



Streamflow-based evaluation of climate model sub-selection methods

Jens Kiesel^{1,2} · Philipp Stanzel³ · Harald Kling³ · Nicola Fohrer² ·
Sonja C. Jähnig¹ · Ilias Pechlivanidis⁴

Received: 30 December 2019 / Accepted: 30 August 2020 / Published online: 10 October 2020
© The Author(s) 2020

Abstract

The assessment of climate change and its impact relies on the ensemble of models available and/or sub-selected. However, an assessment of the validity of simulated climate change impacts is not straightforward because historical data is commonly used for bias-adjustment, to select ensemble members or to define a baseline against which impacts are compared—and, naturally, there are no observations to evaluate future projections. We hypothesize that historical streamflow observations contain valuable information to investigate practices for the selection of model ensembles. The Danube River at Vienna is used as a case study, with EURO-CORDEX climate simulations driving the COSERO hydrological model. For each selection method, we compare observed to simulated streamflow shift from the reference period (1960–1989) to the evaluation period (1990–2014). Comparison against no selection shows that an informed selection of ensemble members improves the quantification of climate change impacts. However, the selection method matters, with model selection based on hindcasted climate or streamflow alone is misleading, while methods that maintain the diversity and information content of the full ensemble are favorable. Prior to carrying out climate impact assessments, we propose splitting the long-term historical data and using it to test climate model performance, sub-selection methods, and their agreement in reproducing the indicator of interest, which further provide the expectable benchmark of near- and far-future impact assessments. This test is well-suited to be applied in multi-basin experiments to obtain better understanding of uncertainty propagation and more universal recommendations regarding uncertainty reduction in hydrological impact studies.

Keywords Hindcast · Climate uncertainty · Ensemble selection · EURO-CORDEX · Climate change impact

This article is part of a Special Issue on “How evaluation of hydrological models influences results of climate impact assessment,” edited by Valentina Krysanova, Fred Hattermann, and Zbigniew Kundzewicz

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10584-020-02854-8>) contains supplementary material, which is available to authorized users.

✉ Jens Kiesel
kiesel@igb-berlin.de

Extended author information available on the last page of the article

1 Introduction

It is a common practice to analyze multiple ensemble members for climate change impact studies, which is important to account for the variability and uncertainty in the projections (Melsen et al. 2018; Krysanova et al. 2017). Use of a large model ensemble of climate projections is favored for impact modeling with the aim to quantify the inherent uncertainties in the projections (Clark et al. 2016; Pechlivanidis et al. 2017; Samaniego et al. 2017). However, in impact studies, the use of a large number of climate models can be computationally and methodologically intensive, while a smaller number of climate-impact models may be necessary to provide users with a manageable set for decision-making (Evans et al. 2013). Therefore, it is not always possible to present all available ensemble members, e.g., when complex model cascades are used (Kiesel et al. 2019a) or when multiple other sources of uncertainty are of interest as well (Clark et al. 2016). While the practice of sub-selecting large model ensembles has been criticized in the past (Mote et al. 2011; Christensen et al. 2010), it is now a generally accepted approach (Eyring et al. 2019; Herger et al. 2018; Knutti et al. 2017). This is mainly due to the fact that informed model sampling and reduction of informational redundancy in the ensemble are expected to improve climate change impact assessments (Eyring et al. 2019; Pechlivanidis et al. 2018). A number of methods have been proposed to deal with the multi-ensemble problem, e.g., selections based on best-performing climate depiction (Ruane and McDermid 2017); best representation of the target variable, e.g., streamflow (Kiesel et al. 2019b); keeping model diversity and independence (Abramowitz et al. 2019); weighted ensemble members driven by the model performance (Knutti et al. 2017); and trading-off the information content, gain, and redundancy in the ensemble set (Pechlivanidis et al. 2018).

Whether to exclude or weigh the available models according to different criteria still remains a subjective exercise. Eyring et al. (2019) provides a recent overview of suitable measures to assess the climate model performance in order to exclude and/or weigh the available models. However, to date no comparison of these selection methods using historical climate change impacts has been performed. This is significant since the effect on climate impact results depends on the selection method, i.e., exclusion/weighting method, available models, and performance criteria used for setting the weights (Knutti et al. 2010; Evans et al. 2013).

A comparison of selection methods is possible if evaluation data are available. Such data has to include the combined influence of meteorological forcing, spatial heterogeneity of the climate model's performance (Kotlarski et al. 2014), and the long time periods over which change occurs. Streamflow of large and complex (e.g., alpine) river basins integrates climatic factors due to the spatiotemporal interaction between meteorological and hydrological processes. Even though future climate is expected to be more pronounced than what has already been observed, from a hydrological viewpoint, climate change has already had an observable impact in various basins (Blöschl et al. 2017). We argue that historic streamflow observations can be used as an evaluation criterion to assess (1) hindcasted climate model skill and (2) ensemble selection methods. Such an evaluation requires the simulation of streamflow based on hindcasted temperature and precipitation data from climate models. On one hand, this simplifies the climate model assessment from two parameters (precipitation, temperature) to one (streamflow), while on the other hand, this also adds an additional layer of uncertainty—the hydrological model and its associated data (Clark et al. 2016).

Pechlivanidis et al. (2017) and Thober et al. (2018) show that uncertainty (both related to climate and hydrological models) depends on the climatic conditions and is generally higher in the dry than in the wet regions, while uncertainty in high (low) flows is related more (slightly less) to the climate models compared with the hydrological models. Melsen et al. (2018) come to similar conclusions, but at the same time show that hydrological model choice and parameterization can influence the direction of the projected change. Therefore, attributing the simulated streamflow change signal clearly to a change in climate either requires the consideration of the hydrological model and parameterization uncertainty in the analysis (Addor et al. 2014; Melsen et al. 2018; Vetter et al. 2017) or a rigorous five-step evaluation of the model's performance (Krysanova et al. 2018).

Considering the above, this paper's objectives are (1) to carry out a rigorous hydrological model evaluation to ensure suitability for climate change impact assessments, (2) to assess how well different climate models are able to reproduce actual climate change impact on streamflow, and (3) to provide an evaluation of eight climate model sub-selection methods.

2 Material and data

2.1 Study area

The Upper Danube up to the gauge Vienna has a basin area of 101,810 km² where annual precipitation ranges from below 500 mm in the eastern lowlands to up to 3000 mm in the Alps. Dominant land uses are forest, grassland, and agriculture, whereas bare areas and glaciers are located in the high alpine regions. The basin has a high elevation gradient ranging from about 200 to 4000 m a.s.l. Overall, the basin is very heterogeneous, with alpine sub-basins such as the Inn or lowland basins such as the Naab and their distinct seasonal streamflow regimes (Fig. 1). At Vienna, the alpine influence with higher streamflow in summer prevails. In the alpine basin of the Inn (gauges Oberaudorf and Hofkirchen), streamflow in summer has been reduced, while streamflow in winter has increased, attributed to an increase in warm years within the recent decades (Kling et al. 2012b; Stanzel and Kling 2018). Due to this, the Upper Danube's streamflow has undergone a significant shift in seasonality showing that climate change has already impacted the basin's hydrological processes.

2.2 Hydrological model

The model COSERO (Kling et al. 2015) is a conceptual hydrological model that includes the simulation of snowfall and melt processes, positive and negative glacier mass balances, actual evapotranspiration, fast runoff, and baseflow components, as well as reservoir storages and releases. In the application of the model for the Upper Danube (Kling et al. 2012a), the model was forced with HISTALP (Auer et al. 2007; Böhm et al. 2009) climate data, available at 30 stations of monthly time series of precipitation, temperature, and sunshine duration from 1800 to 2014. For precipitation, HISTALP was merged with annual 1-km²-gridded precipitation data (Kling et al. 2007), while temperature was further regionalized by a linear regression model based on elevation. Potential evapotranspiration (PET) was calculated using HISTALP sunshine duration, the Angstrom equation (Angstrom 1924), and the Turc PET method (Turc 1961). The model was discretized in 16 sub-basins, where each sub-basin outlet is located at a streamflow gauge (Fig. 1). The sub-basins were further distinguished into 61 hydrological

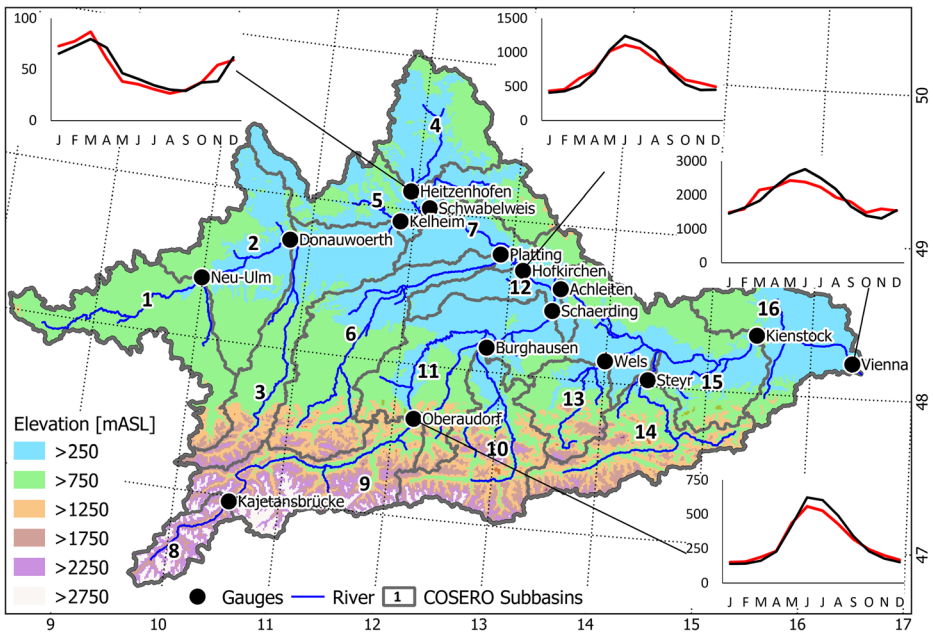


Fig. 1 Basin area of the Upper Danube with the 16 sub-basins, according gauges at the sub-basin outlets and the elevation zones; four diagrams show the streamflow seasonality for Naab (Heitzenhofen), Inn (Hofkirchen), Danube (Vienna), and Inn (Oberaudorf) based on observed streamflow data from 1960 to 1989 (black) and from 1990 to 2007 (red); grid numbers represent latitude and longitude

response units (HRUs) according to elevation bands of 500 m elevation difference. COSERO was applied on a monthly time step from 1800 to 2014. Model parameters were calibrated for the time period 1961–1990 for 16 gauges combining manual and automatic calibration methods using the Kling-Gupta efficiency (KGE^{*}, Kling et al. 2012a) and its sub-components as the objective function. Detailed evaluation of model simulations with observed streamflow data was carried out in the independent evaluation periods 1901–1930, 1931–1960, and 1991–2007 as well as in unusually warm years.

2.3 Climate change data

Monthly precipitation and monthly temperature were acquired from 16 climate models from the EURO-CORDEX initiative (Table 1) (Jacob et al. 2014). We chose the Representative Concentration Pathway (RCP) 8.5 emission scenario since recorded emissions since 2006 have followed the RCP8.5 trajectory closest (Sanford et al. 2014). COSERO was forced with both the raw and bias-adjusted EURO-CORDEX data. For the bias-adjustment, linear scaling was applied on both precipitation and temperature data for the time period 1961–1990 prior to forcing the COSERO model with the data. PET was calculated with the Turc PET method for each climate model temperature dataset, assuming global radiation values of the early twenty-first century. From the model runs, hindcasted streamflow from 1960 to 2014 was obtained (note that 1970–2014 was used for the SMHI-RCA GCM family of models). For each hindcasted streamflow time series, the average annual change in streamflow from the 1960–1989 (reference) to the 1990–2014 (evaluation) period (ΔQ in Table 1) was calculated. This ΔQ is used as the target parameter to select the median model from different sub-ensembles.

Table 1 GCM and RCM data used from the EURO-CORDEX CMIP5 ensemble with change in annual streamflow from the 1960–1979 to the 1980–2014 period; the numbers of the sub-selection methods indicate model is not used (0), model is selected within a subset (> 0), and finally selected models (= median of the subset) for the evaluated models are marked bold: Dem = model democracy, DivG = diversity of GCMs, DivR = diversity of RCMs, MIMR = maximum information minimum redundancy, bCl = best hindcasted climate, bSf = best hindcasted streamflow, sWgt = simple weights, REA = reliability ensemble average

Nr	GCM	RCM	Annual ΔQ (m ³ /s)	Sub-selection methods								
				Dem	DivG	DivR	MIMR	bCl	bSf	sWgt	REA (wgt)	
1	CLMcom-CCLM	CNRM	-67.6	1	0	1	0	0	0	0	13	2.20
2	CLMcom-CCLM	ICHEC	-107.8	1	1	1	0	1	0	0	16	2.18
3	CLMcom-CCLM	MOHC	-60.8	1	0	1	1	0	0	0	11	2.32
4	CLMcom-CCLM	MPI	67.8	1	0	1	0	0	0	0	12	2.09
5	CNRM-ALADIN	CNRM	41.1	1	1	0	0	0	0	0	8	2.01
6	DMI-HIRHAM	ICHEC	-21.4	1	1	0	0	0	0	0	6	1.71
7	IPSL-WRF	IPSL	-36.3	1	1	1	0	0	0	0	1	1.79
8	KNMI-RACMO	ICHEC	-38.5	1	0	0	0	0	0	0	4	1.83
9	KNMI-RACMO	MOHC	7.3	1	0	0	0	0	0	0	9	2.15
10	MPI-REMO	MPIr1	-89.9	1	0	1	0	0	0	0	14	2.19
11	MPI-REMO	MPIr2	149.4	1	1	1	0	0	0	1	15	1.70
12	SMHI-RCA	CNRM	45	1	0	0	0	0	0	0	2	1.92
13	SMHI-RCA	ICHEC	16.3	1	0	0	0	0	0	0	5	2.02
14	SMHI-RCA	IPSL	154.2	1	0	0	0	0	0	0	3	1.02
15	SMHI-RCA	MOHC	-167.5	1	1	0	0	0	0	0	10	1.32
16	SMHI-RCA	MPI	95.5	1	0	0	1	0	0	0	7	1.83

3 Methods

3.1 Hydrological model evaluation

Through a rigorous model evaluation, in case it is successful, the confidence in the results of the climate change impact assessment can be increased (Huang et al. (this SI); Gelfan et al. (this SI)). This is especially important since here we only consider one hydrological model. Therefore, the evaluation of the calibrated model is carried out according to the following five-step criteria proposed by Krysanova et al. (2018):

Quality of observational data (step 1): Observed climate input data used to parameterize the model, particularly temperature and precipitation, has to be spatially representative, homogeneous over the whole time period, and of high quality. Streamflow data has to be observed with up-to-date rating curves, be consistent between upstream and downstream gauges, and available for the full calibration and validation period.

Good performance for periods with different climate (step 2): The hydrological model is able to reproduce the streamflow situation in exceptionally dry and wet as well as hot and cold years. Model performance is therefore assessed separately for the 10 warmest, 10 coldest, 10 wettest, and 10 driest years within the period 1901 to 2007 at the gauge Vienna.

Good performance at multiple sites and for multiple variables (step 3): The Upper Danube is a heterogeneous basin ranging from alpine, pre-alpine to lowland regions. Since the climate model's spatial representation of precipitation and temperature is of key

importance in mountainous basins as the Upper Danube, the hydrological model has to adequately perform in all regions to not cause a spatial bias in the evaluation.

The mass balance of the glaciers in the high alpine region of the Upper Danube is an additional variable that has been evaluated. While the contribution of glacier mass balance is less than 1% of the annual streamflow at Vienna, it is nevertheless an important indicator to assess the model's process depiction in high-alpine regions that generally contribute most of the streamflow. Having a realistic simulation of glacier mass balance ensures that the distinction of precipitation in rainfall and snowfall, as well as simulation of snow- and ice-melt, is plausible.

Reproducibility of the hydrological indicator of interest (step 4): Due to increasing temperatures within the last 60 years, streamflow seasonality at Vienna has been flattened with a reduction in summer streamflow and an increase or no change in the remaining seasons. Here, we test the potential of the hydrological model forced with observed climate data to reproduce the long-term inter-annual seasonal shift.

Reproducibility of significant trends in the streamflow (step 5): An ordinary Mann–Kendall trend test (Mann 1945; Kendall 1955; implemented in Hussain and Mahmud 2019) is applied on the observed and simulated streamflow data at Vienna. It is then compared whether and to what extent the significant trends and insignificant trends are matching in both or not.

3.2 Impact of climate change on precipitation, temperature, and streamflow

We investigate the impact of climate change that occurred from the reference to the evaluation period for observed and hindcasted data. Therefore, we calculate the average change as well as the spatial variability (standard deviation of change over the 16 sub-basins) in temperature, precipitation, and streamflow. This analysis provides insights into how strong temperature, precipitation, or the combination of both are linked to streamflow change.

3.3 Ensemble selection methods

Eight different sub-selection methods are applied using data from the reference period and from the 16 raw (not bias-adjusted) EURO-CORDEX datasets (Table 1):

Democracy (Dem): all models are treated equally (ensemble of opportunity, see e.g., IPCC 2013) where the median model is selected from the full ensemble and, if the number of ensemble members is even, is calculated as the mean between two members;

Diversity of GCM (DivG): from each GCM group, selecting the model that best represents hindcasted climate seasonality leads to a subset of seven models from which the median is selected;

Diversity of RCM (DivR): same as for DivG but here RCMs are used instead of GCMs;

Maximum Information Minimum Redundancy method (MIMR): the MIMR (Li et al. 2012) method is implemented to select the members that maximize independence and minimize redundancy and contain more than 90% of the hydro-climatic spatial information;

Best performing hindcasted climate depiction (bCl): the model that best represents hindcasted climate seasonality is chosen;

Best performing hydrological model (bSf): from all COSERO realizations driven by the raw hindcasted climate data, the model that best represents the hindcasted streamflow seasonality is chosen;

Simple climate model weighing (sWGT): models are ranked according to hindcasted climate performance where the worst model is used once in the ensemble and, the subsequently better model one more time, leading to an ensemble set of 136 members from which the median model is selected;

Reliability Ensemble Average (REA): model weights (see details regarding weighting calculations in the [Appendix](#)) are increased depending on historical climate performance of the raw climate model data as well as the consensus of models in the future regarding the hydrological indicator of interest (Tebaldi and Knutti 2007).

3.4 Evaluation of model ensembles

Each of the 16 EURO-CORDEX models, all medians from all possible model combinations, and the eight ensemble selection and weighing methods (Table 1) are evaluated according to how well they capture the difference in simulated streamflow change within the four seasons. The change in simulated streamflow based on the observed climate is used as benchmark to ensure that the evaluated difference is only attributed to the climate change signal, while any hydrological model impacts are excluded. For the evaluation, the change in streamflow simulated with hindcasted climate (henceforth designated “hindcasted”) within each season from the reference period (1960–1989) and the evaluation period (1990–2014) is therefore compared with the change in streamflow simulated with observed climate (henceforth designated “observed”). The average RMSE over the four seasons between the observed and simulated change is selected as the measure of agreement.

4 Results

4.1 Hydrological model performance assessment

Here, we assess the hydrological model performance using the five-step evaluation approach (see section 3.1).

Step 1: Climate data is based on the high-quality HISTALP dataset, whose development started in the early 1990s, involved 33 organizations and data providers from 15 countries. The HISTALP initiative focuses on long time series utilizing all available systematically observed data, a dense network adequate for the spatial heterogeneity, outliers removed, gaps filled, and accounting for changes in measurement techniques. Regarding streamflow, the 16 gauges used are key gauges of the most important rivers in the Upper Danube basin. Hydrographic Services carry out multiple streamflow measurements per year to keep the rating curve up-to-date. In addition, streamflow values are verified with upstream- and downstream gauges to avoid any possible inconsistencies. The period of

available data is different for different gauges, with data available for all gauges from 1954 to 2007. See also Kling et al. (2012a and b). We therefore conclude that data quality is acceptable and does not hamper achieving good model performance.

Step 2: For different climate conditions, the model performs consistently well over the different climatic conditions of the 10 driest, 10 warmest, 10 coldest, and 10 wettest years at the gauge Vienna, with the lowest KGE' of 0.85 for the driest years and the highest KGE' of 0.9 for the warmest years (see Table 2).

Step 3: The model performs well across the 16 sub-basins with the KGE' varying from 0.81 to 0.97 (Table 2). The model shows a tendency to perform better for the snow-dominated alpine sub-basins (south of the main stem of the Upper Danube; Fig. 1) with all KGE' values being greater than 0.9 in comparison to the pre-alpine and lowland region (all KGE' values < 0.9). Besides streamflow, the agreement between observed and simulated change in glacier mass is investigated. Observations of glacier mass are available for the Vernagtferner glacier (Weber 2003) which is located in subbasin 9 (Fig. 1). The COSERO model simulates glacier processes on the full subbasin scale and, besides the Vernagtferner, includes multiple other glaciers in subbasin 9. Therefore, a comparison between observed and simulated glacier mass is only indirectly possible. Figure 2 a shows a comparison between relative change in glacier mass from 1840 to 2000. The underestimation of simulated glacier mass loss can be due to the lumped glacier depiction which is compared with data from one glacier only. This higher observed mass change compared with other glaciers in the subbasin seems realistic since the Vernagtferner shows characteristics which increase its vulnerability to melting processes: The Vernagtferner has largely lost its protective and comparably shallow and wide firn-body and has significant areas with south-

Table 2 Performance statistics for the full time series of the 16 gauges and for the 10 warmest, 10 coldest, 10 wettest, and 10 driest years of the 107-year period at Vienna (KGE' = modified Kling-Gupta efficiency, r = correlation, β = bias ratio, γ = variability ratio; all dimensionless; ideal value at unity, equations given in Kling et al. 2012a; PBIAS = percent bias with ideal value at zero)

Gauge	Sub	Years	KGE'	r	β	γ	PBIAS
Neu-Ulm	1	54	0.89	0.89	0.98	0.99	-1.60
Donauwörth	2	84	0.88	0.88	0.99	0.96	-1.09
Kelheim	3	107	0.87	0.88	0.99	0.95	-0.75
Heitzenhofen	4	87	0.81	0.84	0.96	0.91	-4.10
Schwabelweis	5	84	0.86	0.88	1.00	0.94	0.35
Platten	6	82	0.84	0.85	0.96	1.03	-3.73
Hofkirchen	7	107	0.84	0.85	1.00	0.93	-0.23
Kajetansbrücke	8	57	0.95	0.96	0.98	1.01	-1.80
Oberaudorf	9	107	0.97	0.97	1.01	1.01	0.85
Burghausen	10	107	0.93	0.94	0.99	1.01	-0.75
Schärding	11	57	0.96	0.96	1.00	1.01	0.25
Achleiten	12	107	0.92	0.92	0.99	1.00	-1.40
Wels	13	57	0.91	0.93	1.00	1.05	0.14
Steyr	14	57	0.91	0.91	0.99	0.98	-0.60
Kienstock	15	77	0.92	0.92	1.01	0.99	1.24
Vienna	16	107	0.91	0.91	1.01	0.97	0.48
Vienna	16	10 warmest	0.90	0.93	1.02	0.93	2.08
Vienna	16	10 coldest	0.88	0.89	1.01	0.94	0.59
Vienna	16	10 wettest	0.86	0.89	1.01	0.91	0.84
Vienna	16	10 driest	0.85	0.91	0.95	0.89	-5.38

facing slopes and multiple, rather short and wide, glacier tongues (Braun and Escher-Vetter 2013).

Step 4: The model is able to reproduce well the observed streamflow seasonal shift from the reference to the evaluation period; however, there is a slight overestimation of streamflow change in the winter period (Dec.–Feb.) and a slight underestimation of the change in the autumn period (Sep.–Nov.; Fig. 2b). In addition, on the monthly time step, the model is able to adequately reproduce the absolute streamflow as well as the change from the reference to the evaluation period (Fig. 2c).

Step 5: Based on results from the Mann Kendall test, no trend is present in the observed and simulated time series of streamflow at Vienna for both the time periods 1901–2014 and 1960–2014 ($p = 0.31$ to 0.66).

Driven by the above results, we conclude that the Upper Danube COSERO model is suitable for a climate change impact assessment.

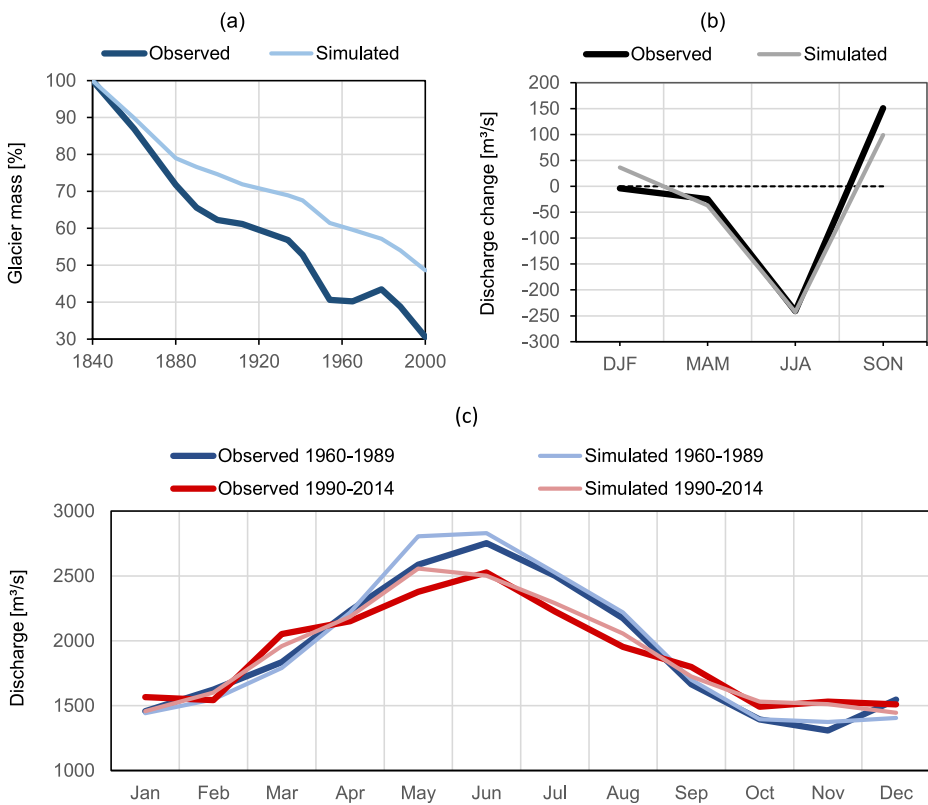


Fig. 2 a Normalized glacier mass development for observations of the Vernagtferner glacier (Weber 2003) and simulations of subbasin 9. b Streamflow of the Danube at Vienna for observations and simulations with observed climate for change in seasonality (DJF = winter-, MAM = spring-, JJA = summer-, SON = autumn months) from 1960 to 1989 period to the 1990–2014 period; dashed line indicates 0 change. c Observed and simulated absolute long-term monthly average streamflow for the two periods

4.2 Impact of climate change

Here, we investigate the changes in temperature, precipitation, and streamflow using both observed and simulated data (Fig. 3).

Temperature increase is consistently shown in all 16 models and the observations across all seasons as well as for the annual average (Fig. 3a). Compared with the observations, most models predict a low-temperature increase in spring (Mar.–May). The increase in temperature magnitudes varies between the models; e.g., model 11 (MPI-REMO + MPIr2) indicates the lowest temperature increase, while model 7 (IPSL-WRF + IPSL) indicates the highest temperature increases across the Upper Danube basin.

Precipitation is subject to a more heterogeneous change from the reference to the evaluation period than temperature. Observed precipitation increases in summer and autumn and decreases slightly in winter and spring. However, no model is capable of fully reproducing this change. In spring, all models predict increased precipitation, while for the remaining seasons and the average annual changes, precipitation both increases and decreases depending on the climate model. Model 8 (KNMI-RACMO + ICHEC) shows a similar pattern in precipitation as the observations (increase in autumn and no significant change over the remaining seasons) and also a consistent increase in temperature during all seasons. Consequently, this model performs best in predicting the streamflow change.

Most climate model-based simulations predict a reduction in streamflow for summer and autumn and an increase in winter and spring. For the annual average, the projected streamflow change varies between climate models. Observed streamflow changes show a strong reduction in summer, a slight reduction in spring (MAM), and increases in the other seasons. The observed annual average streamflow slightly decreases from the reference to the evaluation period.

Reduction in precipitation in winter (e.g., models 4, 10, and 16) does not directly impact winter streamflow, while rising temperatures and subsequently more rainfall and less snowfall in winter cause an increase in winter streamflow. Increasing temperatures mean less accumulation of water in the snowpack and earlier onset of snowmelt in spring, with the overall effect

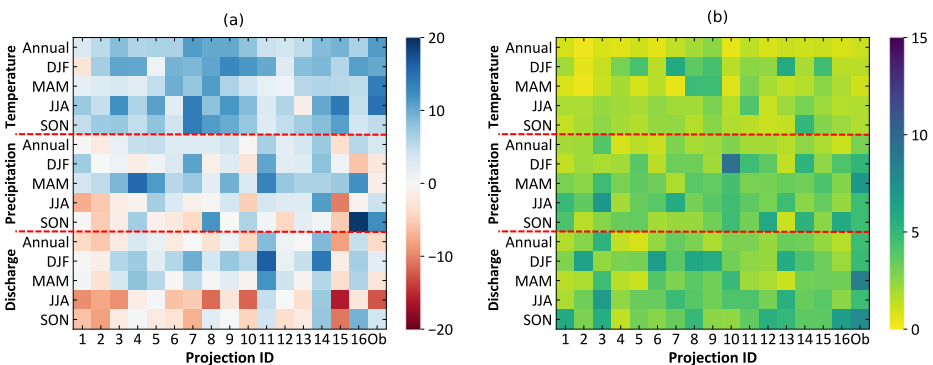


Fig. 3 Information about the GCM and RCM for the mean statistic of the hydro-climatic variables and for different seasons: **a** mean relative change (%) from the 16 sub-basins from reference to evaluation period (note for temperature the actual change ($^{\circ}\text{C}$) $\times 10$ is used to fit the color scale) and **b** standard deviation in relative change (%) from reference to evaluation period which shows the heterogeneity across the 16 sub-basins (note for temperature the actual standard deviation ($^{\circ}\text{C}$) $\times 30$ is used to fit the color scale) (Projection ID on x-axis refers to model nr.-column in Table 1; Ob refers to observed data)

of decreased streamflow in summer. The increasing temperatures and consequently increasing potential evapotranspiration in the summer period causes a further reduction in streamflow.

Spatial heterogeneity of observed streamflow and precipitation change across the Upper Danube is highest during spring and lowest during summer, while that of observed temperature change is generally low (Fig. 3b). This indicates that temperatures from reference to evaluation period have changed similarly across the basin, while precipitation has changed in a more spatially distributed manner. Modeled temperature change is more variable across the 16 sub-basins during winter and less variable during summer and autumn and also for the annual average. Models 2, 10, and 16 show the lowest temperature variability over the seasons. Apart for model 3 and the observation, there is no clear relationship projected between the change in precipitation and streamflow. Overall, all models have difficulties on reproducing the observed spatial change pattern across the basin.

4.3 Hindcasted climate sensitivity

Here, we investigated if change in precipitation and temperature can be linked to a change in streamflow. The spatial pattern of change for temperature, precipitation, and streamflow in the Upper Danube is complex due to alpine processes and the spatial physiographic heterogeneity. These processes are valuable for the model selection, as for instance, for a model to perform well in predicting the observed streamflow change in the Upper Danube, it seems important that the projected increase in winter and spring temperature is correct (models 6 and 8). However, inferring the ability in predicting streamflow change based on temperature and precipitation change is challenging for the remaining models (models 1–5, 7, 8–16). These results highlight the importance to evaluate streamflow as a target variable directly and may explain why precipitation and temperature alone provide insufficient information to select climate models in a complex basin such as the Upper Danube. We therefore investigated the link between a model's potential to reproduce historical climate seasonality and streamflow change. The RMSE values of the observed against hindcasted climate seasonality for the period 1961–1990 in the Upper Danube for all EURO-CORDEX models are available from Stanzel and Kling (2018). These RMSE values for precipitation and temperature were ranked and compared with the rank of the RMSE of the streamflow change for all GCM-RCM (Table 3) to assess if models that depict past climate well are more likely to depict the streamflow change. Figure 4 shows that no relation exists between a good model for temperature, precipitation, or both combined, to the models that best predict the streamflow change.

4.4 Identification of model subsets

Applying the eight different sub-selection methods leads to a different selection of EURO-CORDEX models for each method (Table 1, bold marked numbers). Overall, average seasonal RMSE values range between 38 and 241 m³/s across all models, which is quite significant given that the actual change signal lies in the same order of magnitude (+36.2 to –242.7 m³/s).

While for the methods DivG, DivR, bCI, bSf, and sWgt exactly one median model is selected, the remaining methods lead to a selection of multiple models. The ensemble of opportunity (Dem) method selects model 6 (DMI-HIRHAM+ICHEC) and model 9 (KNMI-RACMO+MOHC), while model 3 (CLMcom-CCLM + MOHC) and model 16 (SMHI-RCA

Table 3 RMSE between the observed (simulated streamflow with observed climate) and hindcasted streamflow change, according rank (1 = best) and seasonal streamflow change from the reference to the evaluation period for each climate model and sub-selection method; bold font indicates the correct depiction of the observed direction of change within the season, italic font indicates the correct depiction of the size proportion of the change over all seasons

ID	GCM	RCM	Method	RMSE	Rank	Streamflow change (m ³ /s)			
						DJF	MAM	JJA	SON
-	-	-	MIMR	68	2	95.3	64.3	-178.3	70.0
-	-	-	REA	96	7	63.6	44.0	-112.8	-14.2
-	-	-	Dem	99	8	85.2	-24.1	-74.2	9.6
1	CLMcom-CCLM	CNRM		94	5	-7.5	23.5	-198.5	-68.8
2	CLMcom-CCLM	ICHEC	bCI	130	11	-48.2	-47.1	-136.5	-121.5
3	CLMcom-CCLM	MOHC	DivR	78	3	59.8	2.8	-307.7	-36.1
4	CLMcom-CCLM	MPI		157	14	84.3	170.3	-31.9	3.1
5	CNRM-ALADIN	CNRM		169	17	13.6	114.9	40.1	-2.4
6	DMI-HIRHAM	ICHEC	DivG	91	4	83.1	-32.1	-93.8	4.8
7	IPSL-WRF	IPSL		142	13	106.4	72.6	-145.8	-135.0
8	KNMI-RACMO	ICHEC	sWgt	38	1	57.6	21.5	-237.9	54.2
9	KNMI-RACMO	MOHC		107	9	87.3	-16.2	-54.7	14.4
10	MPI-REMO	MPIr1		95	6	56.4	-51.9	-321.4	-71.1
11	MPI-REMO	MPIr2	bSf	238	18	231.5	237.3	90.7	61.3
12	SMHI-RCA	CNRM		157	15	51.2	100.1	25.8	9.9
13	SMHI-RCA	ICHEC		117	10	39.3	26.1	-38.6	1.3
14	SMHI-RCA	IPSL		241	19	218.3	78.3	187.7	106.4
15	SMHI-RCA	MOHC		168	16	-59.4	-100.2	-400.7	-174.3
16	<i>SMHI-RCA</i>	<i>MPI</i>		<i>141</i>	<i>12</i>	<i>130.8</i>	<i>125.8</i>	-48.8	<i>176.1</i>
	OBS	-		-	-	36.2	-37.0	-242.7	99.0

+ MPI) are selected by MIMR (note that according to the latter method those two models contain about 90% of the information content of the large ensemble set; see the sum of red dots per column in Fig. 5). Finally based on the REA method, change is calculated using all models with the according weights (right column in Table 1).

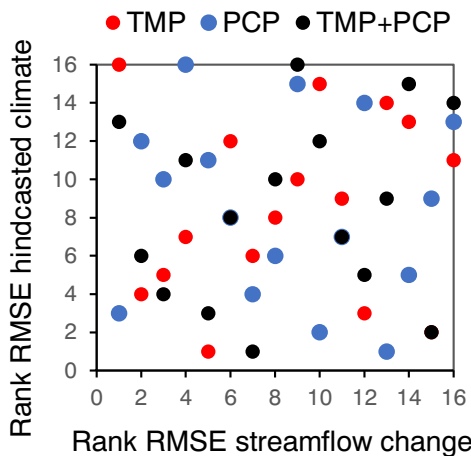


Fig. 4 Ranks of all GCM and RCM models for RMSE between observed and simulated streamflow change (x-axis) and for RMSE between hindcasted and observed climate seasonality; rank 1 denotes best representation, TMP temperature, PCP precipitation

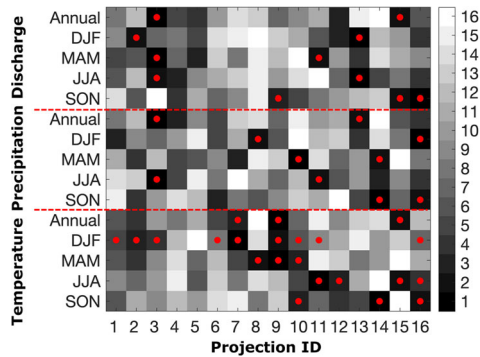


Fig. 5 Information about the GCM and RCM for the mean statistic of the hydro-climatic variables and for different seasons (Projection ID on x-axis refers to nr.-column in Table 1). Boxes with a red dot indicate membership to a subset representing >90% of the total information from the large ensemble of 16 members. Shades of gray relate to the ranking of the climate model in terms of information content for the respective parameter and season (from 1 to 16 with 1 being the most important)

Streamflow change from the reference to the evaluation period was calculated for all selected subset models (Table 3). Streamflow change results of the sub-selection methods MIMR, model democracy (Dem), and REA are not directly related to one specific GCM-RCM model (first 3 rows in Table 3). The medians of all other methods correspond to results of one model from the subsets of all GCM-RCM models. None of the medians and means selected from the EURO-CORDEX models reproduce the observed change better than model 8 (KNMI-RACMO + ICHEC, Table 3). Model 9 (KNMI-RACMO + MOHC), model 6 (DMI-HIRHAM + ICHEC), and the median resulting from the democracy (Dem) selection method are the only ones able to correctly reproduce the direction of change in all seasons. Moreover, only model 16 (SMHI-RCA + MPI) is capable of reproducing the correct relative change, with positive (negative) change being highest in autumn (summer), and also the change in winter being higher than the change in spring (marked italics in Table 3).

The streamflow simulations from the EURO-CORDEX models and all their possible medians (colored lines in Fig. 6) show a wide range of simulated streamflow change between the two periods and within the different seasons. Some models are able to generally reproduce the seasonal change pattern. However, most models predict very high positive streamflow changes from winter to summer and very low or even negative streamflow changes in autumn.

All selection methods—apart from the best streamflow (bSf) and best climate (bCl) methods—seem to constrain the simulations to the ones that are capable of reproducing the seasonal change pattern (visual inspection of Fig. 6). The best performing methods are simple weights (sWgt), MIMR, and diversity of RCMs and GCMs (DivG and DivR) (Table 3). The simple weights (sWgt) method selects model 8 (KNMI-RACMO + ICHEC) as the best-performing model with an RMSE of 38 m³/s. The best streamflow (bSf) method led to model 11 (MPI-REMO + MPIr2), which has the second-worst streamflow change depiction (RMSE of 238 m³/s). Between ranks 4 and 8, the RMSE model performance deteriorates from 91 to 99 m³/s, indicating that the DivG, REA, and Dem methods result in similar errors. Finally, RMSE significantly increases for the remaining methods (bCl and bSf) and models.

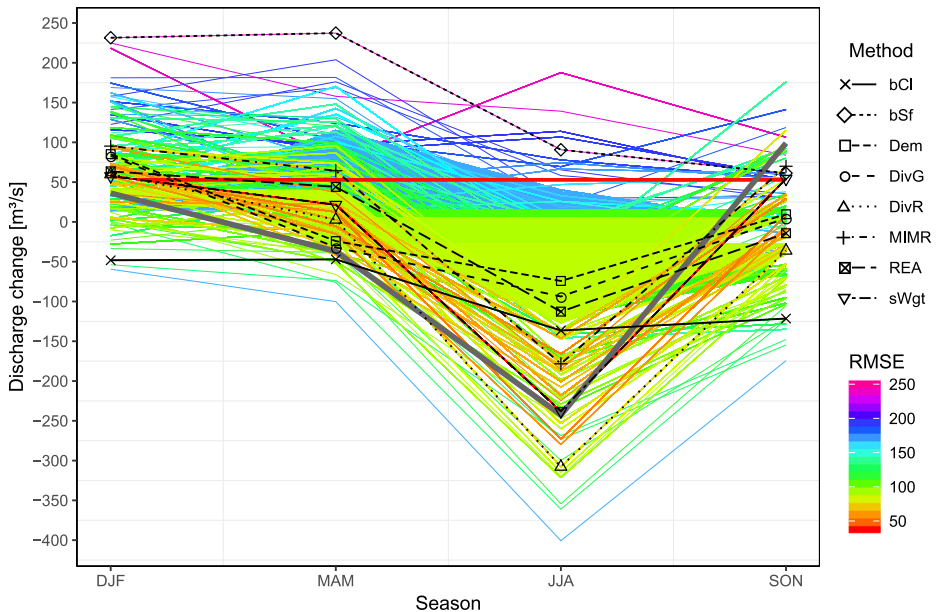


Fig. 6 Change in streamflow of the Danube at Vienna from 1960–1989 to 1990–2014 for simulated streamflow with observed climate (thick gray line), all climate models with all possible medians (colored lines), and the eight sub-selection methods (thin dashed and solid black lines with markers); RMSE values represent the error between the observed and simulated change

5 Discussion

5.1 Inter-comparison of sub-selection methods

We generally show that it is not recommended to select climate models for impact studies based on the depiction of past climate alone. This is also in agreement with expectations from Knutti et al. (2010) concluding that historically best performing models may not be the best for reproducing the climate change signal. On the contrary, Padron et al. (2018) conclude that simulations that spatially agree better with observations of long-term mean precipitation are likely more reliable for future projections. We found that no relation exists between how well the models are able to reproduce past climate and how well they depict the change signal from the reference to the evaluation period (see also Giorgi and Coppola 2010).

We followed a simplified approach to maintain diversity by selecting different GCM and RCM without analyzing common algorithms, boundary conditions, and model assumptions (Abramowitz et al. 2019). However, methods that considered multiple models or maintained the diversity of models (weighing all models based on climate seasonality, diversity of RCM, MIMR) generally performed better than the selection of single models, even if those models reproduce well the indicator of interest (in this case streamflow seasonality).

The concept of respecting model diversity originates from the key assumption that errors level out if models are diverse, and hence a projection can be better represented through averaging, i.e., taking the mean, median, or weighting (Tebaldi and Knutti 2007). This conclusion is also supported by Thober and Samaniego (2014). Pechlivanidis et al. (2018) discuss this issue in more detail and argue that it is important to consider both diversity and

information content in ensemble model sub-selection. In addition, they argue that down-weighting or excluding models that are introducing redundancy (i.e., due to model structure similarities) can lead to robust climate change projections. Results from this paper support these statements considering also that the MIMR method performed second-best overall.

Being able to reproduce the indicator of interest in the past is no guarantee that the models will be able to predict future changes similarly well, as results from the inferior performance of the best streamflow (bSf) method show. Conclusions largely depend on how well past climate change processes, model boundary conditions, and forcing datasets are implemented in the climate models. We believe that the dynamic error between observed and simulated shift in streamflow cannot be attributed to the static biases inherent in the EURO-CORDEX projections (Kotlarski et al. 2014). Our results support this given that (1) biases were adjusted prior to the streamflow simulation and (2) models with higher biases did not perform worse than others.

In addition, we conclude that models that are not able to predict the shift in streamflow seasonality in the past are not particularly well suited to depict an expected, similar or probably aggravated shift in the future. Consequently, the methods evaluated in this study set boundaries to exclude certain models, and favoring others, for impact assessments.

Overall, we recommend splitting historic observations into a reference and evaluation period to understand recent impacts of climate change on the indicator of interest. The observed change in the indicator of interest can then be used to test hindcasted climate data, assess and rank the quality of the climate models, and eventually weigh their contribution for impact assessment, following the suggestion of Eyring et al. (2019).

5.2 Regionalization of methodology beyond case study

The spatiotemporal complexity of the Upper Danube basin allows rigorous testing of the climate models' suitability to force a hydrological model. In this river system, there is a gradient in the hydrological response with both temperature and precipitation strongly affecting streamflow. Consequently, assessing the performance of the climate models is both important and possible. However, we refrain from general recommendations on which methods to use for ensemble sub-selection or which climate models perform best overall due to the single case study and the use of only one hydrological model. It thus remains unclear if and how the ranked sub-selection methods would differ if another (a set of) hydrological model(s) would be used or if a different study area would be chosen. Nevertheless, by selecting a complex basin and carrying out the rigorous hydrological model evaluation, we expect to have minimized these impacts as much as possible. Our approach is simple and suitable to be tested beyond the study region and in multi-basin experiments, but requires the availability of multiple suitable basins with high-quality, long-term datasets and hydrological models to be sure the climate change signal can be distinguished from observation errors, hydrological model uncertainty, and the natural variability. Such data could be, for instance, based on Blöschl et al. (2017), who found that climate change has caused hydrological change across Europe in thousands of basins, or using multi-basin datasets with long-term hydro-meteorological datasets for the USA (Newman et al. 2015; Addor et al. 2017).

In addition to the choice of the climate model selection tested here, our presented evaluation method can be used to test inherent uncertainties, i.e., driven by the choice of the emission scenario (Addor et al. 2014; Samaniego et al. 2017), initial conditions, downscaling (Wagner et al. 2015) and bias-adjustment methods (Teutschbein and Seibert 2013), hydrological model structure and its parameterization (Vetter et al. 2017), and the indicator and region of interest (Clark et al. 2016).

6 Conclusions

This study focuses on the analysis of eight methods for sub-selection of an ensemble of 16 bias-adjusted EURO-CORDEX climate models used in hydrological impact assessments. We test the methods in the Upper Danube basin, where climate change has already affected the streamflow response. To reduce the uncertainty introduced by the hydrological model selection and its parameterization, the model was rigorously evaluated. In particular, we found that:

- (1) The range of changes projected by the different sub-selection methods varies as much as the actual change signal, hence, choosing the best performing sub-selection method matters.
- (2) Of the eight ensemble sub-selection methods tested, those maintaining/maximizing diversity and information content outperformed those that relied on the reproduction of historical climate or streamflow.
- (3) None of the climate models used here was able to correctly reproduce the spatial heterogeneity of historical precipitation and temperature change in the Upper Danube.
- (4) Change in precipitation and temperature impact streamflow through complex hydrological processes. These can hardly be disentangled through looking at precipitation and temperature alone but can be reproduced by a thoroughly evaluated hydrological model that adequately represents the hydrological processes.
- (5) For river systems in which climate change already impacted the hydrological response and sufficient data are available, it is valuable to use this information about historical climate change impact to test climate model performance. For such regions, the proposed methods could provide a possible option to classify and assess the expected boundaries of climate change impact studies for the near and far future.

Finally, we state that future research should focus on testing the proposed methods using multiple hydrological models in multiple basins located under a strong hydro-climatic gradient (e.g., arid regions, monsoon influenced regions, snow-dominated or temperate climates) to further yield conclusions in relevance to sub-selection methods, climate/hydrological model performance, and uncertainty propagation.

Acknowledgments We thank the German Federal Institute of Hydrology (BfG), who funded the hydrological model development and climate change impact assessment, for sharing the data. We also thank the World Climate Research Programme's Working Group on Regional Climate, and the Working Group on Coupled Modeling, former coordinating body of CORDEX and responsible panel for CMIP5 including the climate modeling groups. Climate model data was obtained from the data portal of the Earth System Grid Federation (ESGF) via the node provided by the German Climate Computing Centre (Deutsches Klimarechenzentrum DKRZ).

Data archiving The data that support the findings of this study are not publicly available, but can be obtained from the authors upon reasonable request and with permission of BfG (German Federal Institute for Hydrology).

Authors' contribution JK led and designed the study. JK performed the analysis with essential support of HK and PS who carried out the hydrological simulations and EURO-CORDEX data processing and IP who carried out the MIMR analysis. All authors substantially contributed to the interpretation of the results. JK wrote the initial draft of the manuscript and all co-authors discussed the results and revised the work carefully.

Funding Open Access funding enabled and organized by Projekt DEAL. This study was funded through the “GLANCE” project (Global change effects on river ecosystems; 01LN1320A) supported by the German Federal Ministry of Education and Research (BMBF). This work was also partially funded by the project AQUACLEW, which is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFV (AT), IFD (DK), MINECO (ES), and ANR (FR) with co-funding by the European Commission (grant agreement 690462).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abramowitz G, Herger N, Gutmann E et al (2019) ESD reviews: model dependence in multi-model climate ensembles: weighting, sub-selection and out-of-sample testing. *Earth Syst Dynam* 10:91–105
- Addor N, Rössler O, Köpflin N, Huss M, Weingartner R, Seibert J (2014) Robust changes and sources of uncertainty in the projected hydrological regimes of Swiss catchments. *Water Resour Res* 50: 7541–7562
- Addor N, Newman AJ, Mizukami N, Clark MP (2017) The CAMELS data set: catchment attributes and meteorology for large-sample studies. *Hydrol. Earth Syst Sci* 21:5293–5313
- Angstrom A (1924) Solar and terrestrial radiation. *Quart J Roy Met Soc* 50:121–125
- Auer I et al (2007) HISTALP – historical instrumental climatological surface time series of the greater Alpine region 1760–2003. *Int J Climatol* 27:17–46
- Blöschl G, Hall J, Parajka J et al (2017) Changing climate shifts timing of European floods. *Science* 357:588–590
- Böhm R et al (2009) Eine neue Webseite mit instrumentellen Qualitäts-Klimadaten für den Großraum Alpen zurück bis 1760. In: *Wiener Mitteilungen* Band, 216th edn. Bemessung, Risikoanalyse und Vorhersage, Hochwässer, pp 7–20
- Braun L, Escher-Vetter H (2013) Gletscherforschung am Vernagtferner. *Zeitschrift für Gletscherkunde und Glaziologie*, Band 45/46, 398p
- Christensen JH, Kjellström E, Giorgi F, Lenderink G, Rummukainen M (2010) Weight assignment in regional climate models. *Clim Res* 44:179–194
- Clark MP, Wilby RL, Gutmann ED, Vano JA, Gangopadhyay S, Wood AW, Fowler AJ, Prudhomme C, Arnold JR, Brekke LD (2016) Characterizing uncertainty of the hydrologic impacts of climate change. *Curr Clim Change Rep* 2:55–64. <https://doi.org/10.1007/s40641-016-0034-x>
- Evans JP, Ji F, Abramowitz G, Ekström M (2013) Optimally choosing small ensemble members to produce robust climate simulations. *Environ Res Lett* 8:4p. <https://doi.org/10.1088/1748-9326/8/4/044050>
- Eyring V et al (2019) Taking climate model evaluation to the next level. *Nat Clim Chang*. <https://doi.org/10.1038/s41558-018-0355-y>
- Giorgi F, Coppola E (2010) Does the model regional bias affect the projected regional climate change? An analysis of global model projections. *Clim Chang* 100(3–4):769–815
- Herger N, Angéilil O, Abramowitz G, Donat M, Stone D, Lehmann K (2018) Calibrating climate model ensembles for assessing extremes in a changing climate. *J Geophys Res-Atmos* 123:5988–6004. <https://doi.org/10.1029/2018JD028549>
- Hussain MM, Mahmud I (2019) pyMannKendall: a python package for non parametric Mann Kendall family of trend tests. *J Open Source Softw* 4(39):1556. <https://doi.org/10.21105/joss.01556>
- IPCC (2013) *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) . Cambridge University Press, Cambridge, p 1535
- Jacob D, Petersen J, Eggert B, Alias A et al (2014) EURO-CORDEX: new high-resolution climate change projections for European impact research. *Reg Environ Chang* 14:563–578
- Kendall MG (1955) Rank correlation methods. Griffin, London

- Kiesel J, Gericke A, Rathjens H, Wetzig A, Kakouei K, Jähnig SC, Fohrer N (2019a) Climate change impacts on ecologically relevant hydrological indicators in three catchments in three European ecoregions. *Ecol Eng* 127:404–416
- Kiesel J, Guse B, Bormann H (2019b) Projecting the consequences of climate change on river ecosystems. In: Sabater S, Elosegí A, Ludwig R (eds) *Multiple stressors in river ecosystems, status, impacts and prospects for the future*, chapter 16. Elsevier, Netherlands, pp 281–301
- Kling H, Fürst J, Nachtnebel HP (2007) Seasonal water balance. In: Fuw BML (ed) *Hydrological Atlas of Austria*, 3rd edn. map sheet 7.2, Wien ISBN 3-85437-250-7
- Kling H, Fuchs M, Paulin M (2012a) Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J Hydrol* 424–425:264–277
- Kling H, Lagler M, Stanzel P, Fuchs M (2012b) Wasserbilanzmodellierung und Auswirkungen der Klimaänderung für den deutschen Abschnitt der Donau – Erweiterte Simulationen. *Hydrologische Studie*, BFG, Endbericht, 248p
- Kling H, Stanzel P, Fuchs M, Nachtnebel HP (2015) Performance of the COSERO precipitation–runoff model under non-stationary conditions in basins with different climates. *Hydrol Sci J* 60(7–8):1374–1393. <https://doi.org/10.1080/02626667.2014.959956>
- Knutti R, Furrer R, Tebaldi C, Cernak J, Meehl GA (2010) Challenges in combining projections from multiple climate models. *J Clim* 23:2739–2758. <https://doi.org/10.1175/2009JCLI3361.1>
- Knutti R, Sedláček J, Sanderson BM, Lorenz R, Fischer EM, Eyring V (2017) A climate model projection weighting scheme accounting for performance and interdependence. *Geophys Res Lett* 44:1909–1918. <https://doi.org/10.1002/2016GL072012>
- Kotlarski S, Keuler K, Christensen B, Colette A et al (2014) Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geosci Model Dev* 7:1297–1333
- Krysanova V, Vetter T, Eisner S, Huang S, Pechlivanidis IG, Strauch M et al (2017) Intercomparison of regional-scale hydrological models in the present and future climate for 12 large river basins worldwide - a synthesis. *Environ Res Lett* 12:105002. <https://doi.org/10.1088/1748-9326/aa8359>
- Krysanova V, Donnelly C, Gelfan A, Gerten D, Arheimer B, Hattermann F, Kundzewicz ZB (2018) How the performance of hydrological models relates to credibility of projections under climate change. *Hydrol Sci J* 63(5):696–720
- Li C, Singh VP, Mishra AK (2012) Entropy theory-based criterion for hydrometric network evaluation and design: maximum information minimum redundancy. *Water Resour Res* 48(5):WR011251. <https://doi.org/10.1029/2011WR011251>
- Mann HB (1945) Nonparametric tests against trend. *Econometrica* 13:245–259
- Melsen L, Addor N, Mizukami N, Newman A, Torfs P, Clark M, Uijlenhoet R, Teuling R (2018) Mapping (dis)agreement in hydrological projections. *Hydrol. Earth Syst Sci* 22:1775–1791
- Mote P, Brekke L, Duffy PB, Maurer E (2011) Guidelines for constructing climate scenarios. *EOS Trans Am Geophys Union* 92(31):257–264
- Newman AJ, Clark MP, Sampson K et al (2015) Development of a large-sample watershed-scale hydrometeorological dataset for the contiguous USA: dataset characteristics and assessment of regional variability in hydrologic model performance. *Hydrol Earth Syst Sci* 19:209–223
- Padron RS, Gudmundson L, Seneviratne SI (2018) Observational constraints reduce likelihood of extreme changes in multidecadal land water availability. *Geophys Res Lett* 46:736–744
- Pechlivanidis IG, Arheimer B, Donnelly C, Hundecha Y, Huang S, Aich V et al (2017) Analysis of hydrological extremes at different hydro-climatic regimes under present and future conditions. *Clim Chang* 141(3):467–481
- Pechlivanidis IG, Gupta H, Bosshard T (2018) An information theory approach to identifying a representative subset of hydro-climatic simulations for impact modelling studies. *Water Resour Res* 54:5422–5435
- Ruane AC, McDermid SP (2017) Selection of a representative subset of global climate models that captures the profile of regional changes for integrated climate impacts assessment. *Earth Perspectives* 4:1. <https://doi.org/10.1186/s40322-017-0036-4>
- Samaniego L, Kumar R, Breuer L, Chamorro A, Flörke M, Pechlivanidis IG et al (2017) Propagation of forcing and model uncertainty into hydrological drought characteristics in a multi-model century-long experiment in continental river basins. *Climate Change* 141(3):435–449
- Sanford T, Frumhoff PC, Luers A, Gulleddge J (2014) The climate policy narrative for a dangerously warming world. *Nat Clim Chang* 4:164–166
- Stanzel P, Kling H (2018) From ENSEMBLES to CORDEX: evolving climate change projections for upper Danube River flow. *J Hydrol* 563:987–999
- Tebaldi C, Knutti R (2007) The use of the multi-model ensemble in probabilistic climate projections. *Phil Trans R Soc A* 365:2053–2075. <https://doi.org/10.1098/rsta.2007.2076>
- Teutschbein C, Seibert J (2013) Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions? *Hydrol Earth Syst Sci* 17:5061–5077

- Thober S, Samaniego L (2014) Robust ensemble selection by multivariate evaluation of extreme precipitation and temperature characteristics. *J Geophys Res Atmos* 119:594–613
- Thober S, Kumar R, Wanders N, Marx A, Pan M, Rakovec O et al (2018) Multi-model ensemble projections of European river floods and high flows at 1.5, 2, and 3 degrees global warming. *Environ Res Lett* 13(1):1–11
- Turc L (1961) Estimation of irrigation water requirements, potential evapotranspiration: a simple climatic formula evolved up to date. *Ann Agron* 12(1):13–49
- Vetter T, Reinhardt J, Flörke M, van Griensven A, Hattermann F, Huang S et al (2017) Evaluation of sources of uncertainty in projected hydrological changes under climate change in 12 large-scale river basins. *Clim Chang* 141(3):419–433
- Wagner PD, Reichenau TG, Kumar S, Schneider K (2015) Development of a new downscaling method for hydrologic assessment of climate change impacts in data scarce regions and its application in the Western Ghats, India. *Reg Environ Chang* 15:435–447
- Weber M (2003) Gletscherschwund und Klimawandel an der Zugspitze und am Vernagtferner (Ötztaler Alpen). Informationen zum Gletscherschwund, Kommission für Glaziologie der Bayerischen Akademie der Wissenschaften (KfG), 10 pp.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Jens Kiesel^{1,2} · Philipp Stanzel³ · Harald Kling³ · Nicola Fohrer² · Sonja C. Jähnig¹ · Ilias Pechlivanidis⁴

¹ Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany

² Institute for Natural Resource Conservation, Department of Hydrology and Water Resources Management, Christian-Albrechts-University Kiel, Kiel, Germany

³ AFRY Austria GmbH, Hydro Consulting, Vienna, Austria

⁴ Swedish Meteorological and Hydrological Institute, Norrköping, Sweden