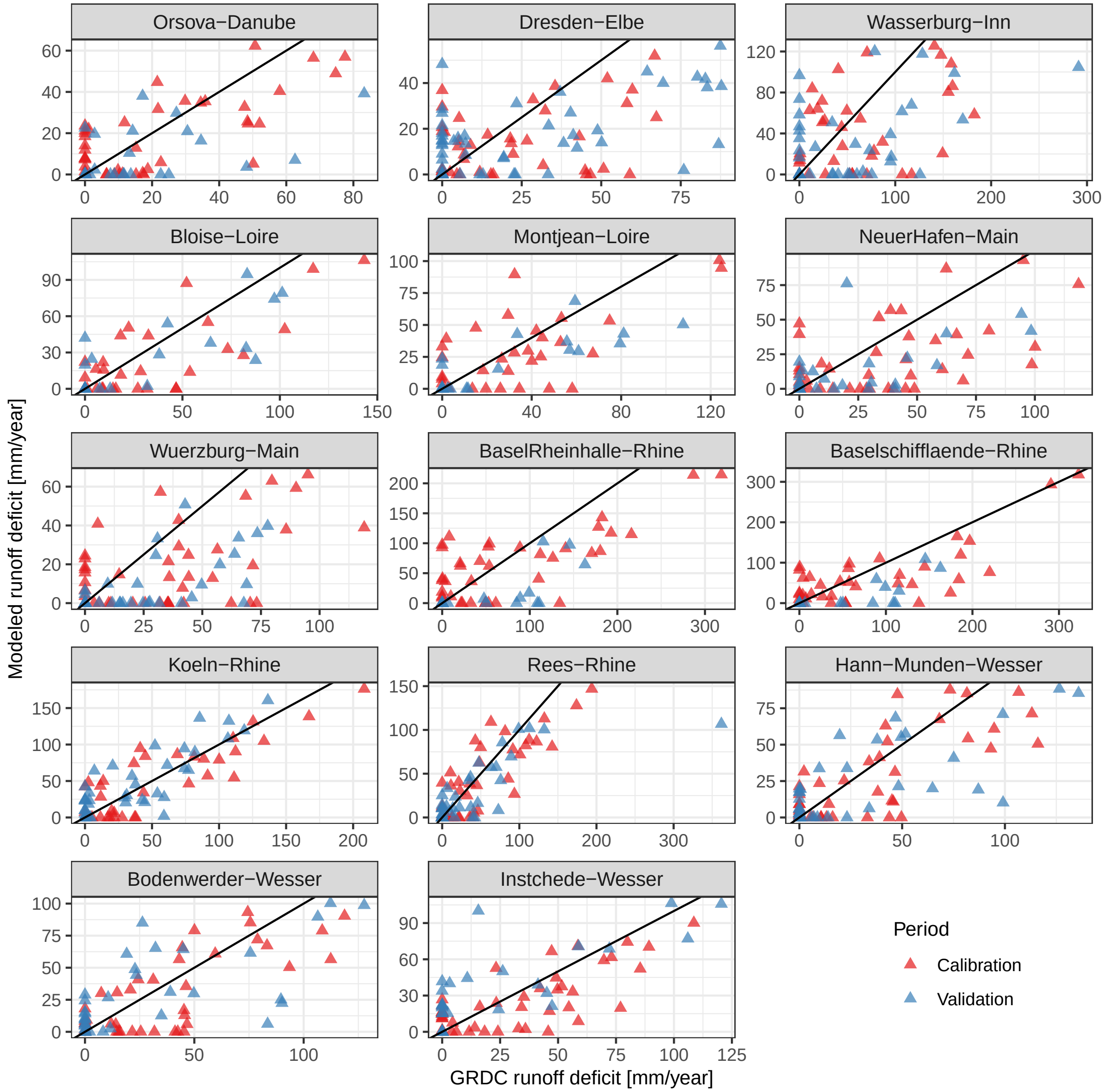
Figure 7. Distribution functions for BRNN(Lag), LSTM(PDSI), i.e. the best two models, GR1A and observed data (OBS) for the periods 1500–1800, 1800–1900 and 1900–2000 over Basel Rheinhalle-Rhine catchment. The values on the horizontal axis are transformed using the "probit" function. The colored labels indicate the most extreme drought years according to each model.

Figure 8. Reconstruction of runoff series for Köln- Main and Hann-Muenden Wesser Rivers. Blue line corresponds to the reconstructed series, the black and red lines represent the observed runoff for the calibration and validation period, respectively.



-0.65cm

Figure 9. Observed and simulated runoff for 14 selected catchments in the calibration and validation periods



-0.65cm

Figure 10. ~~Observed~~ The observed and simulated runoff deficit ~~of~~ based on the 33rd percentile threshold for 14 selected catchments ~~in~~ during the calibration and validation period.

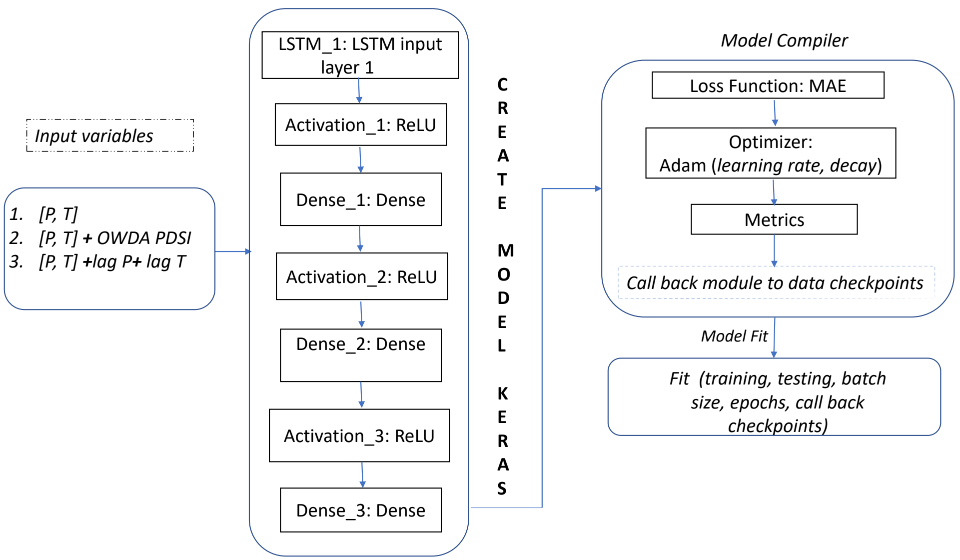


Figure A1. Structure of LSTM neural network model in KERAS environment for runoff predictions

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al.: Tensorflow: A system for large-scale machine learning, in: 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16), pp. 265–283, 2016.

Anonymous(2020): , https://www.euronews.com/green/2020/08/03/flowing-low-how-can-europe-use-climate-information-to-manage-dry-river-spells.

Armstrong, M. S., Kiem, A. S., and Vance, T. R.: Comparing instrumental, palaeoclimate, and projected rainfall data: Implications for water resources management and hydrological modelling, Journal of Hydrology: Regional Studies, 31, 100 728, 2020.

Arnold, T. B.: kerasR: R interface to the keras deep learning library, Journal of Open Source Software, 2, 296, 2017.

Ayzel, G., Kurochkina, L., and Zhuravlev, S.: The influence of regional hydrometric data incorporation on the accuracy of gridded reconstruction of monthly runoff, Hydrological Sciences Journal, pp. 1–12, 2020.

Boch, R. and Spötl, C.: Reconstructing palaeoprecipitation from an active cave flowstone, Journal of Quaternary Science, 26, 675–687, 2011.

Brázdil, R. and Dobrovolný, P.: Historical climate in Central Europe during the last 500 years, The Polish Climate in the European Context: An Historical Overview, p. 41, 2009.

Brázdil, R., Dobrovolný, P., Trnka, M., Kotyza, O., Řezníčková, L., Valášek, H., Zahradníček, P., and Štěpánek, P.: Droughts in the Czech Lands, 1090-2012AD., Climate of the Past, 9, 2013.

Brázdil, R., Kiss, A., Luterbacher, J., Nash, D. J., and Řezníčková, L.: Documentary data and the study of past droughts: a global state of the art, Climate of the Past, 14, 1915–1960, 2018.

Brázdil, R., Demarée, G. R., Kiss, A., Dobrovolný, P., Chromá, K., Trnka, M., Dolák, L., Řezníčková, L., Zahradníček, P., Limanowka, D., et al.: The extreme drought of 1842 in Europe as described by both documentary data and instrumental measurements., Climate of the Past, 15, 2019.

Breiman, L.: Random forests, Machine learning, 45, 5–32, 2001.

Brodie, F. J.: The great drought of 1893, and its attendant meteorological phenomena, Quarterly Journal of the Royal Meteorological Society, 20, 1–30, 1894.

Büntgen, U., Frank, D. C., Nievergelt, D., and Esper, J.: Summer temperature variations in the European Alps, AD 755–2004, Journal of Climate, 19, 5606–5623, 2006.

Büntgen, U., Franke, J., Frank, D., Wilson, R., González-Rouco, F., and Esper, J.: Assessing the spatial signature of European climate reconstructions, Climate Research, 41, 125–130, 2010.

Caillouet, L., Vidal, J.-P., Sauquet, E., Devers, A., and Graff, B.: Ensemble reconstruction of spatio-temporal extreme low-flow events in France since 1871, Hydrology and Earth System Sciences, 21, 2923–2951, 2017.

Casas-Gómez, P., Sánchez-Salguero, R., Ribera, P., and Linares, J. C.: Contrasting Signals of the Westerly Index and North Atlantic Oscillation over the Drought Sensitivity of Tree-Ring Chronologies from the Mediterranean Basin, Atmosphere, 11, 644, 2020.

Casty, C., Wanner, H., Luterbacher, J., Esper, J., and Böhm, R.: Temperature and precipitation variability in the European Alps since 1500, International Journal of Climatology: A Journal of the Royal Meteorological Society, 25, 1855–1880, 2005.

Chen, X., Huang, J., Han, Z., Gao, H., Liu, M., Li, Z., Liu, X., Li, Q., Qi, H., and Huang, Y.: The importance of short lag-time in the runoff forecasting model based on long short-term memory, Journal of Hydrology, 589, 125 359, 2020.

Contreras, P., Orellana-Alvear, J., Muñoz, P., Bendix, J., and Célleri, R.: Influence of Random Forest Hyperparameterization on Short-Term Runoff Forecasting in an Andean Mountain Catchment, Atmosphere, 12, 238, 2021.

Cook, E. R., Seager, R., Kushnir, Y., Briffa, K. R., Büntgen, U., Frank, D., Krusic, P. J., Tegel, W., van der Schrier, G., Andreu-Hayles, L., et al.: Old World megadroughts and pluvials during the Common Era, Science advances, 1, e1500 561, 2015.

Dobrovolný, P., Moberg, A., Brázdil, R., Pfister, C., Glaser, R., Wilson, R., van Engelen, A., Limanówka, D., Kiss, A., Halíčková, M., et al.: Monthly, seasonal and annual temperature reconstructions for Central Europe derived from documentary evidence and instrumental records since AD 1500, Climatic change, 101, 69–107, 2010.

Emile-Geay, J., McKay, N. P., Kaufman, D. S., Von Gunten, L., Wang, J., Anchukaitis, K. J., Abram, N. J., Addison, J. A., Curran, M. A., Evans, M. N., et al.: A global multiproxy database for temperature reconstructions of the Common Era, Scientific data, 4, 170 088, 2017.

Fathi, M. M., Awadallah, A. G., Abdelbaki, A. M., and Haggag, M.: A new Budyko framework extension using time series SARIMAX model, Journal of Hydrology, 570, 827–838, 2019.

Fekete, B. M., Vörösmarty, C. J., and Grabs, W.: Global, composite runoff fields based on observed river discharge and simulated water balances, 1999.

Foresee, F. D. and Hagan, M. T.: Gauss-Newton approximation to Bayesian learning, in: Proceedings of international conference on neural networks (ICNN'97), vol. 3, pp. 1930–1935, IEEE, 1997.

Ghiggi, G., Humphrey, V., Seneviratne, S. I., and Gudmundsson, L.: GRUN: an observation-based global gridded runoff dataset from 1902 to 2014, Earth System Science Data, 11, 1655–1674, 2019.

Ghiggi, G., Humphrey, V., Seneviratne, S., and Gudmundsson, L.: G-RUN ENSEMBLE: A Multi-Forcing Observation-Based Global Runoff Reanalysis, Water Resources Research, 57, e2020WR028 787, 2021.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Hanel, M., Rakovec, O., Markonis, Y., Máca, P., Samaniego, L., Kyselỳ, J., and Kumar, R.: Revisiting the recent European droughts from a long-term perspective, Scientific reports, 8, 1–11, 2018.

Hansson, D., Eriksson, C., Omstedt, A., and Chen, D.: Reconstruction of river runoff to the Baltic Sea, AD 1500–1995, International Journal of Climatology, 31, 696–703, 2011.

Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset, International journal of climatology, 34, 623–642, 2014.

Haslinger, K. and Blöschl, G.: Space-time patterns of meteorological drought events in the European Greater Alpine Region over the past 210 years, Water Resources Research, 53, 9807–9823, 2017.

Hochreiter, S. and Schmidhuber, J.: Long short-term memory, Neural computation, 9, 1735–1780, 1997.

Hu, C., Wu, Q., Li, H., Jian, S., Li, N., and Lou, Z.: Deep learning with a long short-term memory networks approach for rainfall-runoff simulation, Water, 10, 1543, 2018.

Im, S., Kim, H., Kim, C., and Jang, C.: Assessing the impacts of land use changes on watershed hydrology using MIKE SHE, Environmental geology, 57, 231, 2009.

Ionita, M., Tallaksen, L., Kingston, D., Stagge, J., Laaha, G., Van Lanen, H., Scholz, P., Chelcea, S., and Haslinger, K.: The European 2015 drought from a climatological perspective, Hydrology and Earth System Sciences, 21, 1397–1419, 2017.

Jeong, J., Barichivich, J., Peylin, P., Haverd, V., McGrath, M. J., Vuichard, N., Evans, M. N., Babst, F., and Luyssaert, S.: Using the International Tree-Ring Data Bank (ITRDB) records as century-long benchmarks for global land-surface models, Geoscientific Model Development, 14, 5891–5913, 2021.

Ji, Y., Dong, H.-T., Xing, Z.-X., Sun, M.-X., Fu, Q., and Liu, D.: Application of the decomposition-prediction-reconstruction framework to medium-and long-term runoff forecasting, Water Supply, 21, 696–709, 2021.

Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.

Kolachian, R. and Saghafian, B.: Hydrological drought class early warning using support vector machines and rough sets, Environmental Earth Sciences, 80, 1–15, 2021.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall–runoff modelling using long short-term memory (LSTM) networks, Hydrology and Earth System Sciences, 22, 6005–6022, 2018.

Kress, A., Saurer, M., Siegwolf, R. T., Frank, D. C., Esper, J., and Bugmann, H.: A 350 year drought reconstruction from Alpine tree ring stable isotopes, Global Biogeochemical Cycles, 24, 2010.

Kress, A., Hangartner, S., Bugmann, H., Büntgen, U., Frank, D. C., Leuenberger, M., Siegwolf, R. T., and Saurer, M.: Swiss tree rings reveal warm and wet summers during medieval times, Geophysical Research Letters, 41, 1732–1737, 2014.

Krysanova, V., Vetter, T., and Hattermann, F.: Detection of change in drought frequency in the Elbe basin: comparison of three methods, Hydrological Sciences Journal, 53, 519–537, 2008.

Kuhn, M.: Caret: classification and regression training, Astrophysics Source Code Library, pp. ascl–1505, 2015.

Kwak, J., Lee, J., Jung, J., and Kim, H. S.: Case Study: Reconstruction of Runoff Series of Hydrological Stations in the Nakdong River, Korea, Water, 12, 3461, 2020.

Laaha, G., Gauster, T., Tallaksen, L. M., Vidal, J.-P., Stahl, K., Prudhomme, C., Heudorfer, B., Vlnas, R., Ionita, M., Van Lanen, H. A., et al.: The European 2015 drought from a hydrological perspective, Hydrology and Earth System Sciences, 21, 3001, 2017.

Legates, D. R. and McCabe Jr, G. J.: Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation, Water resources research, 35, 233–241, 1999.

Leggewie, C. and Mauelshagen, F.: Climate change and cultural transition in Europe, Brill, 2018.

Li, Y., Wei, J., Wang, D., Li, B., Huang, H., Xu, B., and Xu, Y.: A Medium and Long-Term Runoff Forecast Method Based on Massive Meteorological Data and Machine Learning Algorithms, Water, 13, 1308, 2021.

Ljungqvist, F. C., Krusic, P. J., Sundqvist, H. S., Zorita, E., Brattström, G., and Frank, D.: Northern Hemisphere hydroclimate variability over the past twelve centuries, Nature, 532, 94–98, 2016.

Ljungqvist, F. C., Piermattei, A., Seim, A., Krusic, P. J., Büntgen, U., He, M., Kirdyanov, A. V., Luterbacher, J., Schneider, L., Seftigen, K., et al.: Ranking of tree-ring based hydroclimate reconstructions of the past millennium, Quaternary Science Reviews, 230, 106 074, 2020.

Luoto, T. P. and Nevalainen, L.: Quantifying climate changes of the Common Era for Finland, Climate Dynamics, 49, 2557–2567, 2017.

Luterbacher, J., Dietrich, D., Xoplaki, E., Grosjean, M., and Wanner, H.: European seasonal and annual temperature variability, trends, and extremes since 1500, Science, 303, 1499–1503, 2004.

MacKay, D. J.: A practical Bayesian framework for backpropagation networks, Neural computation, 4, 448–472, 1992.

Manabe, S.: Climate and the ocean circulation: I. The atmospheric circulation and the hydrology of the earth’s surface, Monthly Weather Review, 97, 739–774, 1969.

Markonis, Y. and Koutsoyiannis, D.: Scale-dependence of persistence in precipitation records, Nature Climate Change, 6, 399–401, 2016.

Markonis, Y., Hanel, M., Máca, P., Kyselỳ, J., and Cook, E.: Persistent multi-scale fluctuations shift European hydroclimate to its millennial boundaries, Nature communications, 9, 1–12, 2018.

Martínez-Sifuentes, A. R., Villanueva-Díaz, J., and Estrada-Ávalos, J.: Runoff reconstruction and climatic influence with tree rings, in the Mayo river basin, Sonora, Mexico, iForest-Biogeosciences and Forestry, 13, 98, 2020.

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., and Houston, T. G.: An overview of the global historical climatology network-daily database, Journal of Atmospheric and Oceanic Technology, 29, 897–910, 2012.

Menne, M. J., Williams, C. N., Gleason, B. E., Rennie, J. J., and Lawrimore, J. H.: The global historical climatology network monthly temperature dataset, version 4, Journal of Climate, 31, 9835–9854, 2018.

Middelkoop, H., Daamen, K., Gellens, D., Grabs, W., Kwadijk, J. C., Lang, H., Parmet, B. W., Schädler, B., Schulla, J., and Wilke, K.: Impact of climate change on hydrological regimes and water resources management in the Rhine basin, Climatic change, 49, 105–128, 2001.

Moberg, A., Mohammad, R., and Mauritsen, T.: Analysis of the Moberg et al.(2005) hemispheric temperature reconstruction, Climate dynamics, 31, 957–971, 2008.

Moravec, V., Markonis, Y., Rakovec, O., Kumar, R., and Hanel, M.: A 250-year European drought inventory derived from ensemble hydrologic modeling, Geophysical Research Letters, 46, 5909–5917, 2019.

Mouelhi, S., Michel, C., Perrin, C., and Andréassian, V.: Linking stream flow to rainfall at the annual time step: the Manabe bucket model revisited, Journal of hydrology, 328, 283–296, 2006.

Murphy, C., Broderick, C., Burt, T. P., Curley, M., Duffy, C., Hall, J., Harrigan, S., Matthews, T. K., Macdonald, N., McCarthy, G., et al.: A 305-year continuous monthly rainfall series for the island of Ireland (1711–2016)., Climate of the past., 14, 413–440, 2018.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I—A discussion of principles, Journal of hydrology, 10, 282–290, 1970.

Nicault, A., Alleaume, S., Brewer, S., Carrer, M., Nola, P., and Guiot, J.: Mediterranean drought fluctuation during the last 500 years based on tree-ring data, Climate dynamics, 31, 227–245, 2008.

Okut, H.: Bayesian regularized neural networks for small n big p data, Artificial neural networks-models and applications, pp. 21–23, 2016.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling, Journal of hydrology, 303, 290–306, 2005.

Pauling, A., Luterbacher, J., Casty, C., and Wanner, H.: Five hundred years of gridded high-resolution precipitation reconstructions over Europe and the connection to large-scale circulation, Climate dynamics, 26, 387–405, 2006.

Peterson, T. C. and Vose, R. S.: An overview of the Global Historical Climatology Network temperature database, Bulletin of the American Meteorological Society, 78, 2837–2850, 1997.

Pfister, C., Brázdil, R., Glaser, R., Barriendos, M., Camuffo, D., Deutsch, M., Dobrovolný, P., Enzi, S., Guidoboni, E., Kotyza, O., et al.: Documentary evidence on climate in sixteenth-century Europe, Climatic change, 43, 55–110, 1999.

Pfister, C., Weingartner, R., and Luterbacher, J.: Hydrological winter droughts over the last 450 years in the Upper Rhine basin: a methodological approach, Hydrological Sciences Journal, 51, 966–985, 2006.

Proctor, C., Baker, A., Barnes, W., and Gilmour, M.: A thousand year speleothem proxy record of North Atlantic climate from Scotland, Climate Dynamics, 16, 815–820, 2000.

Quayle, R. G., Peterson, T. C., Basist, A. N., and Godfrey, C. S.: An operational near-real-time global temperature index, Geophysical research letters, 26, 333–335, 1999.

Reinecke, R., Müller Schmied, H., Trautmann, T., Andersen, L. S., Burek, P., Flörke, M., Gosling, S. N., Grillakis, M., Hanasaki, N., Koutroulis, A., et al.: Uncertainty of simulated groundwater recharge at different global warming levels: a global-scale multi-model ensemble study, Hydrology and Earth System Sciences, 25, 787–810, 2021.

Riechelmann, D. F. and Gouw-Bouman, M. T.: A review of climate reconstructions from terrestrial climate archives covering the first millennium AD in northwestern Europe, Quaternary Research, 91, 111–131, 2019.

Rivera, J. A., Araneo, D. C., and Penalba, O. C.: Threshold level approach for streamflow drought analysis in the Central Andes of Argentina: a climatological assessment, Hydrological Sciences Journal, 62, 1949–1964, 2017.

Sadaf, N., Součková, M., Godoy, M. R. V., Singh, U., Markonis, Y., Kumar, R., Rakovec, O., and Hanel, M.: Supporting data for A 500-year runoff reconstruction for European catchments, figshare[data set], https://doi.org/10.6084/m9.figshare.15178107, 2021.

Seiller, G., Anctil, F., and Perrin, C.: Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions, Hydrology and Earth System Sciences, 16, 1171–1189, 2012.

Senthil Kumar, A., Sudheer, K., Jain, S., and Agarwal, P.: Rainfall-runoff modelling using artificial neural networks: comparison of network types, Hydrological Processes: An International Journal, 19, 1277–1291, 2005.

Smith, K. A., Barker, L. J., Tanguy, M., Parry, S., Harrigan, S., Legg, T. P., Prudhomme, C., and Hannaford, J.: A multi-objective ensemble approach to hydrological modelling in the UK: an application to historic drought reconstruction, Hydrology and Earth System Sciences, 23, 3247–3268, 2019.

Su, W., Tao, J., Wang, J., and Ding, C.: Current research status of large river systems: a cross-continental comparison, Environmental Science and Pollution Research, 27, 39 413–39 426, 2020.

Sun, J., Liu, Y., Wang, Y., Bao, G., and Sun, B.: Tree-ring based runoff reconstruction of the upper Fenhe River basin, North China, since 1799 AD, Quaternary International, 283, 117–124, 2013.

Sung, J. H. and Chung, E.-S.: Development of streamflow drought severity–duration–frequency curves using the threshold level method, Hydrology and Earth System Sciences, 18, 3341–3351, 2014.

Swierczynski, T., Brauer, A., Lauterbach, S., Martín-Puertas, C., Dulski, P., von Grafenstein, U., and Rohr, C.: A 1600 yr seasonally resolved record of decadal-scale flood variability from the Austrian Pre-Alps, Geology, 40, 1047–1050, 2012.

Tejedor, E., de Luis, M., Cuadrat, J. M., Esper, J., and Saz, M. Á.: Tree-ring-based drought reconstruction in the Iberian Range (east of Spain) since 1694, International journal of biometeorology, 60, 361–372, 2016.

Thiesen, S., Darscheid, P., and Ehret, U.: Identifying rainfall-runoff events in discharge time series: a data-driven method based on information theory, Hydrology and Earth System Sciences, 23, 1015–1034, 2019.

Trouet, V., Diaz, H., Wahl, E., Viau, A., Graham, R., Graham, N., and Cook, E.: A 1500-year reconstruction of annual mean temperature for temperate North America on decadal-to-multidecadal time scales, Environmental Research Letters, 8, 024 008, 2013.

Tshimanga, R., Hughes, D., and Kapangaziwiri, E.: Initial calibration of a semi-distributed rainfall runoff model for the Congo River basin, Physics and Chemistry of the Earth, Parts A/B/C, 36, 761–774, 2011.

Uehlinger, U. F., Wantzen, K. M., Leuven, R. S., and Arndt, H.: The Rhine river basin, Acad. Pr., 2009.

van der Schrier, G., Allan, R. P., Ossó, A., Sousa, P. M., Van de Vyver, H., Van Schaeybroeck, B., Coscarelli, R., Pasqua, A. A., Petrucci, O., Curley, M., et al.: The 1921 European drought: Impacts, reconstruction and drivers, Climate of the Past Discussions, pp. 1–33, 2021.

Van Loon, A. F.: Hydrological drought explained, Wiley Interdisciplinary Reviews: Water, 2, 359–392, 2015.

Vansteenberge, S., Verheyden, S., Cheng, H., Edwards, R. L., Keppens, E., and Claeys, P.: Paleoclimate in continental northwestern Europe during the Eemian and early Weichselian (125-97 ka): insights from a Belgian speleothem., Climate of the Past, 12, 2016.

Wetter, O. and Pfister, C.: An underestimated record breaking event–why summer 1540 was likely warmer than 2003, Climate of the Past, 9, 41–56, 2013.

Wetter, O., Pfister, C., Weingartner, R., Luterbacher, J., Reist, T., and Trösch, J.: The largest floods in the High Rhine basin since 1268 assessed from documentary and instrumental evidence, Hydrological Sciences Journal, 56, 733–758, 2011.

Wilhelm, B., Arnaud, F., Sabatier, P., Crouzet, C., Brisset, E., Chaumillon, E., Disnar, J.-R., Guiter, F., Malet, E., Reyss, J.-L., et al.: 1400 years of extreme precipitation patterns over the Mediterranean French Alps and possible forcing mechanisms, Quaternary Research, 78, 1–12, 2012.

Wilson, R. J., Luckman, B. H., and Esper, J.: A 500 year dendroclimatic reconstruction of spring–summer precipitation from the lower Bavarian Forest region, Germany, International Journal of Climatology: A Journal of the Royal Meteorological Society, 25, 611–630, 2005.

Xiang, Z., Yan, J., and Demir, I.: A rainfall-runoff model with LSTM-based sequence-to-sequence learning, Water resources research, 56, e2019WR025 326, 2020.

Xoplaki, E., Luterbacher, J., Paeth, H., Dietrich, D., Steiner, N., Grosjean, M., and Wanner, H.: European spring and autumn temperature variability and change of extremes over the last half millennium, Geophysical Research Letters, 32, 2005.

Ye, L., Jabbar, S. F., Abdul Zahra, M. M., and Tan, M. L.: Bayesian Regularized Neural Network Model Development for Predicting Daily Rainfall from Sea Level Pressure Data: Investigation on Solving Complex Hydrology Problem, Complexity, 2021, 2021.

Yevjevich, V. M.: Objective approach to definitions and investigations of continental hydrologic droughts, An, Hydrology papers (Colorado State University); no. 23, 1967.

Zuo, G., Luo, J., Wang, N., Lian, Y., and He, X.: Two-stage variational mode decomposition and support vector regression for streamflow forecasting, Hydrology and Earth System Sciences, 24, 5491–5518, 2020.