Reply to Reviewer 1

The study presents a statistical analysis of the hydrological cycle in Czechia. To do so the study uses multiple gridded hydrological products, derived using remote sensing and reanalysis. First a ranking scheme regarding the performance of each product and their combination is presented. To me this is the main novelty of the study. Afterwards the best products are analyzed to provide spatially explicit estimates of the change of hydrological dynamics between a past and a present era.

We would like to thank the reviewer for their brief yet insightful comments. We have revised the manuscript based on the reviewer's comments and suggestions. The evaluation data sets will be replaced by high-quality observations only. In addition, we will add new figures to present further results and discuss their implications supporting the hypothesis of re-distribution of terrestrial water, since Czechia is losing water in the long-term (precipitation remains the same while evapotranspiration increases). In the following, we provide detailed replies to all comments and discuss changes to the main manuscript.

Overall, the methodology is mostly solid. My main methodological question concerns the use of GLEAM and GRUN. I am not sure why GLEAM and GRUN were used as benchmark datasets. GLEAM and GRUN are both model based. How can they be used as benchmarks for validation? They themselves carry a lot of uncertainty. For ET, the physical basis of some of the remaining datasets (e.g., ERA5 land) is much more detailed than GLEAM as they integrate a full complexity land surface scheme, rather than simplifying models (e.g., Priestley Taylor). GRUN has even less physical basis, as it is a statistical model. I would be more convinced with the analysis, if only real high-quality observations were included in benchmarking the various datasets.

Initially GLEAM and GRUN were chosen as evaluation benchmarks because both are considered high quality products (E.g., Yang et al. [2017]; Bai and Liu [2018]; Liu et al. [2021]; Hu et al. [2021]; Xiong et al. [2022]; Xu et al. [2022]; Mei et al. [2023]) but their record lengths were not long enough to be part of the main analysis. As correctly pointed out by the reviewer, these data sets do carry considerable uncertainty. Therefore, in order to include only high-quality observational data and to evince the robustness of the ranking method proposed we decided to replace GRUN by GRDC for runoff and perform the ranking without an evapotranspiration reference for evaluation. Note that we selected only three stations from GRDC, namely the Bohumin (Oder), Decin (Elbe), and Moravsky Jan (Danube) stations, which are placed near the borders of the country and their wieighted average was computed using the catchment area as registered by GRDC. The revised benchmarking (revised Figure 2) and top ranking results vary only slightly (revised Table 2), further supporting our initial choice of referential data sets.

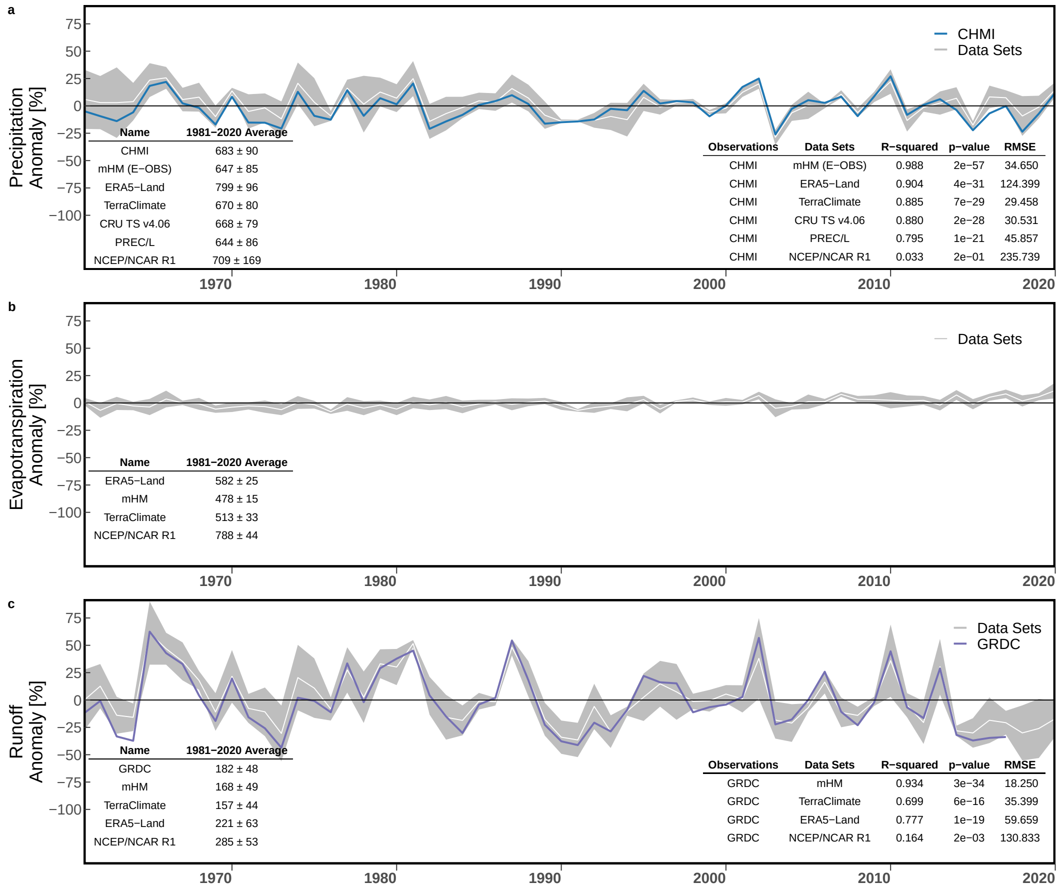

Figure 2. Benchmarking spatial weighted average annual water fluxes over Czechia between 1961 and 2020. For consistency and comparability between different water fluxes, annual anomalies were computed using the 1981-2010 average as a reference, the common period among all data sets. The 1981-2010 average and standard deviation are listed at the bottom left of each panel. Linear correlation summary statistics are displayed at the bottom right of each panel. The spread of the estimates being evaluated is shown in gray, and their mean is in white. (a) Precipitation evaluation. CHMI data is shown in blue. (b) Evapotranspiration evaluation. (c) Runoff evaluation. GRDC (Bohumin, Decin, and Moravsky Jan stations) data is shown in purple.

Table 2. Data set ranking as determined by Equation 3. P is precipitation, E is evapotranspiration, Q is runoff, $\bar{\xi}$ is the mean residual over 60 years, σ_ξ is the standard deviation of the residual over 60 years, $cor(P-E,Q)$ is the correlation between $P-E$ and Q for the i-th ranked combination, $cor(P,P_o)$ is the correlation between P of the i-th ranked combination and CHMI, and $cor(Q,Q_o)$ is the correlation between Q of the i-th ranked combination and GRDC.

Ranking	P	E	Q	$\bar{\xi}$	σ_ξ	$cor(P-E,Q)$	$cor(P,P_o)$	$cor(Q,Q_o)$
1st	TerraClimate	TerraClimate	TerraClimate	-0.346	30.204	0.846	0.941	0.836
2nd	mHM(E-OBS)	mHM	mHM	-0.912	51.231	0.816	0.994	0.967
3rd	CRU TS v4.06	TerraClimate	TerraClimate	-1.749	29.944	0.843	0.938	0.836
4th	TerraClimate	TerraClimate	mHM	-8.861	39.847	0.730	0.941	0.967
5th	CRU TS v4.06	TerraClimate	mHM	-10.265	40.613	0.711	0.938	0.967
6th	ERA5-Land	ERA5-Land	ERA5-Land	-5.554	66.606	0.701	0.951	0.882
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
14th	PRECL/L	mHM	TerraClimate	17.013	60.281	0.658	0.891	0.836
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
24th	ERA5-Land	TerraClimate	mHM	114.628	44.721	0.763	0.951	0.967
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
38th	ERA5-Land	NCEP/NCAR R1	mHM	-166.746	60.420	0.714	0.951	0.967
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
48th	PREC/L	mHM	ERA5-Land	-52.549	82.751	0.382	0.891	0.882
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
72nd	mHM(E-OBS)	mHM	NCEP/NCAR R1	-134.044	87.923	0.237	0.994	0.405
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
87th	NCEP/NCAR R1	NCEP/NCAR R1	NCEP/NCAR R1	-292.024	137.297	0.675	0.181	0.405
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
96th	CRU TS v4.06	NCEP/NCAR R1	NCEP/NCAR R1	-424.772	93.962	-0.019	0.938	0.405

Apart from that, a thorough analysis is presented, which to a large extent is consistent with previous results related to continental Europe. Even though the study is methodologically sound, its novelty is limited in my opinion because of (a) the data products used are all well established and have been extensively previously analyzed at regional and global scales, and (b) the limited geographical extent of the study. I find the paper better suited to journals focusing on regional studies, rather than HESS whose goal is to further advance the fundamental understanding of hydrological processes and their impacts on society and ecosystems.

We thank the reviewer for this comment because it helped us realize that the novelties of our study have not been properly highlighted. Although the data products have been previously analyzed at regional or global scales, this is done under a univariate perspective, that does not consider the ability of the data sets to reproduce the water cycle (and its changes) as a whole in a structurally plausible manner. This comment pushed us to look deeper into the water budget closure, where it became evident that there is a substantial overestimation of the drying in ERA5-Land (Figure 3).

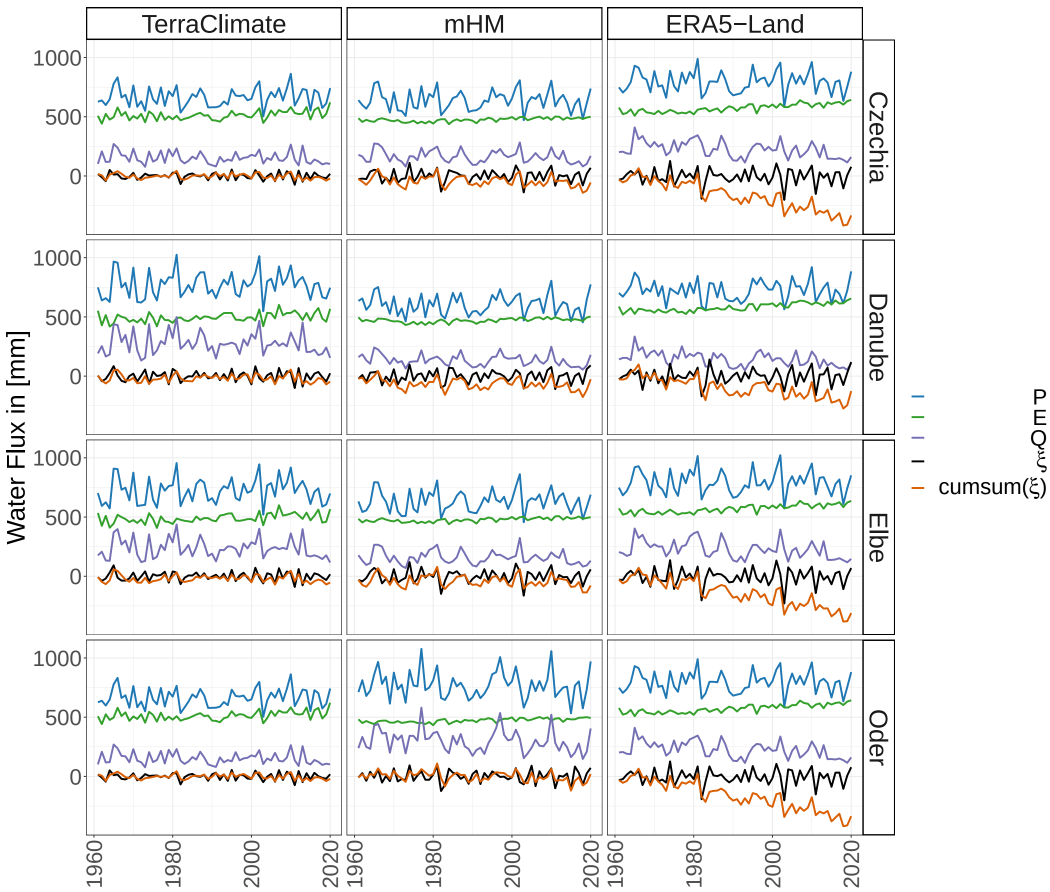

Figure 3. Spatial weighted average annual water fluxes over Czechia (first row), Danube basin inside Czechia (second row), Elbe basin inside Czechia (third row), and Oder basin inside Czechia (fourth row). Where P is precipitation in blue, E is evapotranspiration in green, Q is runoff in purple, ξ is the residual ($P - E - Q$) in black, and cumsum(ξ) is the cumulative sum of the residual in orange. Left column: TerraClimate (P), TerraClimate (E), and TerraClimate (Q). Middle column: mHM(E-OBS) (P), mHM (E), and mHM (Q). Right column: ERA5-Land (P), ERA5-Land (E), and ERA5-Land (Q).

We acknowledge that the geographical extent of the study is small. Nonetheless, a relatively small study domain is not uncommon at HESS, as demonstrated by some of the work we cited [Jenicek and Ledvinka, 2020; Muelchi et al., 2021]. The former discusses the influence of snow storage and snowmelt inter-annual variations effect on seasonal runoff and summer low flows in *Czechia*. The latter addresses projected changes in river runoff regimes in *Switzerland*. Furthermore, there are multiple other publications in recent years with similar geographical extents, some of which are:

- Osuch et al. [2016] Reported possible climate change effects on dryness by assessing the standardized precipitation index on multiple climate projections in *Poland*.
- Silvestro et al. [2018] Analyzed streamflow extremes and long-term water balance in the *Liguria region of Italy*.
- Girons Lopez et al. [2021] Benchmarked the SHYPE operational hydrological model in *Sweden*.
- Hanus et al. [2021] Reported changes in runoff signatures at multiple scales in contrasting Alpine catchments in *Austria*.
- Torelló-Sentelles and Franzke [2022] Presented a random forest model to predict drought impacts in *Spain*.
- Alexopoulos et al. [2023] Evaluated precipitation reanalyses performance for rainfall-runoff modeling using the GR4H model in *Slovenia*

Despite their regional geographical extent, the findings of the above-mentioned have implications to our understanding about the hydrological processes. Likewise, in the revised version we will highlight the main novelties of our study, which is the importance of combining data sources that describe all the components of the terrestrial water cycle and presenting a showcase of inconsistencies that might not be visible if the single components are evaluated as performance metrics (the case of ERA5-Land).

A few minor comments:

Lines 17-20: Not clear what the contradiction is between the 2 statements

For clarity and brevity the text will be rephrased from: ”On the one hand, small changes in total precipitation suggest a shift in precipitation towards more intense and less frequent events [Trenberth, 2011]. On the other hand, it was hypothesized that an increased vertical gradient of atmospheric water vapor would offset atmospheric wind convergence in the tropics making wet regions wetter and dry regions drier [Held and Soden, 2006].”

To: ”It was hypothesized that an increased vertical gradient of atmospheric water vapor would offset atmospheric wind convergence in the tropics making wet regions wetter and dry regions drier [Held and Soden, 2006].”

Line 26: define what you mean by unquantified uncertainties

For clarification the text will be rephrased from: ”... unquantified uncertainties on satellite-based products [Sheffield et al., 2009].”

To: "... unquantified uncertainties on satellite-based products [e.g., the impact of cloud filtering; Povey and Grainger, 2015]."

Line 73: What is the meaning of the roof analogy?

For clarity and brevity, we will remove the roof analogy. Which was meant to be a literary figure for a headwaters region (water falling on top primarily runs away rather than staying in).

Figure 1: the different shading is not clear. I suggest the authors to add in bold colors the catchment boundaries for clarity

Figure 1 will be updated as suggested.

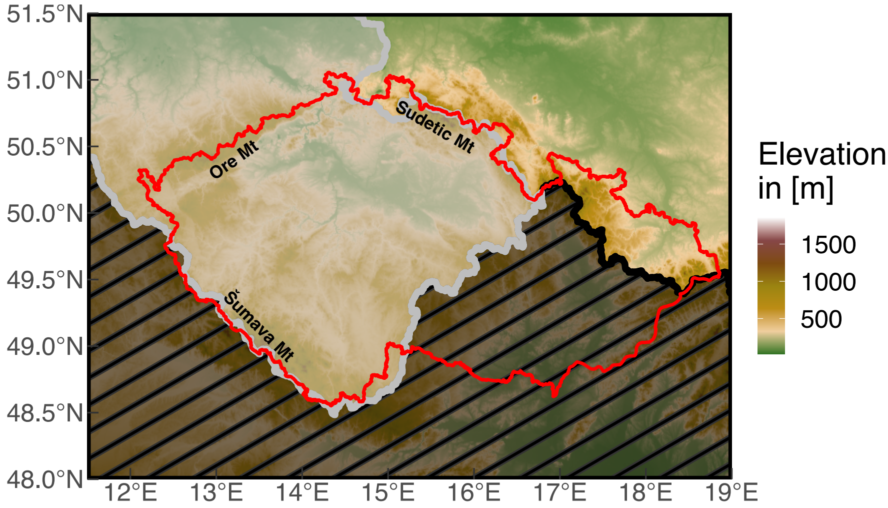

Figure 1. The three drainage basins within Czechia's boundaries. Elbe (light gray shade), Danube (striped dark gray shade), and Oder (no shade).

Line 130: System instead of set of ODEs

Text will be replaced.

I find the definitions of R2 and RMSE a bit redundant.

The definitions will be removed.

In eq 1, 2 I suggest changing the variable name of the residual term from R to something different, e.g. epsilon, to not confuse the reader as R is commonly used for runoff, and previously in the paper as the coefficient of determination

To avoid confusion the variable name will be changed from R to ξ.

Line 211: Why were the authors surprised by the quality of ERA5-Land. Please explain further this statement? The land surface scheme of ERA5-Land (H-TESSEL) has a hydrological component, which is in compatible complexity with the remaining hydrological models of the study.

To explicitly refer to the cause of surprise the text will be rephrased from : "Notwithstanding, we were surprised to see the ERA5-Land exclusive combination (i.e., all flux estimates from the same data set) among the top five ranks."

To: "Notwithstanding, we were surprised to see the ERA5-Land exclusive combination (i.e., all flux estimates from the same data set) among the top six ranks despite non steady water budget residuals (Figure 3) as well as biases 1.7-3.3 and 3.8-4.2 times larger than those of models for runoff (Figure 2c) and precipitation (Figure 2a), respectively" As the previous sentence states: "We expected combinations with hydrological model data to be highly ranked and reanalyses to be poorly ranked due to the above-reported considerable biases of the latter."

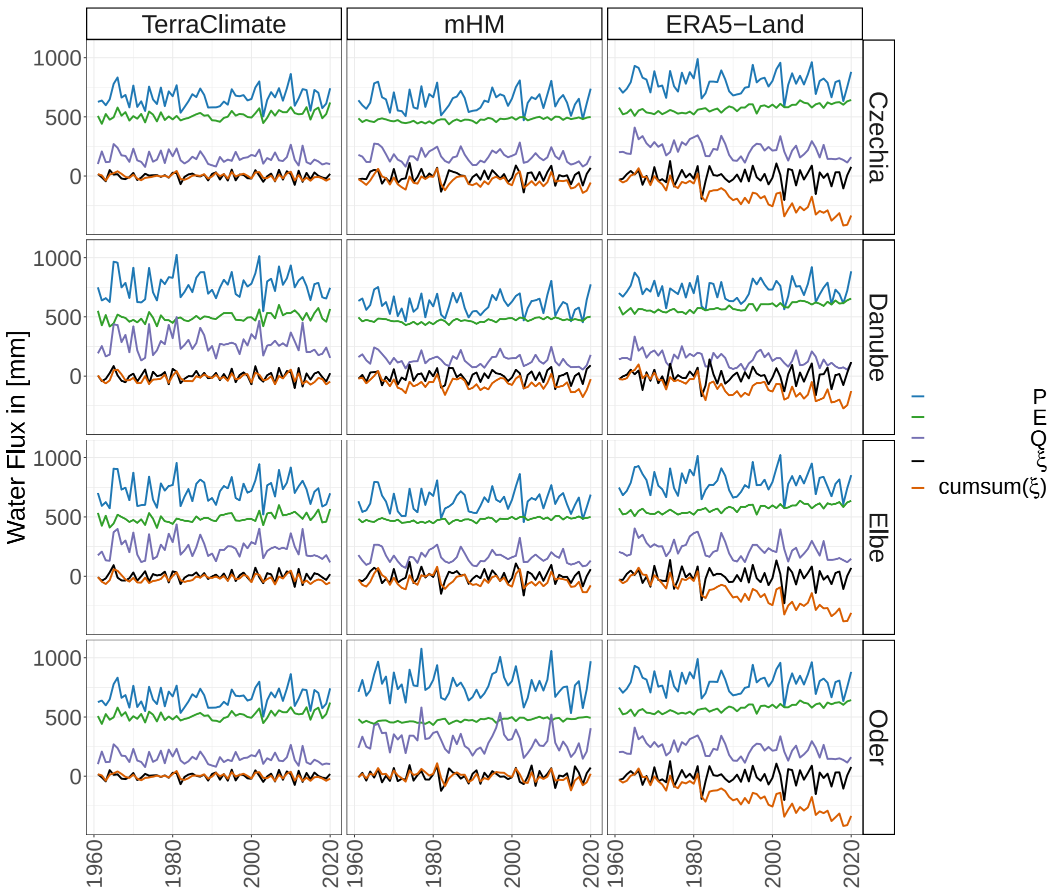

Figure 3. Spatial weighted average annual water fluxes over Czechia (first row), Danube basin inside Czechia (second row), Elbe basin inside Czechia (third row), and Oder basin inside Czechia (fourth row). Where P is precipitation in blue, E is evapotranspiration in green, Q is runoff in purple, ξ is the residual ($P - E - Q$) in black, and cumsum(ξ) is the cumulative sum of the residual in orange. Left column: TerraClimate (P), TerraClimate (E), and TerraClimate (Q). Middle column: mHM(E-OBS) (P), mHM (E), and mHM (Q). Right column: ERA5-Land (P), ERA5-Land (E), and ERA5-Land (Q).

Line 225-228: Does this imply that the models do not close the water balance, or that the integration periods are not long enough, and the discrepancies are due to soil water storage dynamics?

We took 30-year periods, the minimum required to calculate a climate normal, and it would be safe to assume negligible change in water storage. Which is supported by the stationary time series seemingly centered around zero (Figure 3). Moreover, we cannot assert that models do not close the water balance because the discrepancies are considerably small compared to the values of those fluxes.

Line 244: Change Abril to April

Text will be changed.

Figure 4: Might be better if presented as cumulative distribution functions, q-q plots or boxplots.

Figure 4 will be revised from a histogram to a boxplot. Please note that the revised figure numbering is now Figure 5 due to the newly added Figure 3.

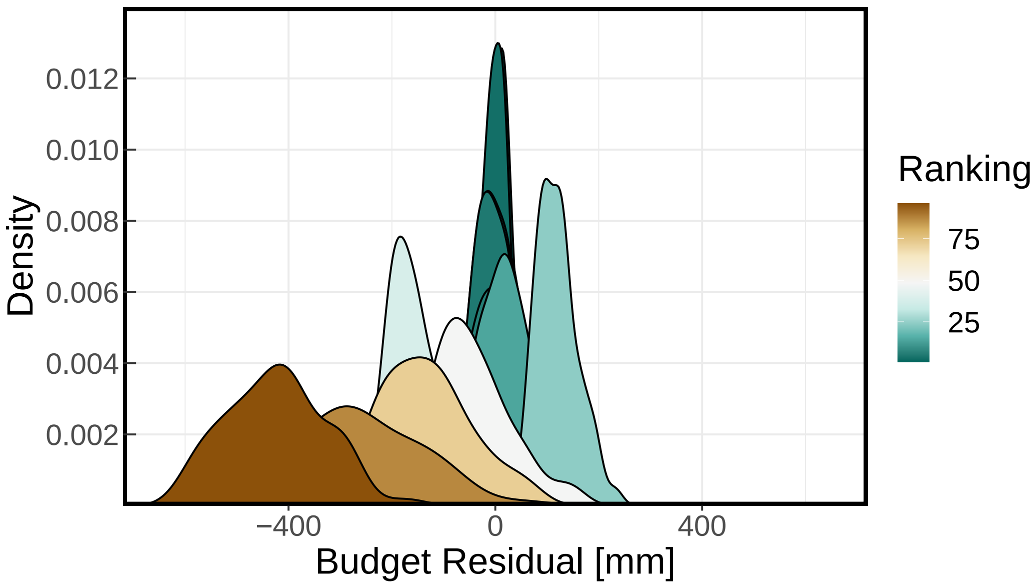

Figure 5. Box plots of spatial weighted average annual water fluxes over Czechia, where P is precipitation, E is evapotranspiration, Q is runoff, and $P - E$ is precipitation minus evapotranspiration. Data are divided into two 30-year periods: 1961-1990 (blue) and 1991-2020 (yellow). Note that outliers are present only in the latter period (i.e., 1991-2020) as expected from the recorded severe drought of 2003.

References

Alexopoulos MJ, Müller-Thomy H, Nistahl P, Šraj M, Bezak N (2023) Validation of precipitation reanalysis products for rainfall-runoff modelling in Slovenia. Hydrology and Earth System Sciences 27(13):2559–2578, DOI 10.5194/hess-27-2559-2023, publisher: Copernicus GmbH

Bai P, Liu X (2018) Intercomparison and evaluation of three global high-resolution evapotranspiration products across China. Journal of Hydrology 566:743–755, DOI 10.1016/j.jhydrol.2018.09.065

Girons Lopez M, Crochemore L, Pechlivanidis IG (2021) Benchmarking an operational hydrological model for providing seasonal forecasts in Sweden. Hydrology and Earth System Sciences 25(3):1189–1209, DOI 10.5194/hess-25-1189-2021, publisher: Copernicus GmbH

Hanus S, Hrachowitz M, Zekollari H, Schoups G, Vizcaino M, Kaitna R (2021) Future changes in annual, seasonal and monthly runoff signatures in contrasting Alpine catchments in Austria. Hydrology and Earth System Sciences 25(6):3429–3453, DOI 10.5194/hess-25-3429-2021, publisher: Copernicus GmbH

Held IM, Soden BJ (2006) Robust Responses of the Hydrological Cycle to Global Warming. Journal of Climate 19(21):5686–5699, DOI 10.1175/JCLI3990.1, publisher: American Meteorological Society Section: Journal of Climate

Hu Y, Duan W, Chen Y, Zou S, Kayumba PM, Sahu N (2021) An integrated assessment of runoff dynamics in the Amu Darya River Basin: Confronting climate change and multiple human activities, 1960–2017. Journal of Hydrology 603:126905, DOI 10.1016/j.jhydrol.2021.126905

Jenicek M, Ledvinka O (2020) Importance of snowmelt contribution to seasonal runoff and summer low flows in Czechia. Hydrology and Earth System Sciences 24(7):3475–3491, DOI 10.5194/hess-24-3475-2020, publisher: Copernicus GmbH

Liu J, Zhang J, Kong D, Feng X, Feng S, Xiao M (2021) Contributions of Anthropogenic Forcings to Evapotranspiration Changes Over 1980–2020 Using GLEAM and CMIP6 Simulations. Journal of Geophysical Research: Atmospheres 126(22):e2021JD035367, DOI 10.1029/2021JD035367, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021JD035367

Mei Y, Mai J, Do HX, Gronewold A, Reeves H, Eberts S, Niswonger R, Regan RS, Hunt RJ (2023) Can Hydrological Models Benefit From Using Global Soil Moisture, Evapotranspiration, and Runoff Products as Calibration Targets? Water Resources Research 59(2):e2022WR032064, DOI 10.1029/2022WR032064, URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2022WR032064, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2022WR032064

Muelchi R, Rössler O, Schwanbeck J, Weingartner R, Martius O (2021) River runoff in Switzerland in a changing climate – runoff regime changes and their time of emergence. Hydrology and Earth System Sciences 25(6):3071–3086, DOI 10.5194/hess-25-3071-2021, URL https://hess.copernicus.org/articles/25/3071/2021/, publisher: Copernicus GmbH

Osuch M, Romanowicz RJ, Lawrence D, Wong WK (2016) Trends in projections of standardized precipitation indices in a future climate in Poland. Hydrology and Earth System Sciences 20(5):1947–1969, DOI 10.5194/hess-20-1947-2016, publisher: Copernicus GmbH

Povey AC, Grainger RG (2015) Known and unknown unknowns: uncertainty estimation in satellite remote sensing. Atmospheric Measurement Techniques 8(11):4699–4718, DOI 10.5194/amt-8-4699-2015, publisher: Copernicus GmbH

Sheffield J, Ferguson CR, Troy TJ, Wood EF, McCabe MF (2009) Closing the terrestrial water budget from satellite remote sensing. Geophysical Research Letters 36(7), publisher: Wiley Online Library

Silvestro F, Parodi A, Campo L, Ferraris L (2018) Analysis of the streamflow extremes and long-term water balance in the Liguria region of Italy using a cloud-permitting grid spacing reanalysis dataset. Hydrology and Earth System Sciences 22(10):5403–5426, DOI 10.5194/hess-22-5403-2018, publisher: Copernicus GmbH

Torelló-Sentelles H, Franzke CLE (2022) Drought impact links to meteorological drought indicators and predictability in Spain. Hydrology and Earth System Sciences 26(7):1821–1844, DOI 10.5194/hess-26-1821-2022, publisher: Copernicus GmbH

Trenberth KE (2011) Changes in precipitation with climate change. Climate Research 47(1-2):123–138

Xiong J, Yin J, Guo S, He S, Chen J, Abhishek (2022) Annual runoff coefficient variation in a changing environment: a global perspective. Environmental Research Letters 17(6):064006, DOI 10.1088/1748-9326/ac62ad, URL https://dx.doi.org/10.1088/1748-9326/ac62ad, publisher: IOP Publishing

Xu D, Bisht G, Sargsyan K, Liao C, Leung LR (2022) Using a surrogate-assisted Bayesian framework to calibrate the runoff-generation scheme in the Energy Exascale Earth System Model (E3SM) v1. Geoscientific Model Development 15(12):5021–5043, DOI 10.5194/gmd-15-5021-2022, publisher: Copernicus GmbH

Yang X, Yong B, Ren L, Zhang Y, Long D (2017) Multi-scale validation of GLEAM evapotranspiration products over China via ChinaFLUX ET measurements. International Journal of Remote Sensing 38(20):5688–5709, DOI 10.1080/01431161.2017.1346400, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01431161.2017.1346400