Response to Reviewer #2's comments:

The authors would like to thank the reviewer for the feedback and the suggestions made to improve the manuscript. The authors generally agree with both the major and minor comments, and revisions were made accordingly. Please see below for the specific responses to the individual comments.

Summary of major changes (based on comments from all reviewers): (1) added a flowchart on the overall methodology; (2) the assumption on negligible inter-watershed groundwater flow (IGF) was clearly stated and discussed within the abstract, methods, discussion and conclusions sections; (3) added Appendix F showing a case study of imbalanced watersheds where IGF may not be the cause of UoP; (4) added monthly time series information on each watershed (in addition of the annual time series, Supplement S6); and (5) tables on watershed characteristics were added to the manuscript (Table 1) and Supplement (Table S6).

The major issues with the manuscript are:

1. The study fails to account for the fundamental assumptions of the catchment water balance method. The primary assumptions of catchment/watershed water balance are of a closed system and overlapping physical and hydrological boundaries, i.e. the watershed boundaries overlap with their respective aquifer boundaries. Both rarely occur in nature. Most aquifers are leaky, and multiple watersheds may fall in a single aquifer, or adjacent watersheds may share a common aquifer. Thus catchments with leaky aquifers (https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wat2.1386) or non-overlapping aquifer-watershed boundaries with significant inter-basin groundwater transfer (https://www.sciencedirect.com/science/article/abs/pii/S0022169420300433), can have significant water imbalance, apart from the limitations mentioned in Section 5 of the manuscript. It may also explain why most off-balance catchments were seen in mountains, as they have significantly high geological connectivity.

We thank the reviewer for bringing up this important caveat which we did not address in the original manuscript. We understand the comment by the reviewer that water balance methods often assume that watersheds are self-contained units and water flowing out of a watershed has been generated within the watershed. The manuscript was revised at several places (including the abstract, methods, discussion and conclusions) to make the reader aware of this important assumption. The literature cited by the reviewer and other relevant references were added to the manuscript.

We acknowledge that some of the observed water imbalance could be due to inter-watershed groundwater flow (IGF). In the absence of field data on the geologic makeup of the watersheds and their groundwater flow patterns it is not possible to definitively know the effect of IGF on streamflow. However, we present examples (in the newly created Appendix F) to show that IGF is unlikely the cause of the observed water imbalance in certain watersheds.

The objective of this paper is not to oversimplify hydrological process representation, but to bring forth the apparent widespread and persistent issue of precipitation underestimation. Towards this goal we are trying to use available information to the extent feasible and are making reasonable assumptions when needed. We acknowledge the limitations of both the information used and the assumptions made within our analysis.

2. The assumption that mountain and/or forested areas have minimal watershed management is not true. Many of the mountain regions and/or forested areas are included in watershed management through extensive government and civil society efforts. The activities include soil and water conservation efforts, which can change local and landscape-level hydrology. I would recommend the authors to have a relook at the assumption and its impact on the results.
https://www.cabidigitallibrary.org/doi/full/10.5555/20123144409 (Chauhan, 2010)

The phrase “minimal watershed management” needs additional clarification in the manuscript since it appears to give the reader the incorrect notion that there is hardly any management. The data on water management from India’s water agencies (discussed in the manuscript) shows that for many watersheds within the mountainous and forested regions of India, the combined annual effect of management – such as surface water diversions (imports and exports), groundwater extraction, and reservoir are storage, is typically less than 20% of annual precipitation. This implies that for a watershed with an annual runoff coefficient of 0.4, the effect of management on streamflow = 20% / 0.4 = 50%. In other words, up to 50% of the annual streamflow could be affected by “minimal watershed management”. Thus, “minimal watershed management” could still have a substantial effect on the annual streamflow.

We agree with the comment that watershed management may not be minimal in an absolute sense in mountainous or forested regions. It is important to note that we are making the above assumption at the watershed scale on an annual basis, and not at the landscape scale on a storm-event basis. The study suggested by the reviewer, Chauhan (2010), was added to the discussion (Section 5.1.1). Additional text has been added to the manuscript to clarify what is meant by “minimal watershed management”. Moreover, within the limitations section we discuss the issues with the water management data used in the study.

Our objective is to find plausible explanations for the observed spuriously high annual runoff coefficients. Towards this objective we use heuristics to differentiate between spurious and non-spurious instances of the annual water balance. The choice of heuristics can affect the results of this study. Hence the sensitivity analyses on such heuristics in Section 4.1 (‘Imbalanced watersheds using IMD-APHRO’).

3. The majority of the watersheds are in central and southern India, which reduces the studies’ representation.

Agree with the comment. Of the 242 watersheds analyzed in this study, 213 are in Peninsular India while 29 are in Northern India. Unfortunately, these are the only watersheds where streamflow data is publicly available. While data is collected by CWC (and other Indian water agencies) at gauging stations throughout India, such data is publicly available only for stations of Peninsular India. Data for the remainder of India is “classified” by these agencies and is unavailable to the authors. The data for Northern Indian watersheds had to be compiled from whatever limited information was available via reports published by water agencies.

This study demonstrates that even with a limited set of streamflow observations, one could identify that approximately a fourth to a third of the total watersheds in both Northern and Peninsular India are affected by water imbalance. It is hoped that this study would help bring forth the needed streamflow and hydrometeorological data.

4. I was excited to see a study on precipitation underestimation/overestimation. Few studies have approached doing hydrology backwards (https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2008WR006912), and rarely any from India. However, the study loses a chance to dwell on the processes behind UoP, which could have been explored by going deeper in a selected watershed like Nethravati. The study shows an overall UoP scenario in central and southern India but doesn't inform us enough on why it might be happening, apart from data issues.

Following the reviewer's suggestion, the monthly and annual water balance of the Nethravati watershed (upstream of Bantwal station), and eight other adjacent watersheds in Southwestern India were further analyzed in the newly created Appendix F. These watersheds, which form the headwaters of several river basins, were shown to simultaneously experience monthly runoff coefficients of greater than1.0 (i.e., $R > P_{obs}$). If inter-watershed groundwater flow played a substantial part in the monthly (or annual) hydrology, some of the watersheds would experience higher runoff coefficients at the expense of the other watersheds. But that is not the case. Considering all other potential causes, including water management and terrestrial water storage changes, UoP appears to be the likely cause of water budget imbalance.

The authors realize that identification of gross UoP using a "reverse hydrology" approach (or based on the water balance approach used here) is a challenging task considering the limitations of available data and the number of proxy datasets that needed to be compiled. Field-scale data on the geological formations underlying each aquifer and the groundwater flow patterns within such formations is not available. In the absence of such data, it is not feasible to delve into the specific mechanisms behind each watershed's annual water imbalance. The authors also acknowledge that UoP may not be the cause of the observed annual water imbalance in some of the watersheds. It is hoped that the scientific community gets access to the needed data and other resources to help understand this very important issue of watershed imbalance and UoP.

Detailed comments

L100 Could you please add a table with salient features of the 242 watersheds. Size distribution, Elevation bands, rainfall characteristics, % forests, %ET, etc.

Yes. A table showing the characteristics of all 242 stations used in this study is added to the supplement (Table S6). The metrics within Table S6 include catchment area, elevation (max and median), percent area covered by forests and cropland, annual average precipitation and annual average ET. A new table showing the number of stations within each major basin and other relevant information on streamflow data is added as Table 1.

There seems to be a lack of watersheds from Himalaya, Western Ghats, Western and Central India. Can the study claim to represent India in the title when most of the watersheds are in central and southern India?

Similar to major comment #3, the watersheds used here are only those where observed streamflow data is publicly available. India's water agencies have additional streamflow data but is not available to the authors. Despite the limited availability of data, imbalanced watersheds occur in almost all of the major basins – particularly in the head waters of these major basins. By shedding light on the limitations of data availability this study hopes to help bring forth the needed data.

L118-119: What is the strength of interpolation between APHRODITE and IMD. Would recommend adding any results to supplementary table.

In this study the gauge-based APHRODITE dataset was used in portions of the study domain where IMD data was unavailable. The similarity between the two datasets has not been explored in this study. However, studies in the literature have compared the two datasets (e.g., Prakash et al. 2015) and found that APHRODITE compares reasonably to IMD across many parts of India. Limitations with APHRODITE have also been discussed in such studies. Based on the literature studies APHRODITE is assumed to be the best gauge-based alternative to IMD. The text in Section 2.2 was revised accordingly.

L149-150: The CGWB groundwater dataset is known to be quite limited by the number of wells and their representativeness. How does that effect the manuscript?

Yes, the CGWB dataset is limited by the number of wells and the recharge estimates from CGWB are dependent on such data and the many assumptions made within their analyses. Some studies have discussed the pitfalls with analyzing trends in groundwater levels using the CGWB dataset (e.g., Hora et al. 2020). However, the CGWB dataset is the only observation-based dataset on groundwater extraction available for the entire study domain. Additional text has been added in Section 5.1.1 ('Limitations with Data').

L162-165: Are there any studies on the validation of TWS from GRACE against ground data? Please do add a section on the limitations of GRACE, especially in complex mountains or geologically diverse systems.

Some of the limitations of GRACE data have been discussed in Section 5.1.1 ("Limitations with data"). The effective resolution of GRACE data is about 300 km x 300 km. As such, GRACE captures only large-scale changes and cannot be directly compared with point-scale (e.g., groundwater wells) or small-scale changes in water table depth. Additional text has been added in Section 5.1.1 ('Limitations with Data').

L175 There seems to be an objective-method mismatch. The methods section describes the theoretical and per-watershed aspects well, but a flowchart mimicking the methods followed to achieve the three objectives would help us understand the study.

Agree with the comment. In order to help the reader better understand the objectives of this study and the methods, a flow chart on the overall methodology was added as Figure 1 in Section 1. This new figure complements Figure 7 (Figure 6, original manuscript; schematic on identification of spurious scenarios) and helps the reader better understand the overall analysis.

L175-180 What about leaky aquifers and inter-connected basins with significant groundwater transfer?

Similar to major comment #1, it is not known to what extent the watersheds analyzed in this study are affected by inter-watershed groundwater flow (IGF) – either because of 'gaining' aquifers or 'losing' (or leaky) aquifers. In the absence of field data on groundwater flow patterns within each watershed it is not possible to quantify the effect of IGF on streamflow. However, we present examples in the newly created Appendix F to show that IGF is not likely the explanation for the observed water imbalance in certain watersheds. Only field-level data can help establish the actual cause of watershed imbalance.

L200-205 Could you please define the exports and imports as per CGWB 2019 and how they are distinct from ∆GW natrual, ∆GW human and ∆Reservoir, with examples if possible?

Yes. Exports and imports are net surface water diversions and represent the net loss of water and net gain of water, respectively. ∆GW natural and ∆GW human are natural and artificial changes to groundwater, respectively. ∆Reservoir is the change in reservoir storage. Text has been added in Section 3.1 to clarify these definitions.

L210-211 The assumption is usually appropriate in catchments with high infiltration and transmissivity, eg Karst aquifers. Do any of the watersheds have such physical conditions?

Since this study does not assume that net change in annual storage is negligible, we have not investigated the watersheds where such change is negligible. However, we identified relevant literature on Karst aquifers (e.g., Dar et al. 2014), and included it in the manuscript (Section 5.1.2, 'Limitations with the methodology').

L240 – 245 This is a big assumption to make. Many of the mountain regions and/or forested areas are included in watershed management through extensive government and civil society efforts. The activities include soil and water conservation efforts which can change local and landscape level hydrology. Please have a relook. https://www.cabidigitallibrary.org/doi/full/10.5555/20123144409

Similar to major comment #2, we agree with the comment that mountainous or forested regions cannot be assumed to have minimal watershed management in an absolute sense of the word "minimal". Additional text was added to the manuscript (Section 3.1) to help the reader understand that "minimal" is used in a relative sense, and one could still have a substantial portion of the annual streamflow affected by "minimal management".

L246 Is off-balance the same as imbalance here? Please maintain consistency if they are the same or clarify if not.

Yes, "off-balance" and "imbalance" were used interchangeably in the original manuscript. As per the reviewer's suggestion, the word "off-balance" was replaced with "imbalance" throughout the revised manuscript and the revised supplement.

L258-260 Please provide details of the spatial syncing methods and any sensitivity analysis to check if there was any loss of information.

In order to compare streamflow at the outlet of the catchment with grid-based precipitation (ET, etc.) over the entire catchment, one has to aggregate the gridded datasets to the domain of the catchment. An area-weighted scheme which accounted for the spatial overlap of the catchment boundaries and individual grid cells was used to perform the aggregation. The GIS procedure employed here is typical to hydrologic studies. There is no loss of information within this aggregation process. The GIS procedure is illustrated through the schematic in Figure 8 within Section 3.2 (Figure 7, original manuscript).