Reviewer #1:

“This manuscript presents a novel probabilistic water balance data fusion approach for calibrating multi-scale hydrological datasets. The methodology is innovative, addressing the challenge of reducing uncertainties in datasets by integrating them through water balance constraints. The approach provides a framework for both basin-scale and pixel-scale applications. The application to the Hindon River Basin demonstrates practical utility, with reasonable error estimates and clear improvements in data consistency. The paper is well-written, structured, and accessible, making a substantial contribution to water resource management and hydrological modeling. However, some areas, such as the clarity of methodological details and validation against independent data, could be strengthened to enhance the robustness and reproducibility of the findings. Suggestions are as follows:”

We would like to thank the referee for the time and effort reviewing our manuscript, and for the valuable feedback received. We are pleased that the reviewer found the paper well written and appreciate the recognition of the novelty of the presented methods. In the following, we address the reviewer’s detailed comments.

Reviewer comment 1:
“In Section 2, beginning on line 117, you describe the Hindon Basin and the separation of two irrigation seasons (Kharif and Rabi), yet it is unclear if the rotated crops use the same land or if they are in adjacent regions. It would be helpful to add a sentence or two clarifying this.”

Reply on comment 1:
We will add the following sentence at the end of section 2 to mention that the distributaries take off from main canals to serve fixed command areas irrigated year-round, with crops rotated between Kharif and Rabi crops:
“Irrigation water is diverted from the canals to supply the basin through off-takes that serve fixed command areas with crops rotated between Rabi or Kharif crops.”

Reviewer comment 2:
In your results, the validation could be strengthened. Are you able to compare your estimates against any in-situ records? Reported standard errors are useful, but which component dominates the uncertainty (precip, evaporation, storage, discharge, canal imports)? Standard errors are provided but there is no discussion of comparisons with independent ground-truth data or other datasets not used in calibration. Including such validation would enhance confidence in the results.

Reviewer comment 3:
Discussion would benefit from a short explanation on generalization. For example, can this approach work in snow dominated or urban catchments or is it basin specific?

Reply on comments 2 and 3:
Indeed, it will be useful to report the order of the standard errors of the different water balance variables, and it might be more interesting to highlight this at the start of the paper. Therefore, we will edit the abstract to reflect the order:

“An application to the irrigated Hindon River basin in India illustrates that the approach generates physically reasonable estimates of all basin-scale variables, with average standard errors decreasing in the following order: 21 mm month^{-1} for storage, 10 mm month^{-1} for evaporation,7 mm month^{-1} for precipitation, 4 mm month^{-1} for irrigation canal water imports, and 2 mm month^{-1} for river discharge".

We also agree that it's good to reflect on model generalizability and validation; therefore, we will add the following separate section:

"6.4. Extensions

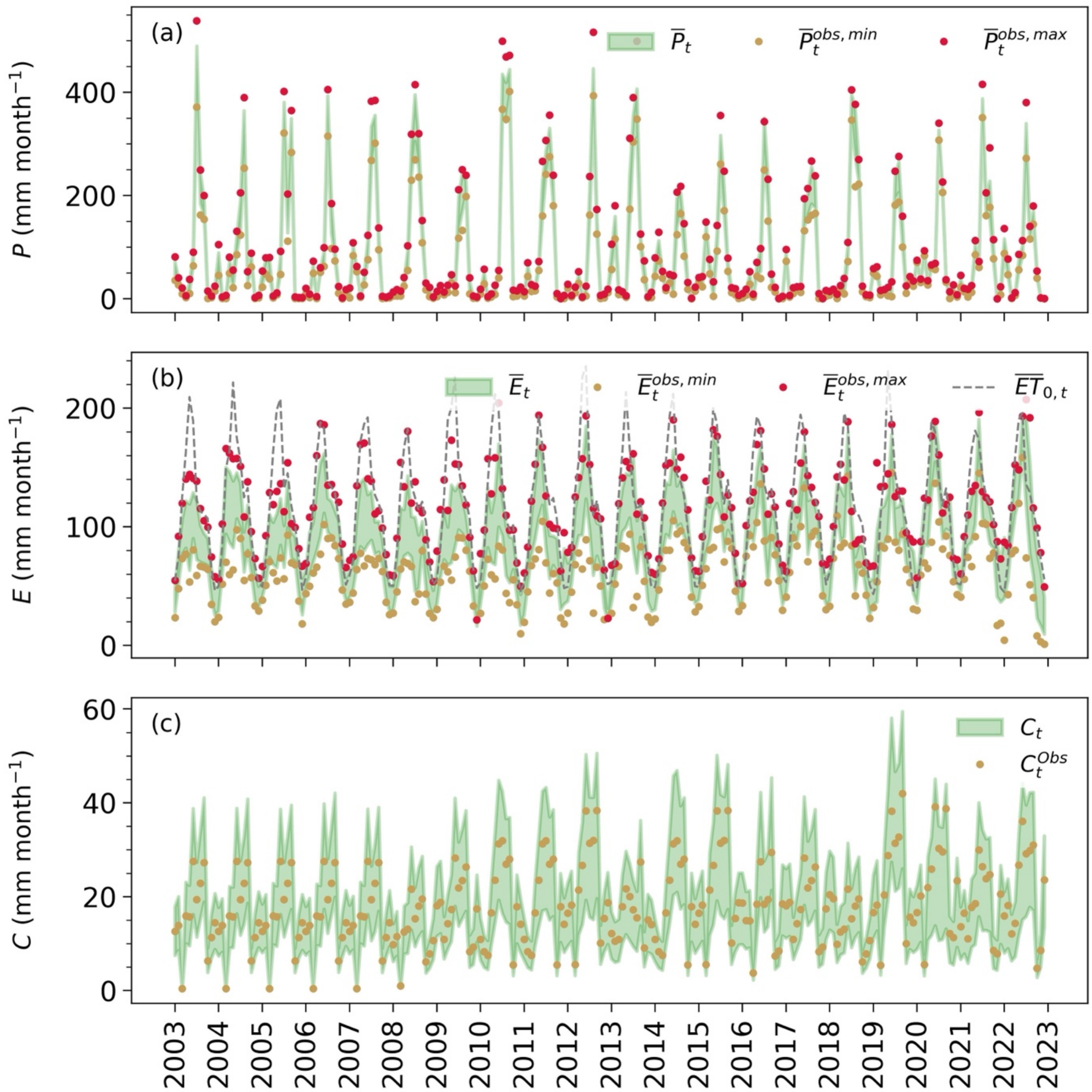
The presented methodology is motivated by the availability of diverse water balance remote sensing data, with very few in situ data available for the Hindon basin, making independent validation of our estimates challenging. For this reason, we evaluated our results using soft validation techniques. For example, for evaporation, in the absence of in situ data, we compared the evaporation posterior estimates with reference evapotranspiration and local irrigation practices (see Sect. 5.1). As for the precipitation posterior estimates, in Sect. 5.1 and Appendix C, these were compared to the spatially interpolated rain gauge dataset for the basin from the Indian Meteorological Department (IMD), keeping in mind the potential underestimation of precipitation by this dataset (Goteti and Famiglietti, 2024). Evaluating the total water storage estimates with independent data is more challenging. In a follow-up study, we will introduce separate rootzone and groundwater balance constraints, with the aim of estimating their contributions to the total water balance. At that point, it will become possible to use available remote sensing soil moisture data, as well as in-situ groundwater level data, for evaluation. Furthermore, adding these additional constraints and data will allow for updating the posterior estimates in this paper.

Additionally, the presented methodology is general and can be applied to other gauged river basins. For example, an application to multiple semi-arid basins was reported in Schoups and Nasseri (2021). The method has several advantages that allow this, including the straightforward and flexible implementation, as it consists of two separate parts: an error model specification that can be customized to fit specific settings, and the model solver that automatically computes the posterior distributions. In addition, the method is set up to rely on in situ and satellite data that inherently capture the hydrological processes. For example, the precipitation data sets estimate the precipitation phase (snow and rain), while evaporation data sets can be used to differentiate between the different land use classes. However, in snow-dominant basins where precipitation data sets might underrepresent this process, or in urban-dominant settings where coarse-resolution evaporation data products might overestimate evaporation, it may be valuable to tailor the error models to local conditions in order to improve the results. Perhaps, this could be achieved by complementing the precipitation error models with other satellite data sets, like the snow cover, snow depth, or temperature products, for better snow detection and mapping. Moreover, evaporation error models can be complemented with land use maps while also considering the use of high-resolution evaporation products or other data sets with improved evaporation estimates of heterogeneous urban surfaces. The major assumption here is that formulating the error models by exploiting ancillary information would allow it to solve for error parameters and water balance variables under varying climatic zones and settings. Alternatively, for this purpose, we could combine the data error models with hydrological models that explicitly account for detailed processes and differentiate between the different hydrological responses".

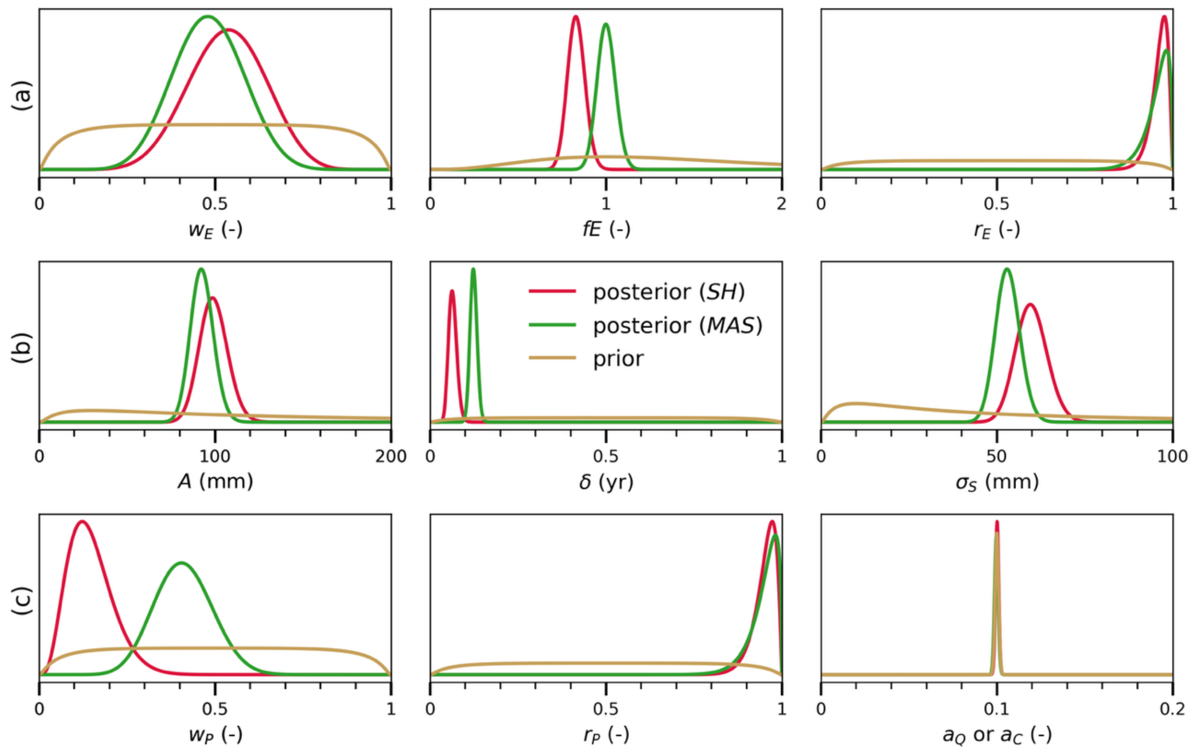
Reviewer comment 4:

"Figures with more than one panel (starting with Figure 2) need tags (a, b, c, etc) and the caption should refer to each panel specifically for clarity (like you did for Figure 4). In Table 2, indicate the meaning of the underlined values. Table 3, Table 4, again indicate the bold and underline importance."

<u>Reply on comment 4:</u>
Thank you for the suggestions regarding figures and tables. We will add the alphabetic labels for the subplots of Figures 2 and 3. The figures and their captions will be modified as follows:



"Figure 2: Monthly water balance posteriors (90% credible intervals in green) for: (a) *P* (precipitation), (b) *E* (actual evapotranspiration), and (c) *C* (canal water imports), and their corresponding observations (dots). Reference evapotranspiration computed here using ERA5 meteorological input variables is shown as (ET_0). The overbars of the labels in *P* and *E* plots indicate that these values are obtained through spatial averaging of the gridded data sets for each month *t*. Each year's label indicates the start of the year (January)."



"Figure 3: Prior and posterior distributions of error parameters for: (a) evaporation, (b) storage, and (c) precipitation, canal water import, and river discharge when using different GRACE storage data sets: CSR Spherical Harmonic (SH) and JPL Mascon (MAS) solutions."

We will also edit the table 2 caption as follows: "Basin-scale posterior covariance matrix of all water balance variables for July 2009, the diagonal elements represent the posterior variances (mm month^{-1}) (shown in bold), while the off-diagonal entries represent the posterior covariances between the variables (mm^2 month^{-2}) (underlined values)"

Also, we will add this line to the table 3 and 4 captions to enhance their clarity: "The bold and underlined values represent the results of the baseline setup of the water balance data fusion model, with GRACE-JPL Mascon data set and fixed correlation length parameters at 50 km".

Goteti, G. and Famiglietti, J.: Extent of gross underestimation of precipitation in India, Hydrology and Earth System Sciences Discussions, 2024, 1-40, 2024.
Schoups, G. and Nasseri, M.: GRACEfully closing the water balance: A data-driven probabilistic approach applied to river basins in Iran, Water Resources Research, 57, e2020WR029071, 2021.