

# A three-quantile bias correction with spatial transfer for the correction of simulated European river runoff to force ocean models

Stefan Hagemann<sup>\*</sup>, Thao T. Nguyen and Ha T. M. Ho-Hagemann

Institute of Coastal Systems – Analysis and Modelling, Helmholtz-Zentrum Hereon, Max-Planck-Str. 1, 21502 Geesthacht, Germany

<sup>\*</sup>Correspondence: Dr. Stefan Hagemann, stefan.hagemann@hereon.de

## 1 Abstract

2 In ocean or Earth system model applications, the riverine freshwater inflow is an important flux  
3 affecting salinity and marine stratification in coastal areas. However, in climate change studies,  
4 the river runoff based on climate model output often has large biases on local, regional or even  
5 basin-wide scales. If these biases are too large, the ocean model forced by the runoff will drift  
6 into a different climate state compared to the observed state, which is particularly relevant for  
7 semi-enclosed seas such as the Baltic Sea. To achieve low biases in riverine freshwater inflow  
8 in large-scale climate applications, a bias correction is required that can be applied in periods  
9 where runoff observations are not available and that allows spatial transferability of its  
10 correction factors. In order to meet these requirements, we have developed a three-quantile bias  
11 correction that includes different correction factors for low, medium and high percentile ranges  
12 of river runoff over Europe. Here, we present an experimental setup using the Hydrological  
13 Discharge (HD) model and its high-resolution (1/12°) grid. First, bias correction factors are  
14 derived at the locations of the downstream stations with available daily discharge observations  
15 for many European rivers. These factors are then transferred to the respective river mouths and  
16 mapped to neighbouring grid boxes belonging to ungauged catchments. The results show that  
17 the bias correction generally leads to an improved representation of river runoff. Especially  
18 over Northern Europe, where many rivers are regulated, the three-quantile bias correction  
19 provides an advantage compared to a bias correction that only corrects the mean bias of the  
20 river runoff. Evaluating two NEMO ocean model simulations in the German Bight indicated  
21 that the use of the bias corrected discharges as forcing leads to an improved simulation of sea  
22 surface salinity in coastal areas. Although in the present study, the bias correction is tailored to  
23 the high-resolution HD model grid over Europe, the methodology is suitable for any high-  
24 resolution model region with a sufficiently high coverage of river runoff observations. It is also  
25 noted that the methodology is applicable to river runoff based on climate hindcasts as well as  
26 on historical climate simulations where the sequence of weather events does not match the  
27 actual observed history. Therefore, it may also be applied in climate change simulations.

28 **Keywords:** Bias correction, river runoff, discharge, high resolution, Europe, sea-surface  
29 salinity

30

31

## 32 1 Introduction

33 River runoff (or discharge/streamflow) is an important component of the global hydrological  
34 cycle, accounting for about one-third of precipitation over land areas. It closes the water cycle  
35 between land and ocean and influences various ocean properties, in particular the salinity of  
36 coastal and semi-enclosed seas (e.g. Väli et al., 2013), the ocean stratification in shelf areas  
37 (e.g. Hordoir and Meier, 2010) such as the German Bight (Becker et al., 1992), and the  
38 thermohaline circulation in different regions (e.g. Hordoir et al., 2008; Lehmann and  
39 Hinrichsen, 2000; Marzeion et al., 2007). In addition, river runoff and associated nutrient loads  
40 are important factors influencing marine ecosystem functioning (Daewel and Schrum, 2017).

41 Consequently, river runoff needs to be adequately represented in studies of the impacts of  
42 climate change on the marine environment or in coupled Earth system studies. In such studies,  
43 the atmospheric data used to force the respective ocean model are usually taken from climate  
44 models, reanalysis products or hydrological models. Here, it is desirable that the river runoff is  
45 consistent with the atmospheric forcing (e.g. Vinayachandran et al., 2015; Hagemann and  
46 Stacke, 2022), i.e. that the impact of weather events and trends in the atmospheric forcing is  
47 transferred via the river runoff into the ocean. In previous modelling studies, runoff was often  
48 taken from climatology or discharge observations, especially when hindcasts were used.  
49 However, this is not a recommended approach for climate change studies where consistently  
50 simulated river runoff should be used. Runoff from the driving climate, land surface or  
51 hydrological model will contain biases, e.g., due to biases in precipitation and/or uncertainties  
52 in the land surface representation of the model. Many simulations of historical daily river runoff  
53 show common biases in the tails of their distributions, with high discharges underestimated and  
54 low discharges overestimated (Farmer et al., 2018, and references therein). If the basin-wide  
55 biases are too large, a bias correction of the simulated discharge would be necessary to avoid  
56 the ocean model drifting into a different climate state compared to the observed state. This is  
57 particularly relevant for semi-enclosed seas such as the Baltic Sea. For example, for Baltic Sea  
58 ocean models, the mean long-term bias of river runoff must be less than 7% (Hagemann and  
59 Stacke, 2022).

60 The bias correction of river runoff is an approach that has been used particularly for short-  
61 term hydrological forecasts and ensemble predictions of up to six months. However, these  
62 approaches (see, e.g., those listed in Kim et al., 2021; Madadgar et al., 2014) are often  
63 specifically trimmed to flood forecasts. Therefore, they often require the existence of observed  
64 values from previous time steps so that that are not applicable in climate change studies, such  
65 as autoregression models (Kim et al., 2021) or components of a Bayesian forecasting system  
66 (Krzysztofowicz and Maranzano, 2004). Others like non-parametric methods based on  
67 Bayesian approaches as proposed by Brown and Seo (2010; 2012) need a large number of  
68 ensemble members (Madadgar et al., 2014).

69 Recently, bias correction of river runoff has also been applied in the context of climate  
70 change. Quantile mapping based approaches are often used for such bias correction, as this  
71 usually leads to a large improvement in the representation of discharge of the considered river.  
72 For example, Budhathoki et al. (2022) used quantile mapping to correct discharge bias in the  
73 Chao Phraya River basin (Thailand), and Daraio (2020) used it for two basins in New Jersey  
74 (USA). A criticism of using quantile-mapping in flood forecasting is that it does not maintain  
75 the pairing of corresponding simulated and observed flows (Madadgar et al., 2014). Madadgar  
76 et al. (2014) also noted that quantile mapping was not always successful in improving the initial  
77 forecast trajectory. In their application for the Sprague River (southern Oregon, USA), the skill

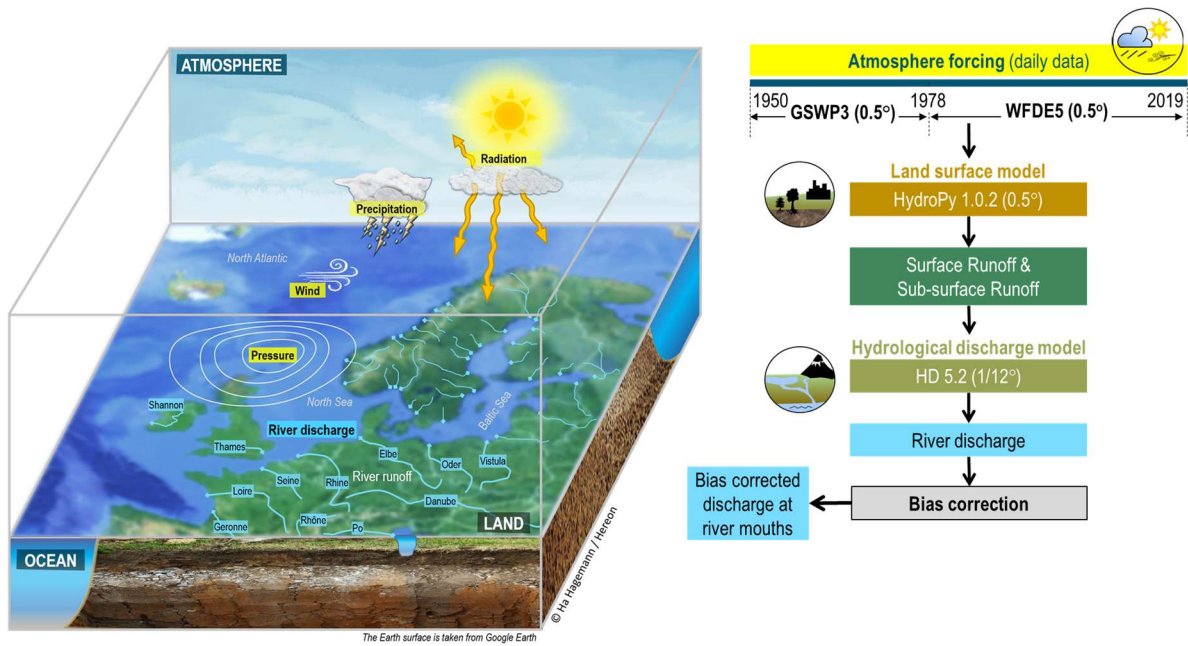
78 of the forecast actually deteriorated when the quantile mapping technique was used. Similarly,  
79 Malek et al. (2022) used a quantile mapping based bias correction of discharge and showed that  
80 ex-post corrections of simulated discharge do not necessarily reduce biases in the simulation of  
81 key processes and in some cases can severely degrade system simulations.

82 Consequently, the aim of the present study was to develop a bias correction method sufficient  
83 to meet the requirements of ocean models in large-scale climate change studies. Note that we  
84 did not aim for the most accurate reproduction of observed discharge characteristics, as required  
85 for short-term hydrological predictions and flood forecasts used by water resource decision  
86 makers (e.g. Shi et al., 2008). In order to maintain a high degree of temporal consistency of  
87 simulated runoff with the meteorological patterns in the driving (on- or offline) climate model  
88 (or data), a bias correction with as little fitting or modification of the daily sequence of runoff  
89 curves as possible is desired. Thus, our target is a simple bias correction that corrects the mean  
90 bias and the tail biases of the discharge distribution in climate change applications of ocean or  
91 coupled system models. The bias correction factors should be transferable from downstream  
92 stations to river mouths as well as to neighbouring ungauged catchments. Furthermore, it should  
93 be applicable to climate model or Earth system model data that lack the observed sequence of  
94 actual discharge events. Therefore, we decided to not apply methods that employ detailed  
95 modifications of the discharge curves for specific rivers such as those methods that use complex  
96 matrix arithmetic of observed and simulated discharge time series (e.g. Zhao et al., 2011), or  
97 the common quantile-mapping approaches, The latter are conducted using a lot of bins, so that  
98 the bias in the discharge curve of a specific river can be strongly reduced. However, these  
99 detailed correction factors for every bin may likely not be transferred to other locations. It may  
100 work for the same river if station and river mouth are relatively close to each other, but certainly  
101 may not be valid for the transfer to neighbouring catchments.

102 The manuscript is organised as follows. Section 2 describes how the simulated discharges  
103 were generated and the newly developed bias correction methodology, as well as the data,  
104 models and metrics used in this study. Sections 3 and 4 evaluate the simulated and bias corrected  
105 discharges and present the effects of the bias correction for station locations and sea basin  
106 inflows, respectively. Finally, Section 5 concludes with a summary and conclusions.

## 107 **2 Data and Methods**

108 To generate the freshwater inflow from rivers to the ocean, we used an experimental setup  
109 analogous to Hagemann and Stacke (2022). Here we used two atmospheric forcing datasets  
110 (Sect. 2.1) and the same model chain of two large-scale hydrological models. The global  
111 hydrological model HydroPy (Sect. 2.2) was used to generate the input to the Hydrological  
112 Discharge (HD) model (Sect. 2.3) at the resolution of the atmospheric forcing data ( $0.5^\circ$ ). These  
113 input data of surface and sub-surface runoff were then interpolated onto the HD model grid and  
114 the HD model was used to simulate daily discharges from land to sea. Subsequently, we bias  
115 corrected these time series as described in Section 2.4 to generate bias corrected discharges at  
116 coastal ocean boxes of the European HD model domain from 1901-2019. Note that we  
117 combined the simulations based on two different atmospheric forcing datasets to cover the  
118 whole 20<sup>th</sup> century and to include the more recent years in the bias corrected discharge time  
119 series. Such an approach was also used in the second phase (ISIMIP, 2023) of the Inter-Sectoral  
120 Impact Model Inter-Comparison Project (ISIMIP; Warszawski et al., 2014). Figure 1  
121 summarises the experimental setup. Section 2.5 refers to the observational data that are used in  
122 the evaluation of the model results. Finally, the evaluation metrics used in the analysis of the  
123 results are presented in Sect. 2.7.



124

125

126 **Figure 1.** Overview on the main steps of generating bias corrected river discharge at HD  
 127 river mouths.

128

## 129 2.1 Atmospheric forcing

130 We used two atmospheric datasets comprising daily data of various near-surface atmospheric  
 131 variables. They have been used as meteorological forcing datasets in several climate impact  
 132 assessments and are recommended by ISIMIP (2023). Both datasets were specifically generated  
 133 to force global hydrological models for hindcast simulations. They are based on re-analysis  
 134 products from different weather forecast centres and bias-correction procedures were applied  
 135 by the respective creators to improve their data.

136 The Global Soil Wetness Project Phase 3 (GSWP3; Dirmeyer et al., 2006; Kim, 2017)  
 137 dataset is available at 0.5° resolution from 1901-2014. To generate the GSWP3 dataset, Kim  
 138 (2017) dynamically downscaled the 20<sup>th</sup> Century Reanalysis (Compo et al., 2011) onto the T248  
 139 (~0.5°) grid using a spectral nudging technique (Yoshimura and Kanamitsu, 2008) in a Global  
 140 Spectral Model. Observation-based bias correction procedures were then applied to the  
 141 downscaled data to obtain daily time series.

142 To generate the WFDE5 dataset, Cucchi et al. (2020) applied the WATCH Forcing Data  
 143 methodology (Weedon et al., 2011) to surface meteorological variables from the ERA5  
 144 reanalysis (Hersbach et al., 2020) to obtain bias corrected time series. ERA5 is the fifth  
 145 generation of atmospheric reanalysis produced by the European Centre for Medium-Range  
 146 Weather Forecasts (ECMWF). WFDE5 is provided at 0.5° spatial resolution from 1979-2019.  
 147 Mengel et al. (2021) stated that WFDE5 is considered as the more realistic dataset, especially  
 148 with respect to day-to-day variability for variables for which the monthly mean values were  
 149 bias corrected, such as precipitation and temperature. For more information on application and



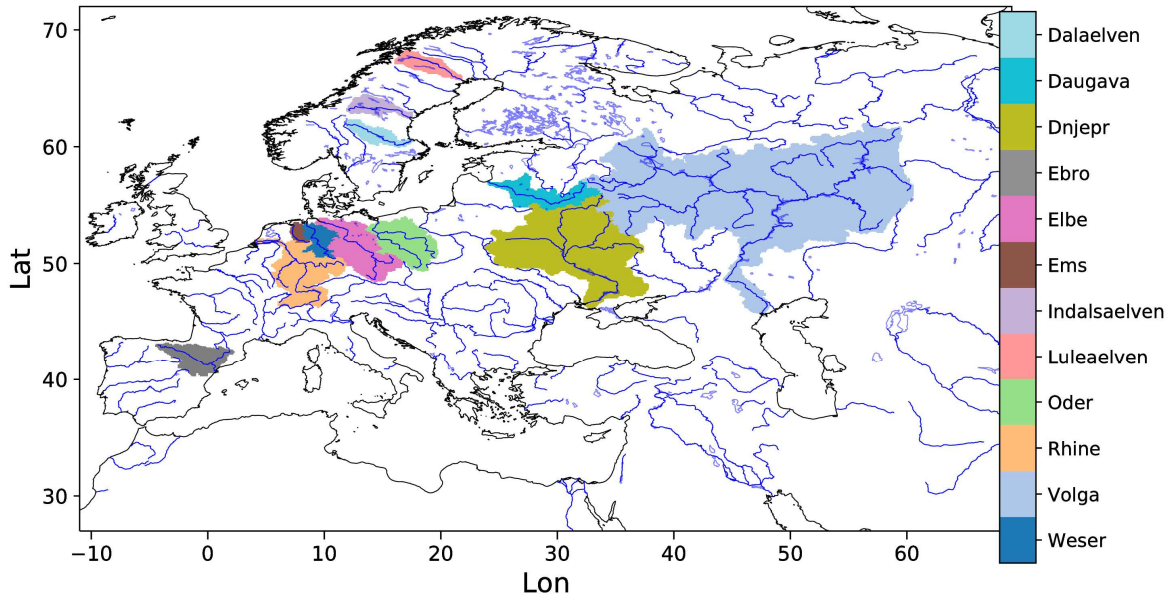
150 evaluation of both datasets, see, e.g., Mengel et al. (2021) and references therein, Hassler and  
151 Lauer (2021), (Arora et al., 2023).

## 152 **2.2 HydroPy setup**

153 HydroPy (Stacke and Hagemann, 2021) is a state-of-the-art global hydrology model for  
154 which no model calibration was performed for its setup. Within global hydrological modelling,  
155 the usage of uncalibrated models is rather common (see, e.g., Haddeland et al., 2011), even  
156 though some models exist that are calibrated for global studies. In the present study, HydroPy  
157 was driven by daily forcing data from 1901-2019. Daily input fields of surface and subsurface  
158 runoff were generated at a resolution of  $0.5^\circ$ . Analogous to the ERA5 forced simulation in  
159 Hagemann and Stacke (2022), precipitation, 2m temperature, downwelling shortwave and  
160 longwave radiation, 2m specific humidity, surface pressure and 10m wind are used as forcing  
161 from the respective forcing dataset. We performed a spin-up simulation over 50 iterations of  
162 the year 1901 with the GSWP3 forcing (cf. Stacke and Hagemann, 2021) to initialize the  
163 storages in the HydroPy model and to avoid any drift during the actual simulation period. We  
164 then forced HydroPy with the GSWP3 data from 1901-1978 and continued with the WFDE5  
165 data from 1979-2019. We also conducted a GSWP3 forced simulation from 1979-2014 in order  
166 to derive bias correction parameters for the earlier period. For our analysis, we focus on the  
167 years from 1950 onwards so that we have an additional transient spin-up of 49 years.

## 168 **2.3 HD model setup**

169 The HD model (Hagemann et al., 2020) is a well-established river routing model that is  
170 implemented in a range of global and regional model systems. As noted in Hagemann et al.  
171 (2020), no river specific parameter adjustments were conducted in the HD model to enable its  
172 applicability for climate change studies and over catchments, where no daily discharges are  
173 available at a downstream station. To simulate discharge with the HD model, we used the daily  
174 input fields of surface and subsurface runoff that were generated by HydroPy from the GSWP3  
175 and WFDE5 data (see Sect. 2.2). As the time step of these runoff data is one day, the time step  
176 of the HD model was also set to one day. However, an internal time step of 0.5 hours is used  
177 for the flow within the river, as the minimum travel time through a grid box is limited by the  
178 chosen time step. The HD model v5.2.0 (Hagemann et al., 2023) was applied over the European  
179 domain, which covers the land areas between  $-11^\circ\text{W}$  to  $69^\circ\text{E}$  and  $27^\circ\text{N}$  to  $72^\circ\text{N}$ . The domain,  
180 along with a number of rivers specifically noted in this study, is shown in Figure 2. In the  
181 following, we refer to the WFDE5-based discharges as HDW and to the GSWP3-based  
182 discharges as HDG. The corresponding bias-corrected discharges are referred to as HD-BC in  
183 general and HDW-BC and HDG-BC in particular.



184

185

**Figure 2.** European HD model domain and catchment areas for selected rivers

186

## 2.4 Bias correction of river runoff

187

188

189

190

191

192

193

194

195

196

197

198

199

We have developed a bias correction method for river runoff that uses correction factors for three quantiles and includes a spatial transfer of these factors. We note that our three-quantile bias correction is similar to a very coarse quantile mapping. The latter has been introduced in climate change impact research to correct for significant biases in data produced by global and regional climate models. Quantile mapping is a distribution mapping in which the distribution function of climate values is corrected to match the observed distribution function. Details of such mapping applied to precipitation and surface air temperature can be found, for example, Piani et al. (2010) and Teutschbein and Seibert (2012). Our bias correction method involves several steps. First, different correction factors for low, medium and high percentiles are calculated at the station locations and then applied at the respective river mouths. Finally, an interpolation is performed to neighbouring coastal mouth points for which no downstream observations are available in the respective catchment. This procedure is summarised in Figure 3. The three percentile ranges for daily discharge  $q_i$  are classified by

200

201

202

- Low (L):  $q_i \leq Q_p$
- Medium (M):  $Q_p < q_i < Q_{100-p}$
- High (H):  $q_i \geq Q_{100-p}$

203

204

205

206

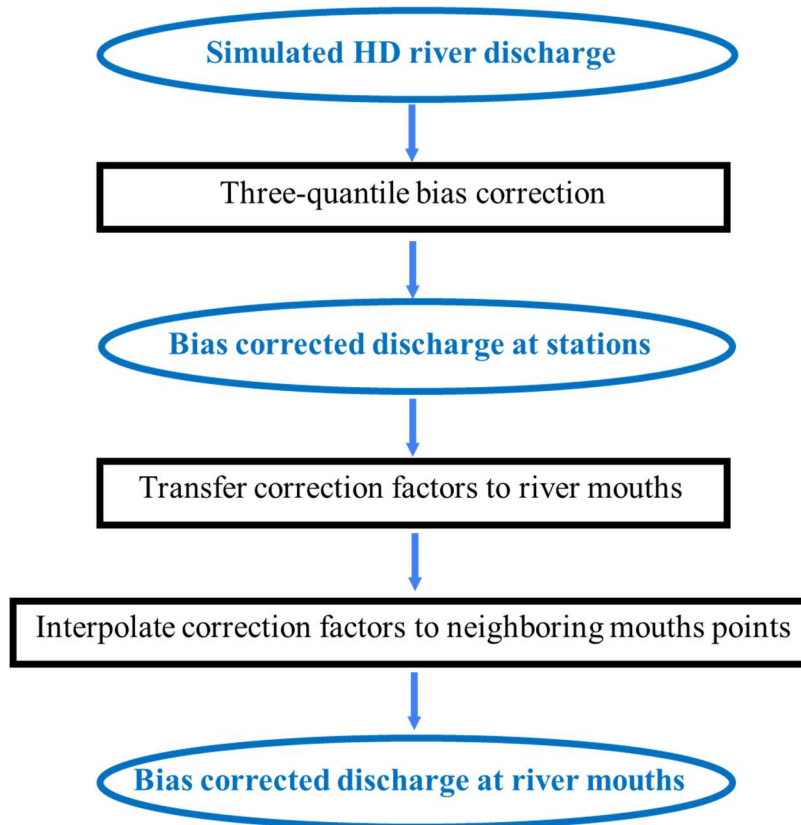
207

208

Here,  $Q_p$  denotes the  $p^{\text{th}}$  percentile of the daily discharge and  $p$  was set to 20. The percentiles  $Q_p$  and  $Q_{100-p}$  were determined separately for the observed and the simulated discharges at the downstream station locations and then the mean discharges  $\bar{q}_R$  were calculated for the three percentile ranges  $R \in \{L, M, H\}$ . Note that for these calculations only those days were considered for which an observed discharge was available. Then, the mean bias  $b_R$  (in %) was calculated for each percentile range and a correction factor  $f_R$  to remove the bias was derived as

209

$$f_R = \frac{100}{b_R + 100}$$



211

212 **Figure 3.** Steps to derive bias corrected discharge at river mouths from simulated  
 213 discharges.

214 For the evaluation of the bias correction in Sect. 3, these correction factors were applied to  
 215 the simulated discharges at the station locations. As the correction factors are independent of  
 216 the absolute amount of discharge, they could be applied to the respective river mouths. For each  
 217 river mouth with more than one inflow ( $j > 1$ ) for which a correction factor  $f_{R,j}$  is determined, a  
 218 combined correction factor is obtained by calculating an average weighted by the respective  
 219 mean inflows  $Q_j$ .

220 
$$\bar{f}_R = \frac{\sum_j f_{R,j} * Q_j}{\sum_j Q_j}$$

221 From these river mouths, an interpolation is performed to neighbouring coastal mouth points  
 222 for which no downstream observations are available in the respective catchment. This  
 223 interpolation was motivated by the fact that the general pattern of bias of neighbouring rivers  
 224 is often similar (cf. Sect. 3.1). The interpolation is performed by inverse distance weighting  
 225 from the four closest (or fewer) river mouths within a search radius of 200 km. If no river mouth  
 226 with a correction factor was found within the search radius, the correction factor was set to one  
 227 (i.e. no correction).

228 Note that the bias correction can lead to spurious daily jumps in discharge when the  
 229 percentile boundary is crossed and the bias correction factors differ between the percentile

230 ranges. In order to reduce this effect, a smoothing radius of  $\Delta s = 0.05$  was introduced around  
 231 the percentile boundaries, which was applied at both station locations and river mouths.

232 For  $(1 - \Delta s) * Q_p < q_i < (1 + \Delta s) * Q_p$ :

$$233 \quad \tilde{q}_i = q_i * (f_L + (f_M - f_L)) * \frac{(q_i - (1 - \Delta s) * Q_p)}{2 * \Delta s * Q_p}$$

234 For  $(1 - \Delta s) * Q_{100-p} < q_i < (1 + \Delta s) * Q_{100-p}$ :

$$235 \quad \tilde{q}_i = q_i * (f_M + (f_H - f_M)) * \frac{(q_i - (1 - \Delta s) * Q_{100-p})}{2 * \Delta s * Q_{100-p}}$$

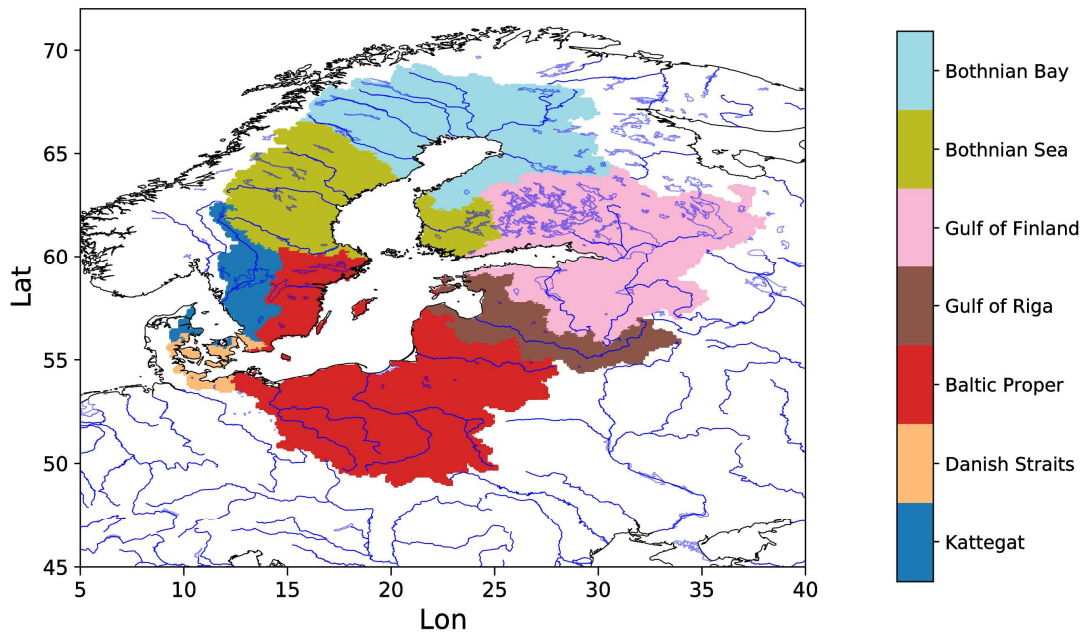
236 The bias correction procedure corrects the days that fall into the different percentile ranges.  
 237 However, this does not necessarily mean that it also corrects the whole distribution into the  
 238 three percentile ranges. Particularly, if the biases in these ranges are quite different, the days  
 239 may change their class and order within the distribution.

240 In order to apply the three-quantile bias correction to the simulated discharge time series  
 241 from 1901-2019, two sets of bias correction factors were derived. The first set uses HDW and  
 242 discharge station observations for the period 1979-2014. This set was used to bias correct the  
 243 simulated discharge at HD river mouths from 1979-2019. The second set uses a further  
 244 discharge simulation where we continued HDG utilizing the GSWP3 forcing up to 2014. Again,  
 245 the set of bias correction factors was derived for the period 1979-2014 using discharge station  
 246 observations. This set was then used to bias correct the simulated discharge at the HD river  
 247 mouths from 1901-1978.

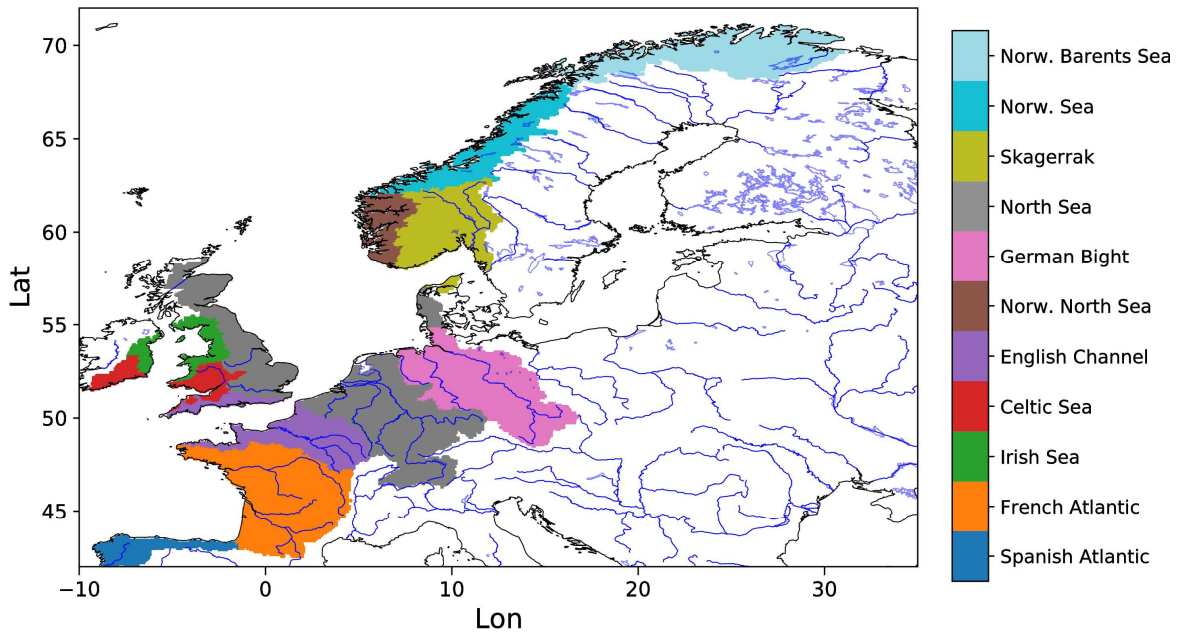
## 248 **2.5 Observed discharge data**

249 We used available daily discharge data from downstream gauges for many rivers across Europe  
 250 with a catchment area of about 1000 km<sup>2</sup> or more. These station data were obtained from Global  
 251 Runoff Data Centre and various agencies and institutions listed in table 2 of Hagemann and  
 252 Stacke (2022). In addition, French discharge data were accessed from the E.U. Copernicus  
 253 Marine Service Information. In order to allow an assessment of the discharge at the river  
 254 mouths, we considered basin-wide estimates from three different sources.

255 For the Baltic Marine Environment Protection Commission – also known as the Helsinki  
 256 Commission (HELCOM), Svendsen and Gustafsson (2022) provided annual waterborne  
 257 inflows into the seven main sub-basins of the Baltic Sea (Figure 4 – upper panel) from 1995 to  
 258 2020. Waterborne inflows comprise the sum of river runoff and direct inflows, i.e. flows from  
 259 point sources discharging directly into the Baltic Sea. These point sources are not included in  
 260 our experimental setup or in the bias correction. However, their contribution to the total  
 261 waterborne inflow to the Baltic Sea is only about 1% (HELCOM, 1998).



262



263

264 **Figure 4.** Selected HELCOM (upper panel) and OSPAR (lower panel) basins for which  
 265 inflows are considered. For OSPAR, the Spanish Atlantic basin is limited to the coast  
 266 of Northern Spain.

267 Under the umbrella of the OSPAR Convention (Convention for the Protection of the Marine  
 268 Environment of the North-East Atlantic), the IGC-EMO (Intersessional Correspondence Group  
 269 for Eutrophication Modelling) database (Lenhart et al., 2010) of daily riverine freshwater  
 270 inflows and nutrient loads was compiled by Van Leeuwen and Lenhart (2021), covering the  
 271 major rivers discharging into the Baltic Sea, the North Sea and the Northeast Atlantic. An  
 272 updated database covering a total of 370 rivers was mapped onto the flow grid of the European  
 273 1/12° domain of the HD model by Van Leeuwen and Hagemann (2023). The associated  
 274 catchment areas of these rivers, which flow into a particular specific sea basin, do not cover the  
 275 entire catchment area of the respective basin (see Table 1) so that the total inflow of the sea

276 basin is underrepresented by the IGC-EMO data. To generate basin-wide estimates, we have  
 277 up-scaled these values by dividing the integrated IGC-EMO river discharges in a basin by the  
 278 fractional coverage of the entire basin catchment on the HD grid. Basin estimates for which the  
 279 fractional coverage is less than 75% are considered to be highly uncertain and are therefore  
 280 provided for completeness only, but are not included in the assessment of simulated inflows.

281 **Table 1.** Sea basin catchment areas on the HD model grid and the fractional catchment  
 282 coverage of the associated IGC-EMO rivers.

Sea basin	HD Area [km <sup>2</sup> ]		
	IGC-EMO	Total	Coverage
Baltic Sea	1513967	1671823	90.6%
Bothnian Bay	238898	258420	92.4%
Bothnian Sea	199908	219375	91.1%
Gulf of Finland	379628	412412	92.1%
Gulf of Riga	124386	134025	92.8%
Baltic Proper	494929	551295	89.8%
Danish Straits	6731	19417	34.7%
Kattegat	69487	76876	90.4%
Norwegian Barents Sea	0	81004	0.0%
Norwegian Sea	0	58408	0.0%
Skagerrak	89060	101787	87.5%
North Sea	514334	599755	85.8%
German Bight	201233	208807	96.4%
Norwegian North Sea	4590	31327	14.7%
English Channel	94327	122235	77.2%
Celtic Sea	41122	44845	91.7%
Irish Sea	29748	35584	83.6%
French Atlantic	207657	257981	80.5%
Northern Spanish Atlantic	17692	46574	38.0%

283

284 In addition, we used estimates of long-term mean sub-basin-wide inflows to the North Sea  
 285 and Northeast Atlantic, published directly by OSPAR (Farkas and Skarbøvik, 2021). Figure 4  
 286 (lower panel) shows the selected OSPAR basins for which the inflows are considered. It should  
 287 be noted that the sea basin inflows provided by the different OSPAR countries are not  
 288 consistent. Some countries include discharge estimates for unmonitored areas, while others do  
 289 not (Table 2). In addition, the sea basin catchment coverage of the monitored areas varies  
 290 between the countries. Note also that we have excluded the Spanish Atlantic from our  
 291 comparisons for the following reason. Here, we limited the Spanish Atlantic basin to the coast  
 292 of northern Spain (see Figure 4 – lower panel) to allow a comparison with the IGC-EMO data  
 293 as the IGC-EMO data only cover rivers in this region, hereafter referred to as NSpA. These  
 294 rivers cover about 38% of the total NSpA area on the HD model grid (Table 1), while the  
 295 OSPAR data for NSpA cover about 50% (23201 km<sup>2</sup>; Farkas and Skarbøvik, 2021). However,  
 296 the associated IGC-EMO discharge from 1961-1990 (629 m<sup>3</sup>/s) is 75 % larger than the OSPAR  
 297 long-term mean average (359 m<sup>3</sup>/s). Therefore, both inflow values are unlikely to be  
 298 representative for the NSpA region and this region is not considered in the following.

299 **Table 2.** Country catchment coverage of OSPAR data and inclusion of estimates for  
 300 unmonitored areas (Borgvang et al., 2008). NI means that no information on the  
 301 coverage was provided.

Country	Coverage	Unmonitored
Belgium	> 90%	No
Denmark	NI	Yes <sup>3</sup>
France	84%	Yes
Germany	>90%	No <sup>1</sup>
Ireland	NI	Yes
Netherlands	>90%	No
Norway	ca. 50%	Yes
Portugal	NI	No
Spain	NI	No
Sweden	88.7%	Yes
United Kingdom	ca. 80% <sup>2</sup>	No

<sup>1</sup> Only for Eider river

<sup>2</sup> 10% in direct discharge

<sup>3</sup> e.g. Farkas and Skarbøvik (2021)

302  
303  
304

## 305 2.6 Ocean model experiments

306 To assess the effect of using bias corrected river discharge on simulated salinity in the German  
 307 Bight, we used version 3.6 of the Nucleus for European Modelling of the Ocean (NEMO;  
 308 Madec et al., 2017) and adopted a domain setup used by Ho-Hagemann et al. (2020). This  
 309 domain covers the region of the north-west European shelf, the North Sea and the Baltic Sea  
 310 between 19.89 E to 30.16 E and 40.07 N to 65.93 N with a resolution of two nautical miles (ca  
 311 3.6 km). We used the atmospheric forcing from ERA5 and the ocean boundary forcing from  
 312 the ECMWF Ocean Reanalysis System 5 (ORAS5; Zuo et al., 2019) to conduct two simulations  
 313 from 2010 to 2018. Initial conditions were taken from a 20-years spin-up simulation driven by  
 314 ERA5 data, so that the deeper ocean layers could adapt to the present-day climate (S. Grayek,  
 315 pers. comm., 2023). Note that for the evaluation of results, we neglected the year 2010 to have  
 316 an additional spin-up where NEMO could adapt to the specific transient conditions within each  
 317 of the two experiments. For the German Bight, this spin-up of one year is sufficient as the  
 318 residence time of water may comprise only up to four months (Becker et al., 1999). In the two  
 319 experiments, the daily riverine inflow into the ocean was taken from the uncorrected and bias  
 320 corrected discharges of HDW, which were converted to the NEMO grid using a procedure of  
 321 Nguyen et al. (2024). For each HD model river mouth box, we associated the nearest coastal  
 322 ocean box on the NEMO grid if such a box was found within a search radius of 200 km. Such  
 323 a large radius is necessary because the NEMO coastline is very smooth, so many estuaries and  
 324 bays in the HD model grid are not resolved by NEMO. If no ocean box was found, the  
 325 corresponding HD model box was not linked. Consequently, the simulated discharge data at the  
 326 river mouths were placed as freshwater inflow into the corresponding NEMO grid boxes.

## 327 2.7 Evaluation metrics

328 The evaluation of the simulated discharge was performed for the grid boxes corresponding to  
 329 the discharge station locations within the river network. For the evaluation at these station  
 330 locations, we used the mean bias, the Pearson correlation coefficient and the Kling-Gupta  
 331 efficiency (KGE; Gupta et al., 2009; Kling et al., 2012). All metrics were calculated with



332 simulated and observed daily discharge time series for the period considered, using only those  
333 days for which observed data are available. The KGE is a quality metric combining bias,  
334 correlation and coefficient of variation. If a simulated discharge time series has a KGE > -0.41,  
335 then it is a better representation of the observations than the use of the observed long-term mean  
336 discharge (Knoben et al., 2019). Note that many ocean model applications still use the latter  
337 method.

338 For the evaluation of simulated salinity in the NEMO experiments, we used daily values and  
339 considered

- 340 • the mean bias
- 341 • the correlation of simulated and observed time series expressed by the Pearson  
342 correlation coefficient
- 343 • the variability ratio defined by the ratio of the simulated and observed coefficients of  
344 variation
- 345 • the normalized root-mean-square-error (RSME)
- 346 • the centered RSME.

347 The first four metrics are described, e.g., in Hagemann et al. (2020), while the centered RSME  
348 is described, e.g., in Taylor (2001).

### 349 **3 Evaluation of the bias correction**

350 Below, various metrics have been calculated at the station locations and at the river mouths.  
351 However, these measures have been assigned to the respective catchment areas for the purpose  
352 of graphical presentation.

#### 353 **3.1 Evaluation of simulated discharge**

354 The distribution of bias and KGE for HDG and HDW during 1979-2014 (Figure 5) is quite  
355 similar to the pattern shown by Hagemann and Stacke (2022) for the ERA5-based discharge.  
356 For both simulations, the general discharge behaviour is well captured (KGE > 0.4) for many  
357 European rivers, especially in Northern Iberia, Western and Central Europe, and over Northern  
358 Russia (Figure 5, lower row). As expected (cf. Hagemann et al., 2020), larger deviations of the  
359 simulated from observed discharges occur for rivers that are heavily influenced by human  
360 activities such as water abstraction, e.g. for irrigation, and regulation, e.g. by dams. This is the  
361 case for many Scandinavian and Turkish rivers as well as the Volga and Don.

362 In general, the HDW discharges are slightly drier than the HDG discharges, as indicated by  
363 larger dry biases in Northern Europe and smaller wet biases in Central Europe. Despite the  
364 differences in bias distribution, the KGEs of HDW are similar to or slightly better than those of  
365 HDG. Compared to the ERA5-based discharge of Hagemann and Stacke (2022), HDW tends  
366 to have smaller discharge biases and better KGEs. This is an expected behaviour caused by the  
367 application of a bias correction methodology to the ERA5 data in the generation of the WFDE5  
368 data (cf. Sect. 2.1). An exception to this general improvement occurs over Northern Europe,  
369 where the dry bias of HDW tends to be slightly larger and the KGEs lower. Note that Hagemann  
370 and Stacke (2022) attributed the dry bias over Northern Europe to an overestimation of the  
371 evapotranspiration simulated by HydroPy.

372 We also note the large-scale patterns of positive and negative discharge biases (Figure 5).  
373 Abrupt changes in bias behaviour along the same coastline are rare. Most of the few cases can  
374 be attributed to large human water abstractions from the river, i.e. especially for the Ebro River

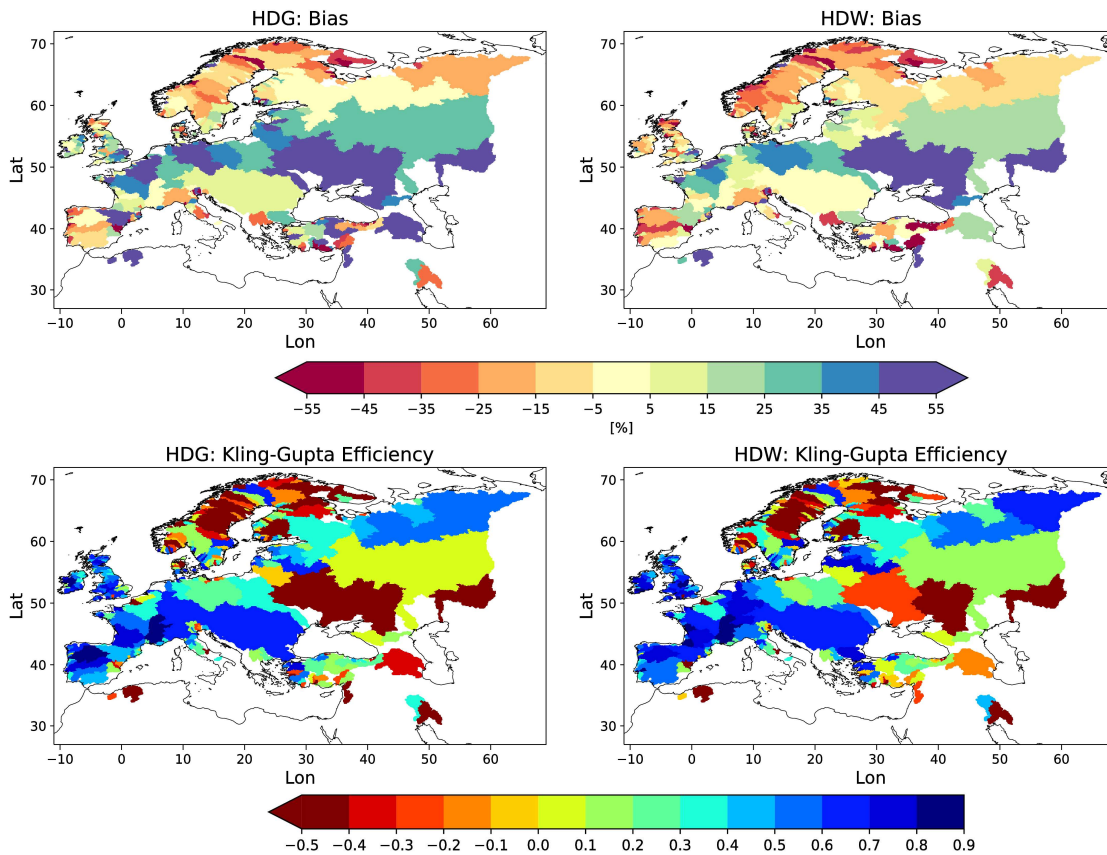
375 (see also Section 3.3) and in Turkey, which are not considered by the model. This supports our  
 376 assumption about the spatial transferability of the three-quantile bias correction factors. The  
 377 bias patterns are related to biases in the atmospheric forcing dataset or biases introduced by the  
 378 HydroPy model.

379

380

381

382



383 **Figure 5.** Mean discharge bias [%] (upper row) and KGE (lower row) for HDG (left) and  
 384 HDW (right) during 1979-2014.

385 In order to analyse how much the bias correction affects the daily sequence of river runoff  
 386 at the station locations, we calculated the correlation between the simulated discharges and the  
 387 observations. Supplementary Figure S1 shows that the correlation patterns of HDW and HDW-  
 388 BC with observed discharges are quite similar. For rivers where differences can be identified,  
 389 the correlation mostly increases for HDW-BC. The correlation between HDW and HDW-BC  
 390 is generally higher than 0.95, and only a very few rivers show correlations lower than 0.9. These  
 391 rivers are usually rivers that are heavily influenced by human activities, such as the Volga and  
 392 the Luleaelven.

### 393 **3.2 Added value of the three-quantile bias correction**

394 In this section, we consider the effect of the bias correction at the station locations and  
 395 investigate whether the three-quantile bias correction adds value compared to using only the  
 396 mean bias correction. For this purpose, we use HDW and the period 1979-2014.

397 Both bias correction methods reduce the mean discharge bias to zero or close to zero in the  
 398 case of the three-quantile bias correction due to the smoothing around the percentile range  
 399 thresholds (see Table 3 for selected rivers). When the mean bias correction is applied, the KGEs

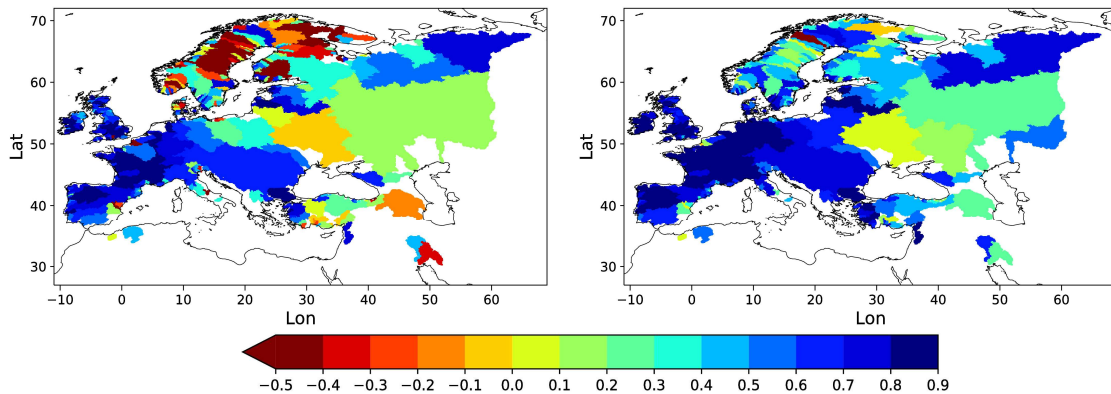
400 (Figure 6 – left panel) are noticeably improved over Western and Central Europe. However,  
 401 with a few exceptions, the KGE pattern over Northern Europe and other areas remains largely  
 402 unchanged. This indicates that a correction of the long-term bias in the annual mean discharge  
 403 over these areas is not sufficient. Only with the three-quantile bias correction does the KGE  
 404 (Figure 6 – right panel, Table 3 for selected rivers) improve considerably over these areas, with  
 405 the largest improvements occurring over Scandinavia. The three-quantile bias correction also  
 406 leads to some further improvements over Western and Central Europe, where the bias corrected  
 407 discharge with the mean bias correction already shows relatively high KGE values, e.g. for the  
 408 rivers Elbe, Rhine and Weser.

409 **Table 3.** Mean bias and KGE of simulated (HDW) and bias corrected discharge during  
 410 1979-2014 for selected rivers, where the three-quantile bias correction led to a  
 411 noticeable KGE improvement in comparison to the mean bias correction.

River	HDW		Mean Bias corr.		3-quantile Bias corr.	
	Bias	KGE	Bias	KGE	Bias	KGE
Dalaelven	-32.02 %	-0.32	0 %	-0.28	0.01 %	0.48
Elbe	36.44 %	0.46	0 %	0.60	-0.06 %	0.85
Indalsaelven	-19.32 %	-0.79	0 %	-0.78	-0.02 %	0.38
Odra	41.30 %	0.14	0 %	0.25	0.01 %	0.75
Rhine	14.60 %	0.74	0 %	0.78	-0.02 %	0.85
Weser	33.15 %	0.55	0 %	0.70	-0.01 %	0.90

412

413

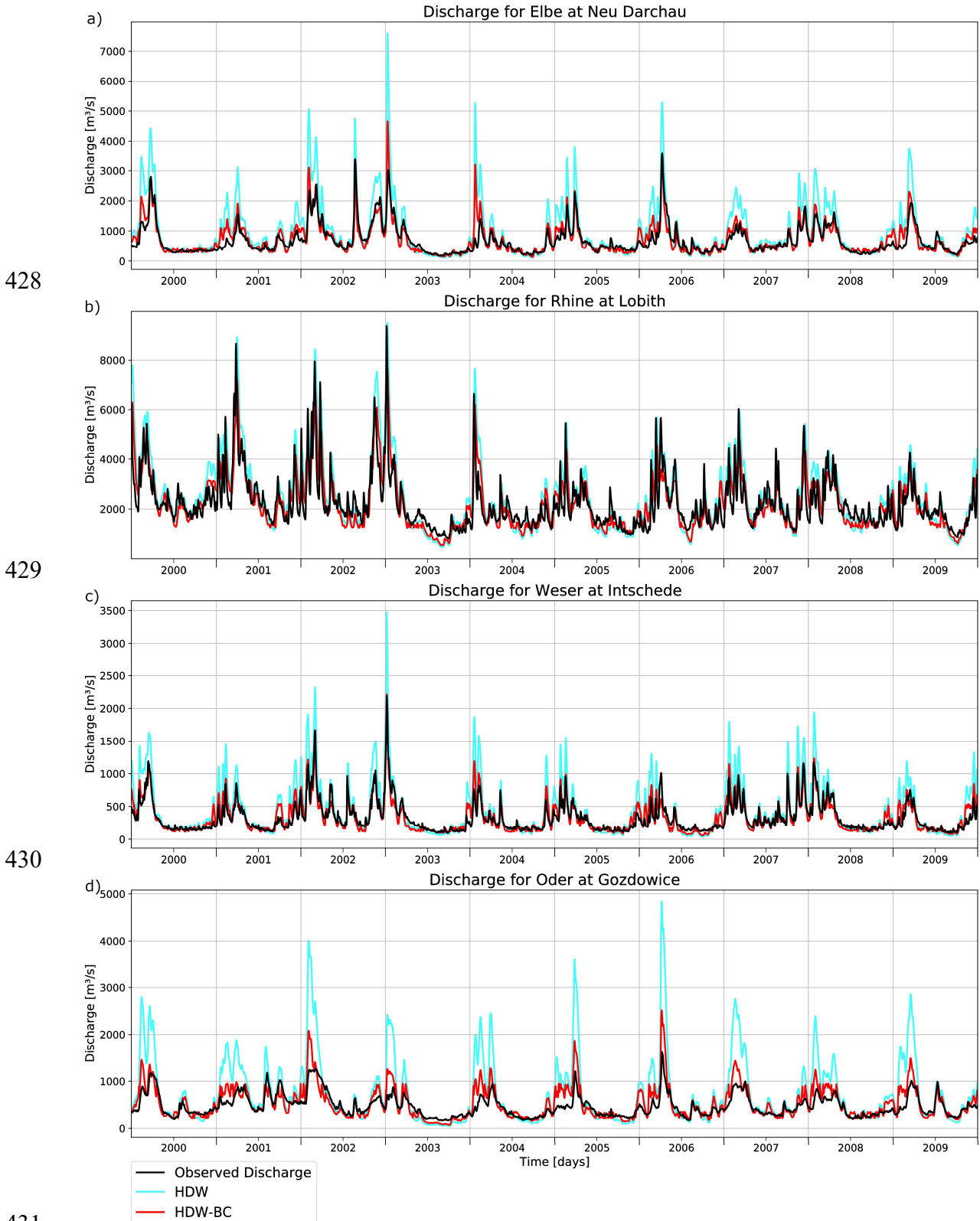


414

415

416 **Figure 6.** KGE for bias corrected HDW discharges using the mean bias correction (left)  
 417 and the three-quantile bias correction (right) during 1979-2014.

418 To visualise the effect of the three-quantile bias correction on the simulated daily discharges,  
 419 we consider the corresponding discharge curves for the period 2000-2009 for selected large  
 420 rivers. The respective biases and KGE are shown in Table 3 for the period 1979-2014. For the  
 421 rivers, Elbe, Weser and Oder, the peak discharges are generally overestimated, while the low  
 422 flows are close to the observed values (Figure 7a,c,d). The correction of the high percentiles  
 423 leads to a considerable improvement in the representation of the peak discharges, while the  
 424 change in the low flows is rather small. The discharge of the Rhine (Figure 7b) is well  
 425 represented by HDW. However, the small downward correction of the peak discharges and the  
 426 slight increase in the low flows still lead to an improved discharge curve, which is also indicated  
 427 by the increased KGE (Table 3).

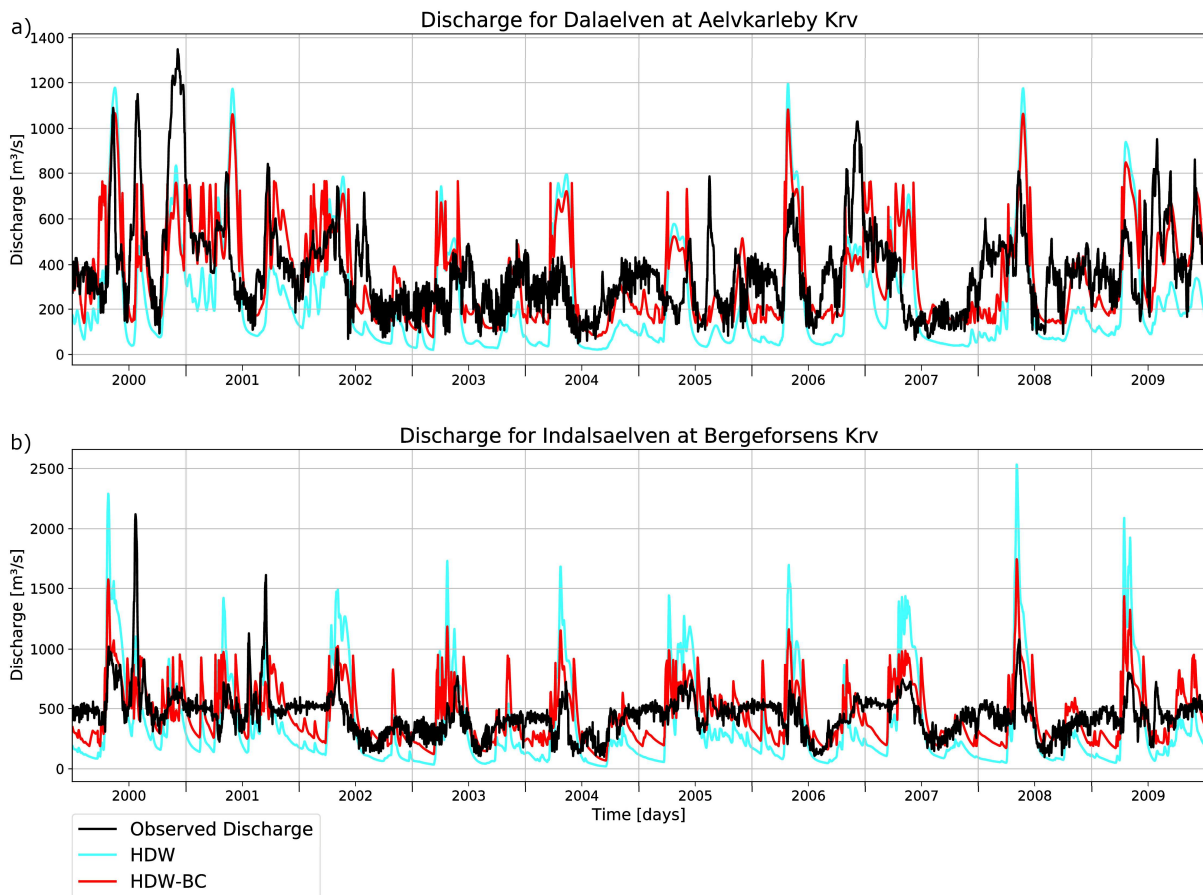


**Figure 7.** Observed and simulated daily discharges for the rivers a) Elbe, b) Rhine, c) Weser and d) Odra during 2000-2009.

As mentioned above, the greatest improvements from the three-quantile bias correction compared to the application of the mean bias correction occur over Scandinavia. Here many

436 rivers are highly regulated. For this reason, the discharge curves of the Daleaelven and  
 437 Indalsaelven rivers are examined in more detail in Figure 8. The observed discharges clearly  
 438 show the effect of the human regulation, where regulation leads to the elimination of peak  
 439 discharges, while maintaining certain flows during low flow periods. Figure 8 shows that, on  
 440 the one hand, peak discharges are often suppressed or reduced, especially during the spring,  
 441 and that, on the other hand, low-flow periods are either almost absent (especially for the  
 442 Indalsaelven) or show a rather noisy, unnatural daily variability. Here, the bias correction  
 443 partially mimics these regulation effects by reducing the peak discharges and increasing the low  
 444 flows.

445



446

447 **Figure 8.** Observed and simulated daily discharges for the rivers a) Daleaelven and b)  
 448 Indalsaelven during 2000-2009.

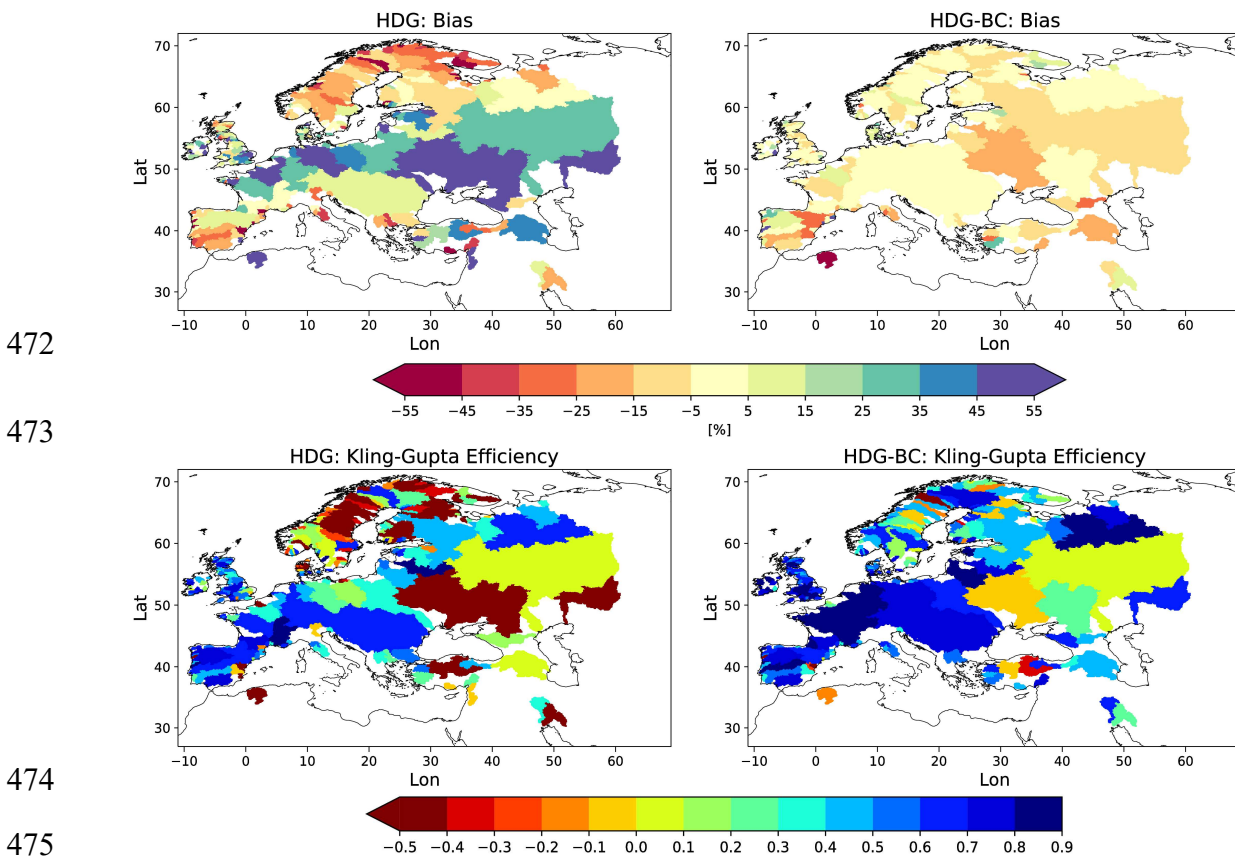
### 449 3.3 Application of the bias correction for a different time period

450 To consider the effect of the bias correction for the applications over different time periods, we  
 451 derived bias correction factors for HDG during 1979-2014 and applied the factors for the period  
 452 1950-1978.

453 For HDG, the distributions of bias and KGE are quite similar between the two periods 1950-  
 454 1978 (Figure 9 – left column) and 1979-2014 (Figure 5 – left column). Consequently, the bias  
 455 correction leads to similar improvements in the KGE (Figure 9) as for the most recent period  
 456 (not shown). The bias also becomes small for most of the rivers. Noticeable exceptions are the  
 457 Dnjepr, Volga and some rivers in Southern Europe. This may be related to differences in the



458 anthropogenic influence on the discharge between the two periods, as is the case for the river  
 459 Ebro. Here, the large wet bias (51.65 %) in the more recent period is contrasted with a small  
 460 wet bias (12.05%) in the earlier period (Figure 10). Since large anthropogenic water  
 461 abstractions occur in the Ebro River (Merchán et al., 2013), this seems to be related to the  
 462 different irrigation activities in the two periods, which are much more pronounced in the more  
 463 recent years. The latter can be seen by looking at the observed daily discharges between 1960-  
 464 1969 and 2000-2009 (Figure 10). In the earlier period, the Ebro discharge still shows some  
 465 variations according to the sequence of weather events in the dry season. However, in the later  
 466 period, the observed discharge includes only very small variations during the dry season,  
 467 indicating more intense human water abstraction than in the earlier period. Consequently, the  
 468 bias correction based on the recent wet bias leads to a dry bias (-25.78 %) in the corrected Ebro  
 469 discharge in the earlier period. However, the KGE decreases only slightly from 0.68 to 0.63, so  
 470 that the deterioration of the mean bias seems to be largely compensated by the correction of the  
 471 different percentile ranges.



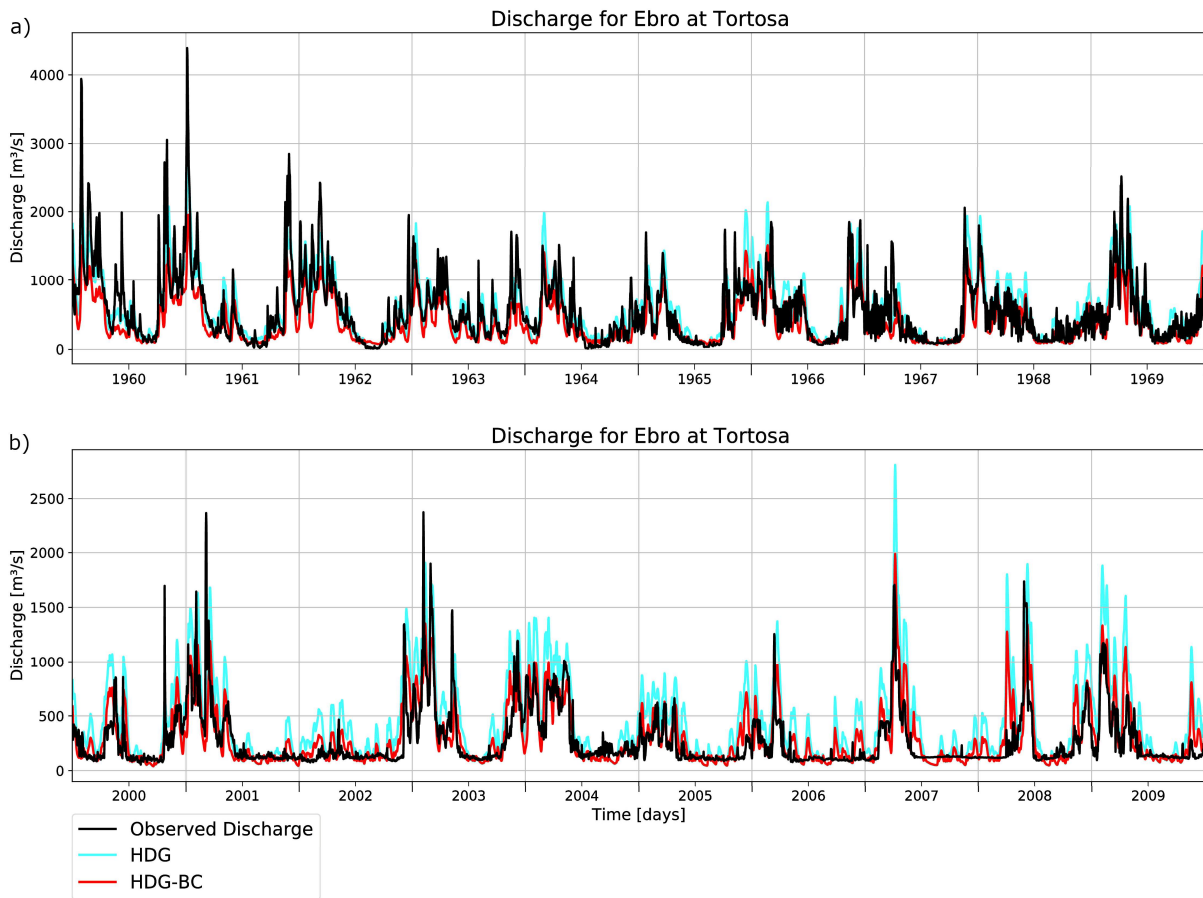
476 **Figure 9.** Mean discharge bias [%] (upper row) and KGE (lower row) for HDG (left) and  
 477 HDG-BC data (right) during 1950-1978.

### 478 3.4 Effect of the bias correction on contemporary trends

479 As mentioned in Sect. 2.4, our three-quantile bias correction is similar to a very coarse quantile  
 480 mapping, and quantile mapping has been flagged as potentially not suitable for climate  
 481 simulations as it has been shown to modify trends (e.g. references in Cannon et al., 2015).  
 482 However, Maraun et al. (2017) pointed out that a debate has arisen about whether trend  
 483 modification by variance-adjusting bias correction methods actually improves or degrades the  
 484 raw climate change signal. They further argued that purely statistical arguments cannot resolve

485 this issue, which requires process understanding. With respect to runoff, the latter needs to take  
486 into account spatial and temporal characteristics of rivers and events, which is beyond the scope  
487 of the present large-scale study.

488



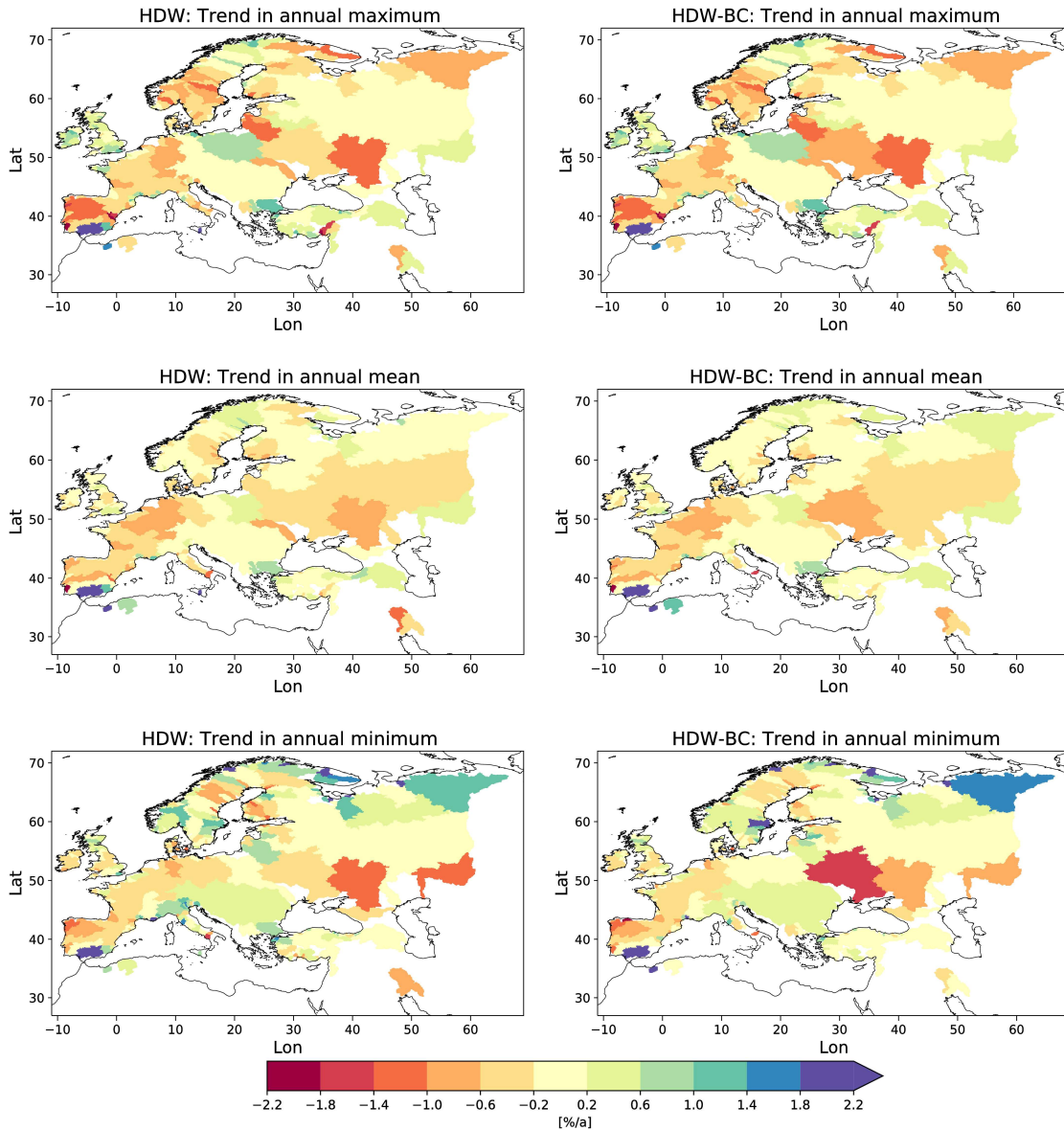
489

490 **Figure 10.** Observed and simulated daily discharge based on HDG for the Ebro river  
491 during a) 1960-1969 and b) 2000-2009.

492 To investigate the effect of the bias correction on contemporary trends, we calculated trends in  
493 the annual maximum, mean and minimum discharge for the period 1979-2014 and compared  
494 the results for HDW and HDW-BC (Figure 11). The trend patterns are generally within the  
495 range spanned by the two datasets considered in Hagemann and Stacke (2022). For the annual  
496 maximum and mean discharge, the trend patterns are only slightly changed by the bias  
497 correction. For the annual minimum discharge, the trend pattern is quite similar in HDW and  
498 HDW-BC. However, there are a few more rivers where the magnitude of the trend is affected  
499 by the bias correction. This is particularly the case over Scandinavia where many rivers are  
500 regulated, so that that the correction of the low percentile range is often strong to account for  
501 the effect of regulation on low flows (cf. Sect. 3.2).



502  
503



504  
505

506

507  
508  
509

**Figure 11.** Trends in annual maximum (1<sup>st</sup> row), mean (2<sup>nd</sup> row) and minimum (3<sup>rd</sup> row) discharge [%/a] for HDW (left column) and HDW-BC (right column) from 1979-2014.

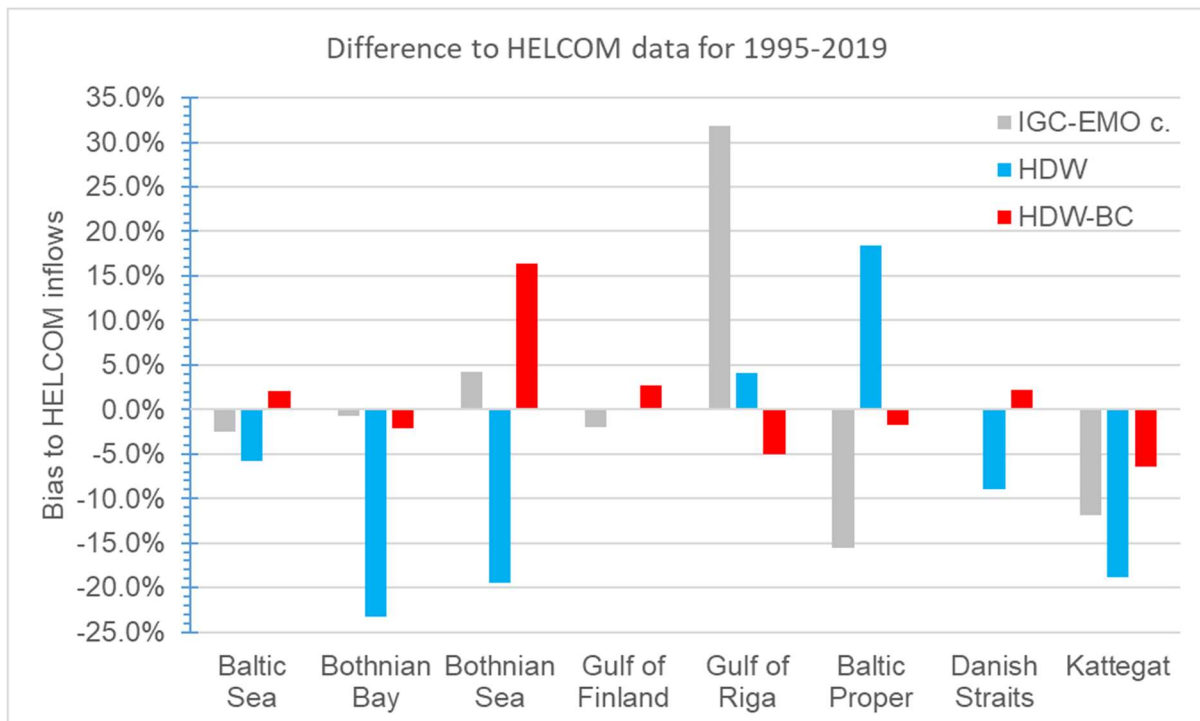
## 510 4 Evaluation of the inflow into sea basins

511 To evaluate the simulated and bias corrected discharges at the river mouths, we considered the  
512 integrated inflow into different sea basins. First, we evaluated the discharges into the Baltic Sea  
513 using HELCOM and IGC-EMO data in Section 4.1. We then compared the discharges to the  
514 North Sea and the Northeast Atlantic with OSPAR and up-scaled (see Section 2.5) IGC-EMO  
515 data in Section 4.2.

### 516 4.1 Baltic Sea

517 In order to achieve a maximum overlap of the simulated discharge time series data with the  
518 HELCOM data (cf. Section 2.5), we considered 1995-2019 as the evaluation period for the  
519 Baltic Sea and its seven sub-basins (Figure 4 – upper panel). For the whole Baltic Sea and most  
520 of its sub-basins, the bias correction improves the basin inflows if compared to the HELCOM  
521 estimates (Table 4, Figure 12). Only for the Gulf of Finland and the Gulf of Riga, the bias

522 correction leads to a slightly larger bias while the biases of HDW in these basins are relatively  
 523 small. When the simulated inflows are compared with the IGC-EMO estimates, similar results  
 524 are obtained, except for the Gulf of Riga. Here, the IGC-EMO estimates are about 32% larger  
 525 than the HELCOM estimates, indicating a larger uncertainty in at least one of these two  
 526 estimates. For the Gulf of Riga basin, the major part of the inflow is contributed by the Daugava  
 527 river. In IGC-EMO, the discharge from the Daugava is based on observed time series from  
 528 1970-1990. These data were extended to earlier and later periods, e.g. by climatological values  
 529 and trend preservation (Van Leeuwen and Hagemann, 2023). For 1970-1990, the mean IGC-  
 530 EMO discharge comprises 623 m<sup>3</sup>/s at the Daugava mouth, while this has increased by ca. 45%  
 531 in 1995-2019 (902 m<sup>3</sup>/s). However, this strong increase cannot be seen in the observed  
 532 discharge time series at the station Daugavpils that covers about three quarter of the Daugava  
 533 catchment. Here, the discharge increases only slightly from 1970-1999 (439 m<sup>3</sup>/s; 95%  
 534 temporal data coverage) to 1995-2019 (452 m<sup>3</sup>/s; 83% temporal data coverage). This indicates  
 535 a large overestimation of the IGC-EMO Daugava discharge during 1995-2019 and, hence, also  
 536 of the respective Gulf of Riga inflow.



537

538 **Figure 12.** Relative difference in basin inflows compared to HELCOM data for 1995-  
 539 2019. Note that no IGC-EMO estimate is provided for the Danish Straits as the  
 540 respective river catchment coverage in IGC-EMO is too small.

541 **Table 4.** Estimated and simulated inflows (unit: m<sup>3</sup>/s) into the Baltic Sea and its major  
 542 sub-basins during 1995-2019. Note that for the Danish Straits no IGC-EMO estimate is  
 543 provided as the respective catchment area coverage of the rivers in IGC-EMO is too low.

Sea basin	HELCOM	IGC-EMO c.	HDW	HDW-BC
Baltic Sea	15676	15286	14764	15995
Bothnian Bay	3444	3420	2642	3369
Bothnian Sea	2913	3038	2347	3391
Gulf of Finland	3519	3448	3520	3612

Gulf of Riga	1071	1411	1114	1017
Baltic Proper	3436	2901	4070	3377
Danish Straits	217	0	198	222
Kattegat	1077	949	873	1008

544

## 545 4.2 North Sea and Northeast Atlantic

546 Due to the different treatment of unmonitored regions by the OSPAR countries (cf. Section  
547 2.5), and thus of the respective sea basin areas, we have not corrected the OSPAR inflows.  
548 Instead, we have also considered up-scaled IGC-EMO data as alternative estimates of basin  
549 inflow (as in Section 4.1). Table 5 shows simulated and estimated basin inflows for the  
550 considered OSPAR regions (cf. Figure 4 – lower panel). Note that IGC-EMO data for the  
551 Norwegian shares of the Barents Sea, Norwegian Sea and North Sea, and the North Spanish  
552 Atlantic are not included in the following comparisons due to their limited area coverage. When  
553 comparing the simulated sea basin inflows with the OSPAR and IGC-EMO data, we found that  
554 the bias correction improves the simulated inflows for most of the OSPAR regions (Figure 13).  
555 Exceptions are the values for the Celtic Sea and the Irish Sea. For the Celtic Sea, the bias  
556 corrected inflows are very close to the uncorrected inflows and the difference to the OSPAR  
557 data is rather small. For the Irish Sea, the bias corrected inflows are somewhat larger than the  
558 uncorrected ones, with both showing large differences (52.5% and 47.5%) to the OSPAR data.  
559 Here both inflows are closer to the IGC-EMO estimate, which exceeds the OSPAR estimate by  
560 about 40%.

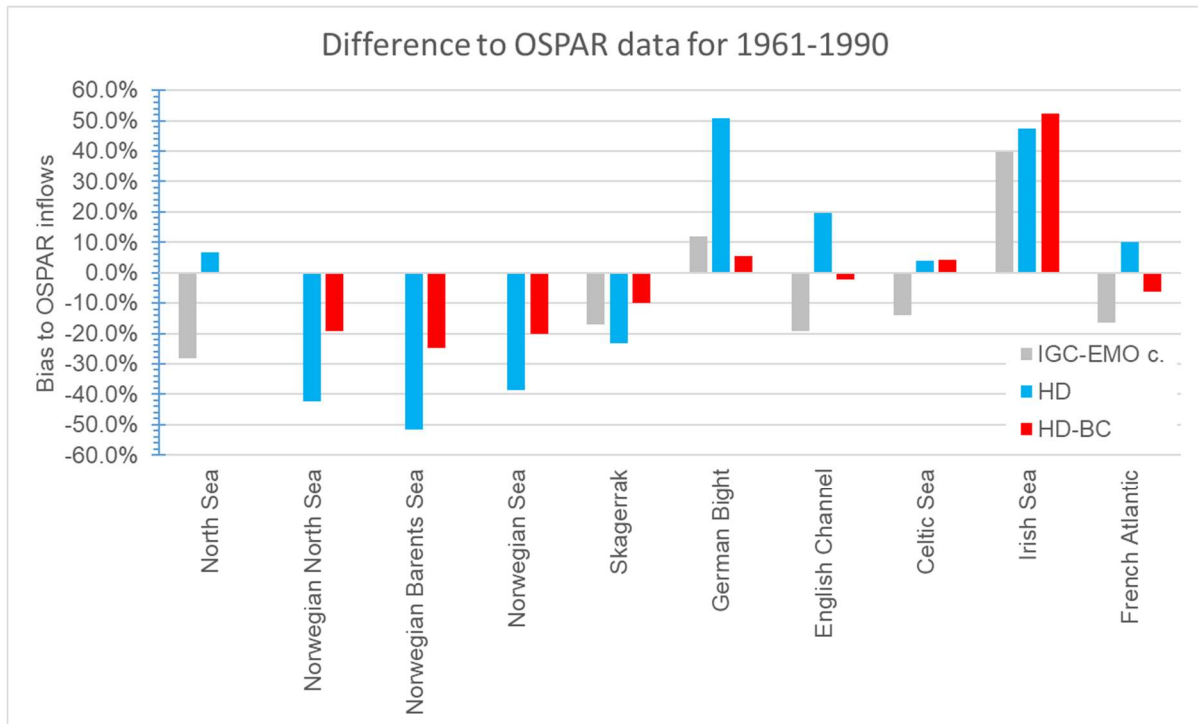
561 **Table 5.** Estimated and simulated inflows (unit: m<sup>3</sup>/s) into major sub-basins of the North  
562 Sea and the Northwest Atlantic during 1961-1990. Note that the North Sea does not  
563 comprise Skagerrak and the English Channel. Up-scaled IGC-EMO basin estimates  
564 for which the fractional catchment coverage (see Table 1) of IGC-EMO rivers is less  
565 than 75% are considered as highly uncertain and are therefore only given in brackets  
566 (cf. Sect. 2.5). The same applies to the OSPAR inflow into the Northern Spanish  
567 Atlantic.

Sea basin	OSPAR	IGC-EMO c.	HD	HD-BC
North Sea	9190	6600	9798	9164
Norwegian North Sea	3534	(1499)	2038	2856
Norwegian Barents Sea	2294	-	1106	1723
Norwegian Sea	3663	-	2242	2922
Skagerrak	2544	2113	1956	2292
German Bight	1344	1505	2025	1419
English Channel	1250	1011	1498	1222
Celtic Sea	976	839	1016	1016
Irish Sea	672	939	992	1025
French Atlantic	2862	2391	3147	2684
Northern Spanish Atlantic	(359)	(1655)	1104	1550

568

569 While the OSPAR values from Ireland include estimates for unmonitored areas, this is not  
570 the case for the United Kingdom (Table 2). Farkas and Skarbøvik (2021) list the rivers

571 contributing to the OSPAR value (560 m<sup>3</sup>/s) from the United Kingdom part of the Irish Sea  
 572 catchment (35000 km<sup>2</sup>). Adding up the catchment areas of each river, obtained from various  
 573 online resources, gives a coverage of about 70%. In order to account for this under-  
 574 representation of the catchment area, an up-scaling can be performed, similar to the treatment  
 575 of the IGC-EMO data. This would give an estimate of about 803 m<sup>3</sup>/s for the Irish Sea inflow  
 576 from the United Kingdom and thus 915 m<sup>3</sup>/s for the whole Irish Sea. The respective IGC-EMO  
 577 inflow is close to this value (+2.6%) and the overestimation of inflows is less pronounced for  
 578 HD and bias corrected discharges with +8.4% and +12% respectively.



579

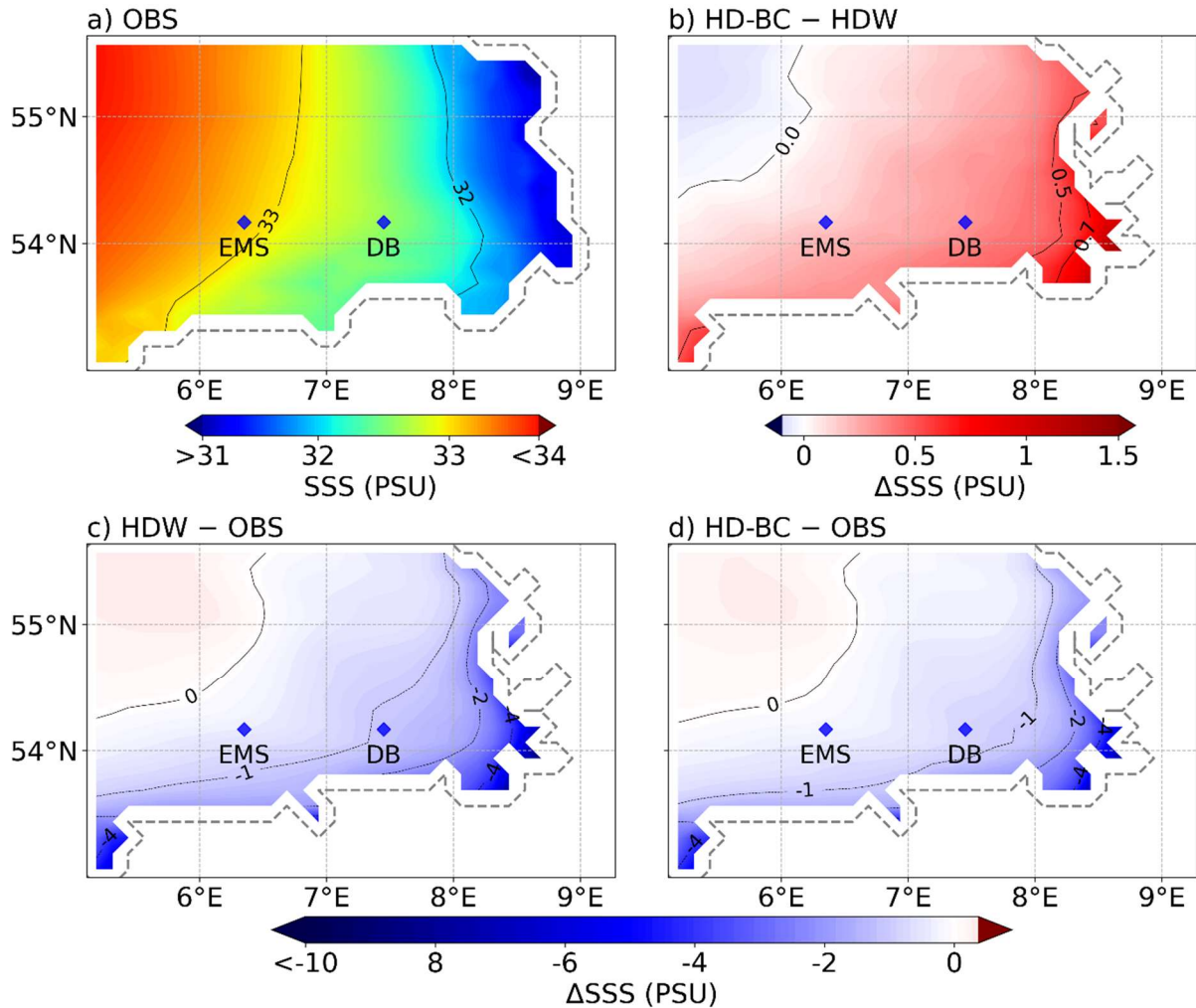
580 **Figure 13.** Relative difference in basin inflows compared to OSPAR data for 1961-1990.  
 581 IGC-EMO basin estimates for which the fractional catchment coverage (see Table 1)  
 582 is less than 75% are not shown.

### 583 4.3 Simulated salinity in the German Bight

584 Using the two experiments described in Sect. 2.6, we evaluated the simulated sea surface  
 585 salinity (SSS) with satellite-based analyses and in-situ observations for the period 2010 to 2018.  
 586 The SSS analyses were derived using a multivariate optimal interpolation algorithm that  
 587 combines sea surface salinity images from several satellite sources, such as the National  
 588 Aeronautics and Space Administration Soil Moisture Active Passive satellite and the European  
 589 Space Agency Soil Moisture Ocean Salinity satellite, with in-situ salinity measurements  
 590 (Droge et al., 2018). These SSS data are available with a spatial resolution of 0.125°.

591 Figure 14a shows the mean analysed SSS in the German Bight for the period 2010-2018,  
 592 with lower salinities near the west coast of Germany and higher salinities towards the west. The  
 593 NEMO simulation using the uncorrected discharges of HDW (Figure 14c) has too low SSS in  
 594 coastal areas, especially near the estuaries. This low bias is reduced using the bias corrected  
 595 discharges (Figure 14d), as the general effect of the bias correction in the German Bight leads  
 596 to reduced riverine inflows (cf. Figure 13) and hence increased SSS in coastal areas (Figure

597 14b). Similar improvements can also be seen in June 2013 when the Elbe flood is strongly  
 598 influences the SSS of the German Bight (Figure S2). Here, the increase in salinity due to the  
 599 bias corrected runoff (Figure S2b) is more pronounced than in the long-term mean (Figure 14b).  
 600 In addition, we found that use of the bias corrected river runoff also improves the SSS variability  
 601 expressed by its coefficient of variation, shown in Figure 15.

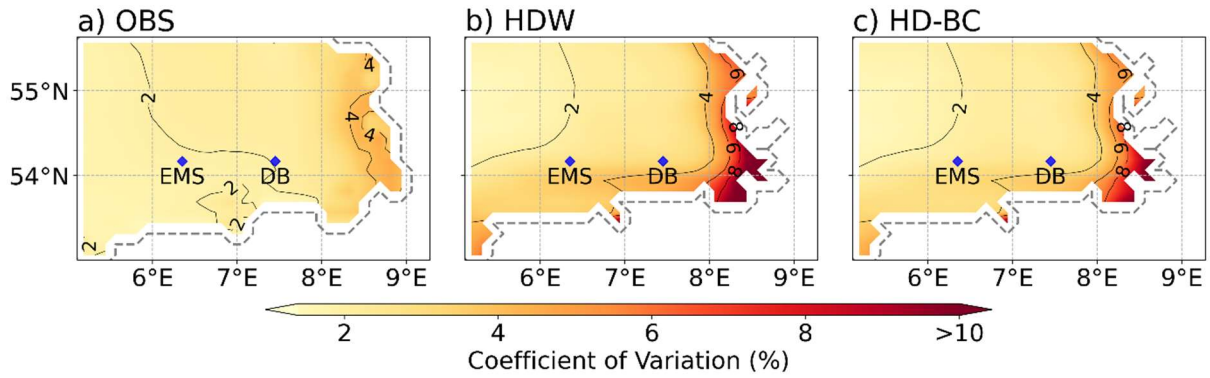


602

603 **Figure 14.** Mean analyzed SSS: a) Droghei et al. (2018) data (OBS) and various SSS  
 604 differences of the NEMO experiments in the German Bight for the period from 2010  
 605 to 2018. The SSS differences comprise b) HD-BC minus HDW, c) HDW minus OBS,  
 606 and d) HD-BC minus OBS.

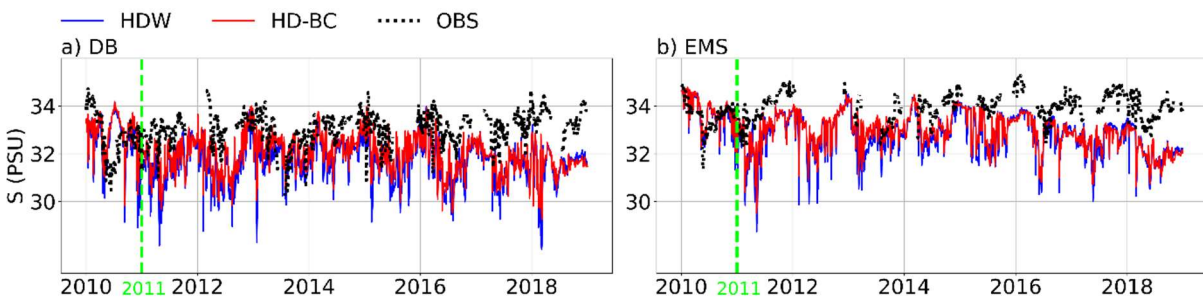
607





608

609 **Figure 15.** Coefficients of variation of SSS in the German Bight for the period from 2011-  
 610 2018: a) OBS, b) HDW and c) HD-BC



611

612 **Figure 16.** Observed (OBS) and simulated daily time series of salinity in 6 m depth for the  
 613 stations a) *Deutsche Bucht* (DB) and b) *EMS*. Unit: PSU. The blue and red solid lines  
 614 correspond to the HDW and HD-BC experiments, respectively. The green line  
 615 separates the spin-up period in 2010 from the evaluation period 2011-2018.

616 In addition, we had access to salinity measurements at two stations in the German Bight  
 617 operated by the German Federal Maritime and Hydrographic Agency as part of the Marine  
 618 Environmental Monitoring Network in the North and Baltic Seas. These two stations are  
 619 *Deutsche Bucht* (DB; located at 54.17°N, 7.45°E) and *EMS* (54.17°N, 6.35°E) and their  
 620 locations are shown in Figure 14. In general, the bias corrected discharges lead to improved  
 621 characteristics of the daily salinity at 6 m depth at the *Deutsche Bucht* and *EMS* stations (Figure  
 622 16, Table 6). Here the bias, normalized and centred RSME are decreased, and the coefficient of  
 623 variation is closer to the salinity observations for HDW-BC. This means that the bias correction  
 624 improves the mean and the variability of the simulated salinity at these stations. However, the  
 625 correlation with the observed salinity measurements is somewhat reduced. Note that temporal  
 626 SSS variations are strongly influenced by local currents, vertical mixing and wind-wave-surface  
 627 interactions. Therefore, signals from an improved river runoff can easily be obscured by the  
 628 noise from these processes, which can also differ at the point scale of the station measurements  
 629 and at the grid scale of the respective NEMO grid box. This is reflected in the relatively low  
 630 correlation values. Furthermore, this can be seen when the gridded SSS data of Droghei et al.  
 631 (2018) are used as a reference for the metrics at the station locations (Supplementary Table S1).  
 632 Here, all metrics improve with HDW-BC, even the correlation. However, the correlation is  
 633 lower than with the station observations, which is also the case for the correlation of the gridded  
 634 SSS data itself with the station observations (*Deutsche Bucht*: 0.15; *EMS*: 0.18). Considering  
 635 only the year 2013, when the influence of the Elbe flood on the salinity at the *Deutsche Bucht*  
 636 station is more pronounced (Nguyen et al., 2024), the correlation also improves when using

637 HDW-BC for both references (Table 6 and S1). It seems that in NEMO the positive effect of  
 638 using bias corrected discharges is limited to near-surface salinities, as there is no noticeable  
 639 effect at 30 m depth (not shown). This is consistent with the fact that the *Deutsche Bucht* and  
 640 *EMS* stations are located in an area where the salinity is temporarily stratified, depending on  
 641 the meteorological conditions and the intensity of river runoff (Klein and Frohse, 2008).

642 In summary, the results of the NEMO experiments indicate the beneficial effect of using bias  
 643 corrected discharges on the simulated SSS in coastal areas. However, although the low SSS  
 644 biases are reduced by using the bias corrected discharges, the simulated SSS is still  
 645 underestimated in coastal areas, especially close to the estuaries of large rivers (Figure 14d).  
 646 This may be attributed to the rather smooth coastline of the NEMO ocean grid. Here, most parts  
 647 of the large estuaries of the rivers Elbe, Ems and Weser are not included. In reality, a major part  
 648 of the mixing of the riverine freshwater inflow and the saline North Sea happens within these  
 649 estuaries. In the NEMO model setup, the freshwater inflow is introduced at the respective river  
 650 mouth points of the smooth NEMO coastline where it starts to mix with the saline North Sea  
 651 water. Consequently, the simulated water at and near those points is much fresher than in reality,  
 652 which leads to the low SSS bias. Note that on the one hand such a smooth coastline is necessary  
 653 in NEMO to avoid numerical instabilities. On the other hand, the spatial resolution of the  
 654 NEMO grid is not high enough to adequately resolve parts of the longer estuaries.

655 **Table 6.** Various metrics (see Sect. 2.7) of the simulated salinity time series in 6 m  
 656 depth compared with the observations at the stations *Deutsche Bucht* and *EMS* for  
 657 2011-2018 and at *Deutsche Bucht* for 2013.

Metric	2011-2018				2013	
	<i>Deutsche Bucht</i>		<i>EMS</i>		<i>Deutsche Bucht</i>	
	HDW	HDW-BC	HDW	HDW-BC	HDW	HDW-BC
Bias [%]	-4.5	-3.7	-4	-3.6	-3.2	-1.8
Variability ratio [%]	142.7	125	151.5	136.1	82.9	74.2
Normalized RMSE [%]	40.1	34.3	51.3	47.6	36.2	27.1
Centered RMSE	0.94	0.89	0.73	0.72	0.97	0.89
Correlation	0.24	0.21	0.48	0.39	0.20	0.28

658

## 659 5 Summary and Conclusions

660 In the present study, we have introduced a methodology for the bias correction of European  
 661 river runoff to provide corrected riverine inflows as forcing for ocean models in offline and  
 662 coupled system model simulations. The central part of this methodology is a three-quantile bias  
 663 correction, which can correct different biases for low, medium and high discharges. The bias  
 664 correction parameters are derived in two steps. First, different correction factors for low,  
 665 medium and high flows are derived for each river considered (cf. Sect. 2.5) at the location of  
 666 the most downstream station for which daily discharge measurements were available. These  
 667 factors were then transferred to the respective river mouth on the HD model grid and to adjacent  
 668 coastal inflow points in its vicinity.

669 The evaluation of the bias corrected discharge at the station location showed that the bias  
 670 correction greatly improved the simulated discharges. For the evaluation of the bias corrected  
 671 discharge at the downstream station locations, we considered the mean bias and the KGE, which  
 672 is a quality metric combining bias, correlation and coefficient of variation. Considering the



673 same period as used to derive the bias correction factors, the mean bias is trivially close to zero.  
674 However, the bias is also substantially reduced for most rivers if a different period is considered.  
675 Irrespective of the period, the KGE pattern generally improves for the bias corrected discharges  
676 and shows high values for many rivers. Exceptions are those rivers with a very strong  
677 anthropogenic distortion of the natural flow, e.g. by many dams or large water withdrawals.  
678 Here, despite of some improvements, the KGE values are still rather low, such as for the rivers  
679 Dnjepr, Volga, Luleälven and a few Turkish rivers flowing into the Black Sea. The KGE also  
680 shows the beneficial effect of the three-quantile bias correction, as correcting only the long-  
681 term mean annual discharge bias is not sufficient in many areas, especially in northern Europe.  
682 We found that the three-quantile bias correction often improves the KGE in regulated rivers, so  
683 that it appears to mimic the effect of regulation, where regulation leads to the elimination of  
684 peak flows while maintaining certain flow levels during low flow periods.

685 The evaluation of riverine inflows to the sea at river mouths with observed daily discharge  
686 is rarely possible as there are usually no river gauges available. Even if there is a gauge at the  
687 mouth of a river, the measurements are often affected by tidal influences from the coast, so that  
688 the measured amounts may not represent the actual river discharge. For obvious reasons, it is  
689 also difficult to compare simulated inflows with observed discharges for unmonitored rivers.  
690 Therefore, we compared the simulated and bias corrected discharges with long-term mean  
691 inflow estimates into different sea basins from HELCOM, OSPAR and IGC-EMO. For most of  
692 the basins considered, the bias correction improves the simulated inflows. This indicates a  
693 reasonable performance of the approach to transfer the bias correction factors obtained at the  
694 downstream stations to the respective river mouths and adjacent coastal areas. The improved  
695 inflows to the sea basins, together with the fact that the discharge bias behaviour tends not to  
696 vary abruptly along the same coastline, underpin the validity of our transferability approach.  
697 Exceptions are the Gulf of Finland, the Gulf of Riga, the Celtic Sea and the Irish Sea. For the  
698 Gulf of Finland and the Celtic Sea, the deviations of the uncorrected and bias corrected inflows  
699 from the inflow estimates are rather small. For the Gulf of Riga, the deviations of the  
700 uncorrected and bias corrected inflows from the HELCOM estimates are also small, but they  
701 significantly underestimate the IGC-EMO estimates. However, this could be due to a large  
702 overestimation of the Daugava discharge during the period 1995-2019 in the IGC-EMO data  
703 and thus also of the corresponding Gulf of Riga inflow. For the Irish Sea, IGC-EMO seems to  
704 be closer to reality as the OSPAR inflow does not cover the unmonitored rivers in the British  
705 part of the catchment.

706 A caveat applies for rivers where the human influence on river flow has changed  
707 significantly over time. Applying bias correction factors derived for 1979-2014 to earlier  
708 periods may lead to errors for regulated rivers in years before these regulatory measures were  
709 implemented. This is the case for the Ebro, where irrigation activities have largely intensified  
710 during the period 1979-2014 compared to earlier periods (see Sect. 3.3). A detailed analysis of  
711 the rivers and periods concerned is beyond the scope of this study. However, at least for the  
712 period 1950-1978, the KGE distribution does not seem to be significantly affected, as there is  
713 no noticeable deterioration.

714 We have shown that our bias correction method works well for Europe at the station  
715 locations as well as for the riverine inflow into northern and western European sea basins. Using  
716 two NEMO simulations in the German Bight, we have also shown that the use of the bias  
717 corrected discharges as forcing leads to an improved simulation of sea surface salinity in coastal  
718 areas especially regarding the mean salinity and its variability. However, for the potential  
719 transfer of the bias correction methodology to other regions, it has to be pointed out that the

720 application of the three-quantile bias correction over a region only makes sense if a large part  
721 of the catchment area is covered by available daily discharge measurements. As the three-  
722 quantile bias correction is based on biases in three percentile ranges of daily flows, it is also  
723 suitable for the use in climate change applications. Here the bias correction factors can be  
724 derived from a historical discharge simulation and then applied to future projections or past  
725 reconstructions. In addition, the bias correction can also be applied in regional coupled system  
726 model simulations, where the bias correction factors can be derived from an initial simulation  
727 and then applied during the run-time of the actual coupled simulation. This capability has been  
728 implemented in the HD model v5.2.2 (Hagemann et al., 2023) and is currently being applied in  
729 the coupled system model GCOAST-AHOI (Ho-Hagemann et al., 2020). Finally, we note that  
730 the bias corrected discharges are available from the World Data Centre for Climate and are  
731 already used within the CoastalFutures project (<https://www.coastalfutures.de>).

### 732 **Data Availability Statement**

733 Many of the observed daily discharge data used can be obtained from the Global Runoff Data  
734 Centre ([https://grdc.bafg.de/GRDC/EN/02\\_srvcs/21\\_tmsrs/riverdischarge\\_node.html](https://grdc.bafg.de/GRDC/EN/02_srvcs/21_tmsrs/riverdischarge_node.html)). Other  
735 data have been retrieved from public websites associated with the sources referred to in Sect.  
736 2.5. GSWP3 data were retrieved from the ISIMIP data portal (<https://data.isimip.org>) and  
737 WFDE5 data were retrieved from the Copernicus Climate Data Store  
738 (<https://cds.climate.copernicus.eu>). OSPAR data were taken from an OSPAR report (Farkas  
739 and Skarbøvik, 2021) or its associated data available on the OSPAR webpage  
740 (<https://odims.ospar.org/en/search/?dataset=rid-data-reports>). This study has been conducted  
741 using E.U. Copernicus Marine Service Information data on SSS ([https://doi.org/10.48670/moi-  
742 00051](https://doi.org/10.48670/moi-00051)) and some French discharge measurements. The daily data of surface runoff and  
743 subsurface runoff as well as the simulated and bias corrected discharge data (Hagemann and  
744 Stacke, 2023) can be accessed via the World Data Centre for Climate at the German Climate  
745 Computing Center.

### 746 **Acknowledgments**

747 This study was conducted within the CoastalFutures project that was funded by the German  
748 Federal Ministry of Education and Research under grant number 03F0911E. TN was supported  
749 by the subproject ‘A6 - The earth system variability and predictability in changing climate’ of  
750 Germany’s Excellence Strategy EXC 2037 ‘CLICCS - Climate, Climatic Change, and Society’  
751 with project no. 390683824, funded by the Deutsche Forschungsgemeinschaft (German  
752 Research Foundation). We thank the German Climate Computing Center for providing the  
753 computing resources to perform the HD simulations. We acknowledge the Copernicus Climate  
754 Data Store and the ISIMIP project for making WFDE5 and GSWP3 datasets available. We are  
755 deeply indebted to all data providers. We are also grateful to Sonja van Leuwen (Royal  
756 Netherlands Institute for Sea Research) for providing us with the latest version of the IGC-  
757 EMO data. We are thankful to Sebastian Grayek (Helmholtz-Zentrum Hereon) for the  
758 discussion on his NEMO results using an initial version of the bias corrected discharges.  
759 Finally, we thank Tobias Stacke (Max Planck Institute for Meteorology) for conducting the  
760 HydroPy simulations published in Hagemann and Stacke (2023).

### 761 **Author Contributions**

762 SH developed and applied the three-quantile bias correction, conducted the discharge  
763 simulations and analysis of results, and wrote the manuscript. HH conducted the NEMO

764 simulations, helped with the analysis of results and revised the manuscript. TN evaluated the  
765 SSS data of the NEMO simulations, helped with the analysis of results and revised the  
766 manuscript.

## 767 **Conflict of Interest Statement**

768 The authors declare that the research was conducted in the absence of any commercial or  
769 financial relationships that could be construed as a potential conflict of interest.

## 770 **References**

771

- 772 Arora, V. K., Seiler, C., Wang, L. B., and Kou-Giesbrecht, S.: Towards an ensemble-based  
773 evaluation of land surface models in light of uncertain forcings and observations,  
774 *Biogeosci.*, 20, 1313-1355, <https://doi.org/10.5194/bg-20-1313-2023>, 2023.
- 775 Becker, G. A., Dick, S., and Dippner, J. W.: Hydrography of the German Bight, *Mar. Ecol.*  
776 *Prog. Ser.*, 91, 9-18, <https://doi.org/10.3354/meps091009>, 1992.
- 777 Becker, G. A., Giese, H., Isert, K., König, P., Langenberg, H., Pohlmann, T., and Schrum, C.:  
778 Mesoscale structures, fluxes and water mass variability in the German Bight as exemplified  
779 in the KUSTOS- experiments and numerical models, *Deutsche Hydrographische*  
780 *Zeitschrift*, 51, 155-179, <https://doi.org/10.1007/bf02764173>, 1999.
- 781 Borgvang, S. A., Skarbøvik, E., and Pengerud, A.: RID 2006 data report: Presentation and  
782 Assessment of the OSPAR Contracting Parties' RID 2006 Data., Norwegian Institute for  
783 Agricultural and Environmental Research, London, No. 376/2008, 373 pp., 2008.
- 784 Brown, J. D., and Seo, D. J.: A nonparametric postprocessor for bias correction of  
785 hydrometeorological and hydrologic ensemble forecasts, *J. Hydrometeorol.*, 11, 642-665,  
786 <https://doi.org/10.1175/2009jhm1188.1>, 2010.
- 787 Brown, J. D., and Seo, D. J.: Evaluation of a nonparametric post-processor for bias correction  
788 and uncertainty estimation of hydrologic predictions, *Hydrol. Process.*, 27, 83-105,  
789 <https://doi.org/10.1002/hyp.9263>, 2012.
- 790 Budhathoki, A., Tanaka, T., and Tachikawa, Y.: Correcting streamflow bias considering its  
791 spatial structure for impact assessment of climate change on floods using d4PDF in the  
792 Chao Phraya River Basin, Thailand, *J. Hydrol.-Reg. Stud.*, 42,  
793 <https://doi.org/10.1016/j.ejrh.2022.101150>, 2022.
- 794 Cannon, A. J., Sobie, S. R., and Murdock, T. Q.: Bias correction of GCM precipitation by  
795 quantile mapping: How well do methods preserve changes in quantiles and extremes?, *J*  
796 *Climate*, 28, 6938-6959, <https://doi.org/10.1175/Jcli-D-14-00754.1>, 2015.
- 797 Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason,  
798 B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Bronnimann, S., Brunet, M.,  
799 Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C.,  
800 Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, O., Ross, T. F., Trigo, R. M., Wang, X.  
801 L., Woodruff, S. D., and Worley, S. J.: The Twentieth Century Reanalysis Project, *Q. J.*  
802 *Roy. Meteor. Soc.*, 137, 1-28, <https://doi.org/10.1002/qj.776>, 2011.
- 803 Cucchi, M., Weedon, G. P., Amici, A., Bellouin, N., Lange, S., Schmied, H. M., Hersbach,  
804 H., and Buontempo, C.: WFDE5: bias-adjusted ERA5 reanalysis data for impact studies,  
805 *Earth Syst. Sci. Data*, 12, 2097–2120-2097–2120, [https://doi.org/10.5194/essd-12-2097-](https://doi.org/10.5194/essd-12-2097-2020)  
806 [2020](https://doi.org/10.5194/essd-12-2097-2020), 2020.

807 Daewel, U., and Schrum, C.: Low-frequency variability in North Sea and Baltic Sea identified  
808 through simulations with the 3-D coupled physical–biogeochemical model ECOSMO,  
809 Earth Syst. Dyn., 8, 801-801, <https://doi.org/10.5194/esd-8-801-2017>, 2017.

810 Daraio, J. A.: Hydrologic Model Evaluation and Assessment of Projected Climate Change  
811 Impacts Using Bias-Corrected Stream Flows, Water, 12,  
812 <https://doi.org/10.3390/w12082312>, 2020.

813 Dirmeyer, P. A., Gao, X., Zhao, M., Guo, Z., Oki, T., and Hanasaki, N.: GSWP-2:  
814 Multimodel Analysis and Implications for Our Perception of the Land Surface, Bull. Amer,  
815 Meteor. Soc., 87, 1381-1398, <https://doi.org/10.1175/bams-87-10-1381>, 2006.

816 Droghei, R., Buongiorno Nardelli, B., and Santoleri, R.: A New Global Sea Surface Salinity  
817 and Density Dataset From Multivariate Observations (1993–2016), Front. Mar. Sci., 5,  
818 <https://doi.org/10.3389/fmars.2018.00084>, 2018.

819 Farkas, C., and Skarbøvik, E.: OSPAR Contracting Parties’ RID 2019 Data Report, NIBIO –  
820 Norwegian Institute for Bioeconomy Research, 57 pp., 2021.

821 Farmer, W. H., Over, T. M., and Kiang, J. E.: Bias correction of simulated historical daily  
822 streamflow at ungauged locations by using independently estimated flow duration curves,  
823 Hydrol. Earth Syst. Sci., 22, 5741-5758, <https://doi.org/10.5194/hess-22-5741-2018>, 2018.

824 Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean  
825 squared error and NSE performance criteria: Implications for improving hydrological  
826 modelling, J. Hydrol., 377, 80–91-80–91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>,  
827 2009.

828 Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N.,  
829 Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N.,  
830 Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P.,  
831 Weedon, G. P., and Yeh, P.: Multimodel estimate of the global terrestrial water balance:  
832 setup and first results, J. Hydrometeorol., 12, 869-884,  
833 <https://doi.org/10.1175/2011jhm1324.1>, 2011.

834 Hagemann, S., Stacke, T., and Ho-Hagemann, H. T. M.: High Resolution Discharge  
835 Simulations Over Europe and the Baltic Sea Catchment, Front. Earth Sci., 8,  
836 <https://doi.org/10.3389/feart.2020.00012>, 2020.

837 Hagemann, S., and Stacke, T.: Complementing ERA5 and E-OBS with high-resolution river  
838 discharge over Europe, Oceanologia, 65, 230-248,  
839 <https://doi.org/10.1016/j.oceano.2022.07.003>, 2022.

840 Hagemann, S., Ho-Hagemann, H. T. M., and Hanke, M.: The Hydrological Discharge Model -  
841 a river runoff component for offline and coupled model applications. Zenodo.  
842 <https://doi.org/10.5281/zenodo.10405875>, 2023.

843 Hagemann, S., and Stacke, T.: Bias corrected high resolution river runoff over Europe. World  
844 Data Center for Climate (WDCC) at DKRZ.  
845 [https://doi.org/10.26050/WDCC/Biasc\\_hr\\_riverro\\_Eu](https://doi.org/10.26050/WDCC/Biasc_hr_riverro_Eu), 2023.

846 Hassler, B., and Lauer, A.: Comparison of reanalysis and observational precipitation datasets  
847 including ERA5 and WFDE5, Atmos., 12, <https://doi.org/10.3390/atmos12111462>, 2021.

848 HELCOM: The Third Baltic Sea Pollution Load Compilation, Balt. Sea Environ. Proc., no 70,  
849 Baltic Marine Environment Protection Commission--Helsinki Commission, Helsinki,  
850 Finland, 134 p. pp., 1998.

851 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J. n.,  
852 Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S.,  
853 Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G.,  
854 Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes,  
855 M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E. a., Janisková, M., Keeley,

856 S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F.,  
857 Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, *Quart. J. Roy. Meteor. Soc.*,  
858 146, 1999-2049, <https://doi.org/10.1002/qj.3803>, 2020.

859 Ho-Hagemann, H. T. M., Hagemann, S., Grayek, S., Petrik, R., Rockel, B., Staneva, J., Feser,  
860 F., and Schrum, C.: Internal Model Variability of the Regional Coupled System Model  
861 GCOAST-AHOI, *Atmos.*, 11, 227-227, <https://doi.org/10.3390/atmos11030227>, 2020.

862 Hordoir, R., Polcher, J., Brun-Cottan, J. C., and Madec, G.: Towards a parametrization of  
863 river discharges into ocean general circulation models: a closure through energy  
864 conservation, *Clim. Dyn.*, 31, 891-908, <https://doi.org/10.1007/s00382-008-0416-4>, 2008.

865 Hordoir, R., and Meier, H. E. M.: Freshwater fluxes in the Baltic Sea: A model study, *J.*  
866 *Geophys. Res.*, 115, C08028-C08028, <https://doi.org/10.1029/2009jc005604>, 2010.

867 ISIMIP: ISIMIP2a Simulation protocol (extended version):  
868 [https://www.isimip.org/documents/647/ISIMIP2a\\_protocol\\_230302.pdf](https://www.isimip.org/documents/647/ISIMIP2a_protocol_230302.pdf), access: 8.3.,  
869 2023.

870 Kim, H.: Global Soil Wetness Project Phase 3 Atmospheric Boundary Conditions  
871 (Experiment 1). Data Integration and Analysis System (DIAS).  
872 <https://doi.org/10.20783/DIAS.501>, 2017.

873 Kim, K. B., Kwon, H. H., and Han, D. W.: Bias-correction schemes for calibrated flow in a  
874 conceptual hydrological model, *Hydrol. Res.*, 52, 196-211,  
875 <https://doi.org/10.2166/nh.2021.043>, 2021.

876 Klein, H., and Frohse, A.: Oceanographic Processes in the German Bight, Heide, Holstein:  
877 Boyens, 60-76, 2008.

878 Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an  
879 ensemble of climate change scenarios, *J. Hydrol.*, 424-425, 264-277-264-277,  
880 <https://doi.org/10.1016/j.jhydrol.2012.01.011>, 2012.

881 Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not?  
882 Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores, *Hydrol. Earth Syst. Sci.*, 23,  
883 4323-4331, <https://doi.org/10.5194/hess-23-4323-2019>, 2019.

884 Krzysztofowicz, R., and Maranzano, C. J.: Hydrologic uncertainty processor for probabilistic  
885 stage transition forecasting, *J Hydrol*, 293, 57-73,  
886 <https://doi.org/10.1016/j.jhydrol.2004.01.003>, 2004.

887 Lehmann, A., and Hinrichsen, H.-H.: On the thermohaline variability of the Baltic Sea, *J.*  
888 *Mar. Syst.*, 25, 333-357, [https://doi.org/10.1016/s0924-7963\(00\)00026-9](https://doi.org/10.1016/s0924-7963(00)00026-9), 2000.

889 Lenhart, H. J., Mills, D. K., Baretta-Bekker, H., van Leeuwen, S. M., van der Molen, J.,  
890 Baretta, J. W., Blaas, M., Desmit, X., Kuhn, W., Lacroix, G., Los, H. J., Menesguen, A.,  
891 Neves, R., Proctor, R., Ruardij, P., Skogen, M. D., Vanhoutte-Brunier, A., Villars, M. T.,  
892 and Wakelin, S. L.: Predicting the consequences of nutrient reduction on the eutrophication  
893 status of the North Sea, *J. Mar. Syst.*, 81, 148-170,  
894 <https://doi.org/10.1016/j.jmarsys.2009.12.014>, 2010.

895 Madadgar, S., Moradkhani, H., and Garen, D.: Towards improved post-processing of  
896 hydrologic forecast ensembles, *Hydrol. Process.*, 28, 104-122,  
897 <https://doi.org/10.1002/hyp.9562>, 2014.

898 Madec, G., Bourdallé-Badie, R., Bouttier, P.-A., Bricaud, C., Bruciaferr, D., Calvert, D.,  
899 Chanut, J., Clementi, E., Coward, A., Delrosso, D., Ethé, C., Flavoni, S., Graham, T.,  
900 Harle, J., Iovino, D., Lea, D., Lévy, C., Lovato, T., Martin, N., and Vancoppenolle, M.:  
901 NEMO ocean engine. Notes du Pôle de modélisation de l'Institut Pierre-Simon Laplace  
902 (IPSL), 27, Zenodo, <https://doi.org/10.5281/zenodo.3248739>, 2017.

903 Malek, K., Reed, P., Zeff, H., Hamilton, A., Wrzesien, M., Holtzman, N., Steinschneider, S.,  
904 Herman, J., and Pavelsky, T.: Bias correction of hydrologic projections strongly impacts



905 inferred climate vulnerabilities in institutionally complex water systems, *J. Water Res.*  
906 *Plan. Man.*, 148, [https://doi.org/10.1061/\(Asce\)Wr.1943-5452.0001493](https://doi.org/10.1061/(Asce)Wr.1943-5452.0001493), 2022.

907 Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M.,  
908 Hagemann, S., Richter, I., Soares, P. M. M., Hall, A., and Mearns, L. O.: Towards process-  
909 informed bias correction of climate change simulations, *Nat Clim Change*, 7, 764-773,  
910 <https://doi.org/10.1038/Nclimate3418>, 2017.

911 Marzeion, B., Levermann, A., and Mignot, J.: The Role of Stratification-Dependent Mixing  
912 for the Stability of the Atlantic Overturning in a Global Climate Model\*, *J. Phys.*  
913 *Oceanogr.*, 37, 2672–2681-2672–2681, <https://doi.org/10.1175/2007jpo3641.1>, 2007.

914 Mengel, M., Treu, S., Lange, S., and Frieler, K.: ATTRICI v1.1-counterfactual climate for  
915 impact attribution, *Geosci Model Dev*, 14, 5269-5284, <https://doi.org/10.5194/gmd-14-5269-2021>, 2021.

917 Merchán, D., Causapé, J., and Abrahao, R.: Impact of irrigation implementation on hydrology  
918 and water quality in a small agricultural basin in Spain, *Hydrol. Sci. J.*, 58, 1400–1413-  
919 1400–1413, 2013.

920 Nguyen, T. T., Staneva, J., Grayek, S., Bonaduce, A., Hagemann, S., Pham, N. T., Kumar, R.,  
921 and Rakovec, O.: Impacts of extreme river discharge on coastal dynamics and  
922 environment: Insights from high-resolution modeling in the German Bight, *Reg. Stud.*  
923 *Mar. Sci.*, 73, <https://doi.org/10.1016/j.rsma.2024.103476>, 2024.

924 Piani, C., Weedon, G. P., Best, M., Gomes, S. M., Viterbo, P., Hagemann, S., and Haerter, J.  
925 O.: Statistical bias correction of global simulated daily precipitation and temperature for  
926 the application of hydrological models, *J Hydrol*, 395, 199-215,  
927 <https://doi.org/10.1016/j.jhydrol.2010.10.024>, 2010.

928 Shi, X. G., Wood, A. W., and Lettenmaier, D. P.: How Essential is Hydrologic Model  
929 Calibration to Seasonal Streamflow Forecasting?, *J. Hydrometeorol.*, 9, 1350-1363,  
930 <https://doi.org/10.1175/2008jhm1001.1>, 2008.

931 Stacke, T., and Hagemann, S.: HydroPy (v1.0): A new global hydrology model written in  
932 Python, *Geosci. Model Dev.*, <https://doi.org/10.5194/gmd-2021-53>, 2021.

933 Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram., *J*  
934 *Geophys Res-Atmos*, 106, 7183-7192, [https://doi.org/Doi 10.1029/2000jd900719](https://doi.org/Doi%2010.1029/2000jd900719), 2001.

935 Teutschbein, C., and Seibert, J.: Bias correction of regional climate model simulations for  
936 hydrological climate-change impact studies: Review and evaluation of different methods, *J*  
937 *Hydrol*, 456-457, 12-29, <https://doi.org/10.1016/j.jhydrol.2012.05.052>, 2012.

938 Väli, G., Meier, H. E. M., and Elken, J.: Simulated halocline variability in the Baltic Sea and  
939 its impact on hypoxia during 1961-2007, *J. Geophys. Res. Oceans*, 118, 6982–7000-6982–  
940 7000, <https://doi.org/10.1002/2013jc009192>, 2013.

941 Van Leeuwen, S., and Lenhart, H. J.: OSPAR ICG-EMO riverine database 2020-05-01 used  
942 in 2020 workshop. NIOZ, V1. <https://doi.org/10.25850/nioz/7b.b.vc>, 2021.

943 Van Leeuwen, S., and Hagemann, S.: Mapping of IGC-EMO nutrient loads on the high  
944 resolution HD model grid (Version 1). World Data Center for Climate (WDCC) at DKRZ.  
945 [https://doi.org/10.26050/WDCC/IGC-EMO\\_HD\\_v1](https://doi.org/10.26050/WDCC/IGC-EMO_HD_v1), 2023.

946 Vinayachandran, P. N., Jahfer, S., and Nanjundiah, R. S.: Impact of river runoff into the ocean  
947 on Indian summer monsoon, *Environ. Res. Lett.*, 10, <https://doi.org/10.1088/1748-9326/10/5/054008>, 2015.

949 Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., and Schewe, J.: The Inter-  
950 Sectoral Impact Model Intercomparison Project (ISI-MIP): Project framework, *Proc. Natl.*  
951 *Acad. Sci. USA*, 111, 3228-3232, <https://doi.org/10.1073/pnas.1312330110>, 2014.

952 Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., Adam, J.  
953 C., Bellouin, N., Boucher, O., and Best, M.: Creation of the WATCH Forcing Data and Its

954 Use to Assess Global and Regional Reference Crop Evaporation over Land during the  
955 Twentieth Century, *J. Hydrometeorol.*, 12, 823-848,  
956 <https://doi.org/10.1175/2011JHM1369.1>, 2011.

957 Yoshimura, K., and Kanamitsu, M.: Dynamical global downscaling of global reanalysis, *Mon.*  
958 *Weather Rev.*, 136, 2983-2998, <https://doi.org/10.1175/2008mwr2281.1>, 2008.

959 Zhao, L., Duan, Q., Schaake, J., Ye, A., and Xia, J.: A hydrologic post-processor for ensemble  
960 streamflow predictions, *Adv. Geosci.*, 29, 51-59, [https://doi.org/10.5194/adgeo-29-51-](https://doi.org/10.5194/adgeo-29-51-2011)  
961 [2011](https://doi.org/10.5194/adgeo-29-51-2011), 2011.

962 Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., and Mayer, M.: The ECMWF  
963 operational ensemble reanalysis–analysis system for ocean and sea ice: a description of the  
964 system and assessment, *Ocean Sci.*, 15, 779-808, <https://doi.org/10.5194/os-15-779-2019>,  
965 2019.

966

967