Reply to Reviewers

Thank you for sending your revised manuscript, which has been read by myself and by two of the same original reviewers. Both reviewers agree the significance and presentation of the manuscript are good and that it should be accepted with some small revisions. This aligns with my own reading of the work, and I would like to invite you to kindly make these revisions. Additionally, please could you add a header at the top of your Supplementary materials with the manuscript details.

We thank the reviewers and the editor for providing further constructive comments and suggestions. We have revised the manuscript accordingly and added a header at the top of the Supplementary material. In the following, we provide detailed replies to all comments and discuss changes to the main manuscript.

Reviewer 2

We thank the reviewer for their detailed comments.

Minor revision

1. Figure 1 caption, what is the meaning of clear for Oder, the red borders?

Figure 1 and its caption were revised to improve its clarity.

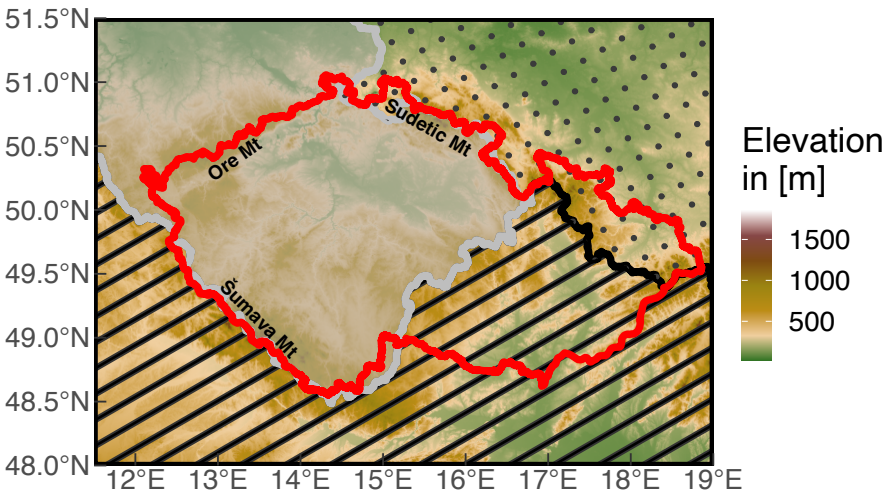

Figure 1. The three drainage basins within Czechia's administrative boundaries (red line). Elbe (light gray shade), Danube (black stripes), and Oder (dark gray points).

2. Line 90-103: The authors introduce CHMI and GRDC as assessment references, can you briefly describe in a sentence or two why these two are more appropriate as references than other datasets in section 2.2.1? Have previous studies been compared or evaluated between these datasets presented in this paper?

To further support the selection of CHMI and GRDC as appropriate references rather than any of the remaining data sets included in our study, the following text was added at the beginning of Section 2.2.1:

”As evaluation references, we relied solely on ground station data sets. A distinct advantage of station data over hydrological models or reanalyses is their capability to capture detailed and localized information. These in-situ measurements directly reflect the local climatic conditions, offering a more accurate representation of the water cycle.”

To the best of our knowledge there are no previous studies comparing all these data sets simultaneously nor in the context of our study.

3. Figure 2: it might be clearer to zoom in on the y-axis in Figure 2b.

The y-axis was modified as suggested:

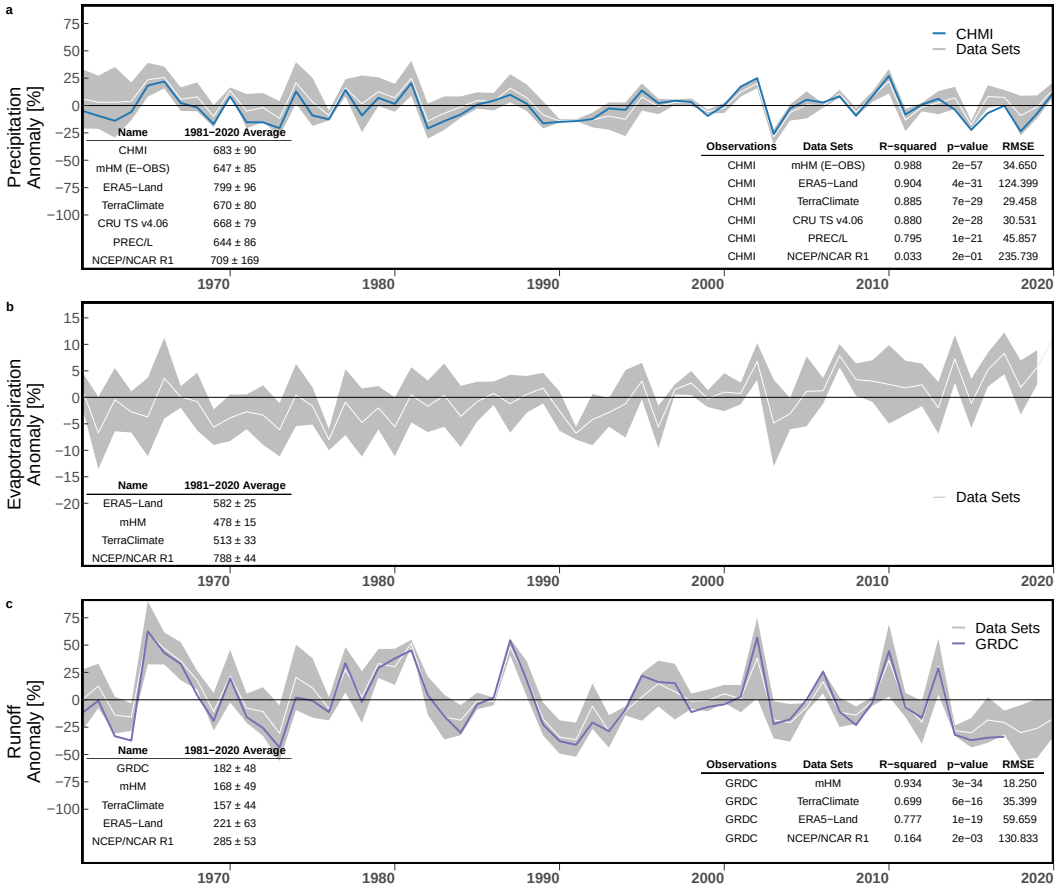

Figure 2. Benchmarking spatial weighted average annual water fluxes over Czechia between 1961 and 2020. For consistency and comparability between different water fluxes, annual anomalies were computed using the 1981-2010 average as a reference, the common period among all data sets. The 1981-2010 average 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 data is shown in purple.

4. Enlarge the font and legend in the figure 7 and figure 8.

Font an legend of Figure 7 and 8 were enlarged (See also minor comment 6).

5. When the authors discuss which combination is the most superior, have they considered that the assessment of different combinations will be different due to the different resolutions of datasets, for example, when calculating water balance if different datasets are used for individual water fluxes can lead to inaccurate results due to different resolutions. Has this bias been considered when weighing the final ranking?

The metric score for ranking the different data set combinations is computed over single time series per flux and not over single grid cells. We mention that it might be a factor for poor performance in L336-337: "Additionally, the poor performance of NCEP/NCAR R1 might be rooted in its coarse spatial resolution (two grid cells cover Czechia).". However, we do not consider the ranking score to be significantly sensitive to different resolutions because we can see that two combinations with CRU TS v4.06 (1°) have higher ranking than ERA5-Land (0.1°) or PREC/L (0.5°).

In addition, for clarity the text was rephrased from: "where P_n is precipitation, E_n is evapotranspiration, and Q_n is runoff for a given year n. Thus, we have 60 annual values for each of the 96 possible combinations. Under steady state conditions the mean of these residuals should tend to zero:"

To: "where P_n is precipitation, E_n is evapotranspiration, and Q_n is runoff for a given year n. Thus, we have 60 annual values for each of the 96 possible combinations. Note that, the water flux time series used to compute the residuals are the spatial weighted average values. Under steady state conditions the mean of these residuals should tend to zero:"

6. For figures 7 and 8, is it possible to add an additional line to show the change in water balance closure (P-E-Q)? Or just a quick additional graph for checking water balance closure.

Figure 7 and 8 were revised as suggested:

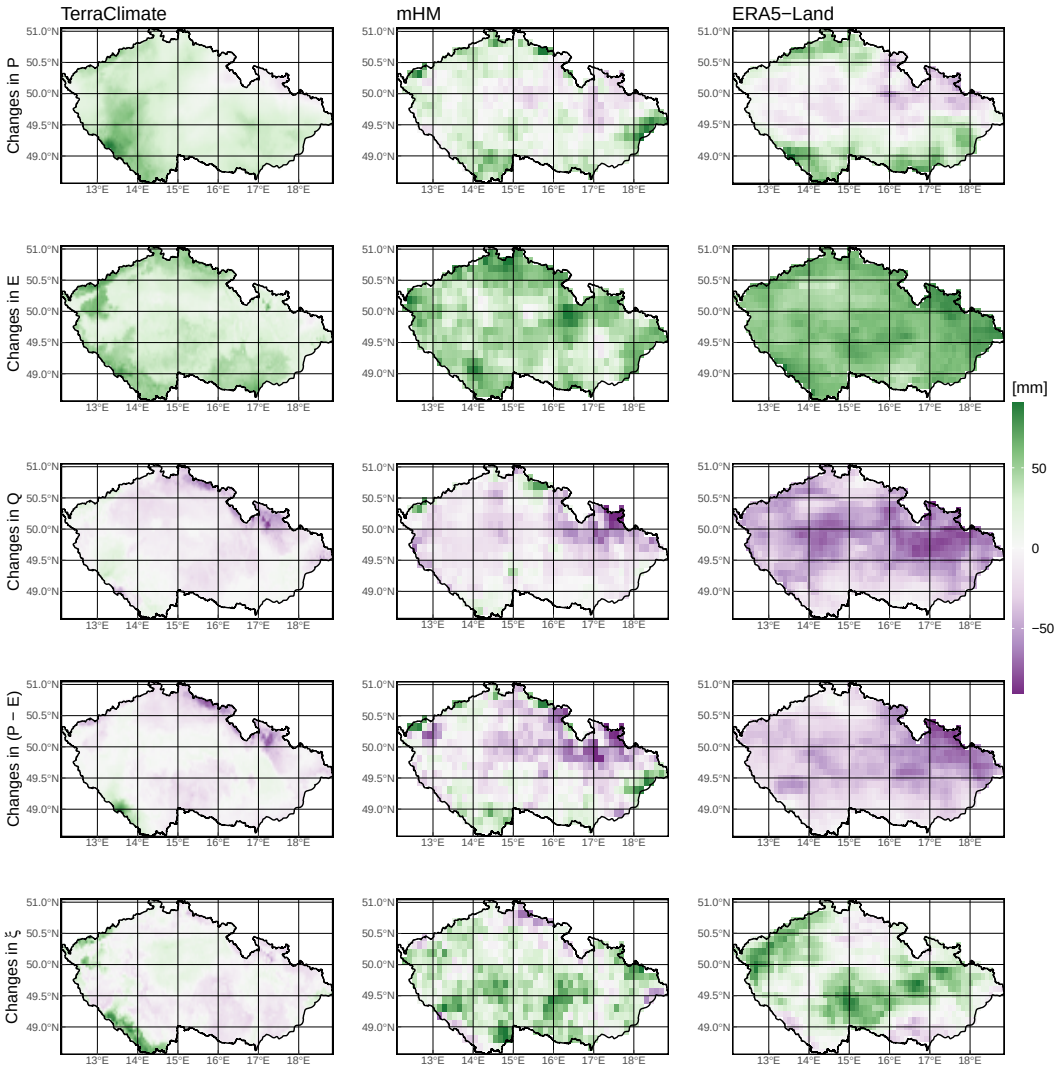

Figure 7. Spatial pattern of changes in median water fluxes over Czechia between two 30-year periods: 1961-1990 and 1991-2020. I.e., the value of each grid cell is equal to the median value of 1991-2020 minus the median value of 1961-1990. P is precipitation, E is evapotranspiration, Q is runoff, $P - E$ is precipitation minus evapotranspiration, and ξ is the residual ($P - E - Q$). 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).

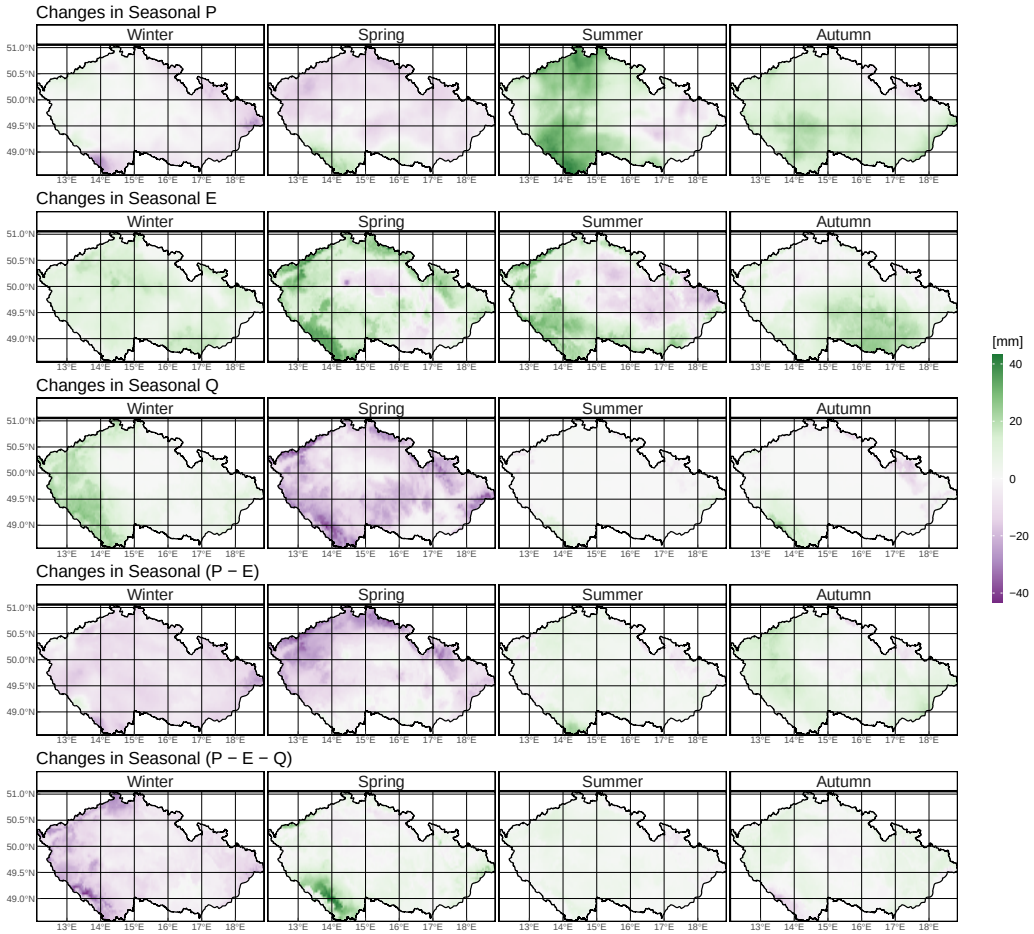

Figure 8. TerraClimate spatial pattern of changes in seasonal median water fluxes over Czechia between two 30-year periods: 1961-1990 and 1991-2020. I.e., the value of each grid cell is equal to the seasonal median value of 1991-2020 minus the seasonal median value of 1961-1990. P is precipitation, E is evapotranspiration, and Q is runoff. The seasons are defined as follows: winter as December, January, and February; spring as March, April, and May; summer as June, July, and August; autumn as September, October, and November.

Reviewer 3

The authors have made a number of revisions based on the comments from all three reviewers and I am satisfied that the manuscript is a useful contribution to understanding the implications of different data source choices on estimating the water budget.

We thank the reviewer for agreeing with the revised manuscript.