<u>Response to Reviewers</u>

‘‘Remotely sensed reservoir water storage dynamics (1984-2015) and the influence of climate variability and management at global scale’’ by Jiawei Hou et al.

We thank the three reviewers for the thoughtful comments and constructive suggestions, which helped us improve the manuscript. We have thoroughly considered all comments and suggestions, and made modifications accordingly below (review comments in blue, our response in black bold font). The major changes include:

(1) We highlighted the advancement of this study over previous ones and clarified the logic of our analysis to determine the extent to which climate variability and human activity each affected global reservoir water volume over the past three decades.

(2) We acknowledged the limitation of our approach to deducing the influence of human activity on changes in reservoir storage and the need of in situ records of reservoir water releases to validate this part of our conclusion in specific cases.

(3) We implemented higher correlation (R) thresholds between A-L and between A-V for reservoir storage estimation and updated the subsequent long-term analysis.

(4) We added a validation of bias and error and included detailed validation results for individual reservoirs globally.

(5) We validated trend analyses of net evaporation against Zhao and Gao (2019).

(6) We included statistics on the number of reservoirs with Landsat observations for each month from 1984 – 2015.

(7) We added a worked example to explain the reliability, resilience, and vulnerability metrics.

In addition to replies to reviewers’ comments in the open discussion:

(8) We included an additional analysis on the influence of human activity on reservoir storage trends using global water withdrawal data.

<u>Reviewer #1 Comments:</u>

In this manuscript, Hou et al. estimated water storage dynamics for more than 6,000 reservoirs worldwide from 1984 to 2015 using a combination of Landsat imagery, radar satellite altimetry, and geostatistical modeling. They also analyzed the patterns of increasing and decreasing trends globally. Finally, they attributed reservoir storage changes to climate and human variables and found that precipitation and river inflows largely dominated reservoir storage changes.

I feel this is a very interesting study. Previous studies provided long-term storage changes for only dozes of reservoirs. It is really great to see a global dataset of more than 6,000 reservoirs, as compiled in this study. Their attributions on the reservoir storage changes can potentially inform local to regional water resources management. However, I have some major concerns on the quality of the global dataset and the methodology that they applied to attribute the storage changes.

R1C1) The Landsat satellites does not provide global coverage in the 1980s and maybe in the 1990s as well (Murray et al 2019). The authors did not acknowledge this limitation while stating they quantified reservoir storage from 1984 to 2015 globally. Is the produced storage time series consistent through 1984 to 2015? Could you provide a figure documenting the number of observations in each year in the time series from 1984 to 2015?

"Murray, N. J. et al. The global distribution and trajectory of tidal flats."

We agree that Landsat-derived products have limited observations in the 1980s, but this issue predominantly occurs in Oceania, Siberia, Greenland and parts of central and eastern Asia (Pekel et al. (2016); https://www.nature.com/articles/nature20584/figures/5). Landsat-derived water observations are available from 1984 onwards for most parts of Northern America, South America, Africa, Europe, and western and eastern Asia.

Furthermore, Zhao and Gao (2018) developed an algorithm to fill gaps in time series when the contamination/occultation in a Landsat image is between 5-95%, and applied interpolation and extrapolation for the missing monthly area estimates (i.e., no images or >95% invalid data). As a result, in their reservoir area product, there are 5,917 reservoirs that have Landsat observations every month from 1984 - 2015 (Fig. R1). To address this point, we changed the sentence in L99-102:

> *"Clouds, cloud shadows and terrain shadows cause errors or missing data for individual months, but Zhao and Gao (2018) developed an automated method to fill gaps in contaminated image classifications and enhance the accuracy and consistency of reservoir surface water extent estimates."*

and modified the sentence in L105-107:

> *"The resulting monthly data are available from 1984 to 2015 and there are 5,917 reservoirs have continuous observations every month over the 32 years. We used this data here as its temporal consistency fits the purpose of this study for long-term trend analysis."*

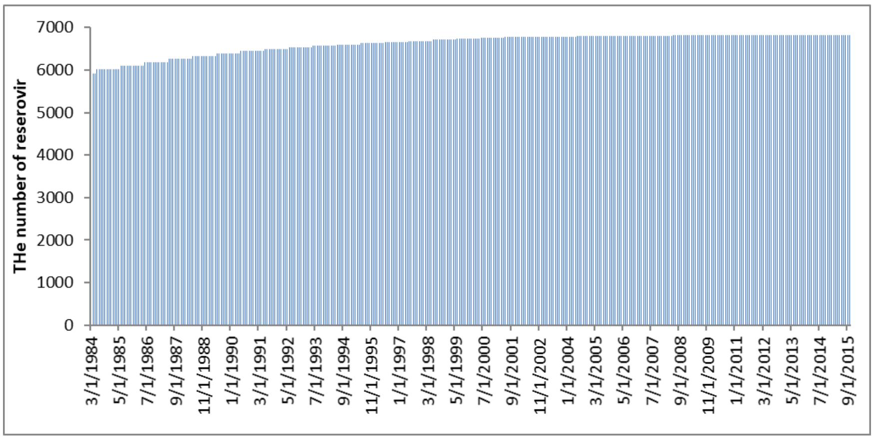

Figure R1 The number of reservoirs with Landsat observations for each month from 1984 - 2015 in the reservoir area product developed by Zhao and Gao (2018).

Despite that, we remained careful in using the reservoir area data for long-term storage trend analysis and included additional criteria. First, we removed reservoirs for which more than five years of (not necessarily consecutive) data were inter- or extrapolated. Second, we filtered out reservoirs with observations for less than 360 months (30 years), e.g. in New Zealand. After these steps, we found that 4,573 reservoirs constructed before 1984 have sufficient Landsat observations and these were used for long-term analysis, compared to the 6,669 reservoirs for which we produced monthly storage dynamics.

We modified the sentences in L183-186 to explain these steps:

> *"To ensure consistency in the 1984-2015 time series used for long-term trend analysis, we ignored reservoirs with less than 360 months (i.e., 30 years) of Landsat-derived observations or for which more than five years of water extent observations were inter- or extrapolated by Zhao and Gao (2018)."*

R1C2) While this study produces storage changes for a greater amount of reservoirs globally, I do not think the authors fully addressed the limitations that prevent previous studies from documenting reservoir storage dynamics with a better spatial coverage. The authors estimated storage changes for the 132 large reservoirs with both water areas and levels without assessing their consistency. Without a high correlation between water areas and levels, it makes no sense to me to combine these two to deduce storage changes. The authors need to refer to existing studies (e.g., Busker et al.) on quality control before simply combining satellite observations. The authors used a geostatistical method to estimate the storage changes in the vast majority of reservoirs, on which I have an even greater concern. The authors need to be aware that the mean depth, as archived in the HydroLakes dataset, is a ratio of the total volume and maximum lake area. The mean depth does not provide any meaningful information of the actual water depth. Additionally, the geostatistical model adopted by Messager et al. is a spatial model measuring the relationship between the total storages and maximum areas for a large group of water bodies. The authors tried to use the outcome (e.g., mean depth) to estimate storage changes in each individual reservoir, which differs from the purpose of the Messager et al. Unless the authors provide a comprehensive validation, I am not convinced the proposed method is feasible to estimate storage changes for the majority of studied reservoirs here.

"Busker, T. et al. A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry. Hydrol. Earth Syst. Sci. 23, 669–690 (2019)."

"Messager, M. L., Lehner, B., Grill, G., Nedeva, I. & Schmitt, O. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. Nat. Commun. 7, (2016)."

We thank the review for this suggestion. Following comments R1C2 and R1C7, we increased correlation (R) thresholds between area and level (A-L) and between area and volume (A-V) for reservoir storage estimation. We regard R values above 0.7 as indicating strong correlation, and used this as the correlation threshold. For group A, we only calculated reservoir storage when the correlation between area and level exceeded 0.7. Storage dynamics between 1984-1993 (when

altimetry data is not available) were estimated from area if the correlation with volume exceeded 0.7 between 1993-2015. We changed sentences in L131-134 to explain these steps:

> *"In total, 132 large reservoirs had records of both surface water extent and height for the overlapping period 1993–2015. Strong correlation (R ≥0.7) between extent and height was found for 58 reservoirs (Group A; Fig. 1). For these, we estimated the height and area at capacity as the maximum observed surface water height and extent, respectively, and calculated reservoir storage volume (V_o in GL or 10^6 m^3) as:"*

in L137-139:

> *"There were 53 reservoirs with a relationship between A_o and V_o for this overlapping period with a Pearson's R≥0.7. For these reservoirs, V_0 was estimated from 1984 onwards using a cumulative distribution function (CDF) matching method based on A_0."*

We updated the subsequent long-term analysis as well (from Fig. 1 to Fig. 11; please see in the revised manuscript). These stricter measures did not in any way affect the conclusions of our study, but arguably made them more statistically robust.

We would like to clarify that we did not directly use the mean depth archived in the HydroLAKES dataset (Messager et al., 2016). Indeed, this value is not related to the geo-statistical model and is simply the ratio of the reported volume and lake area. The geo-statistical model, on the other hand, provides the empirical relationship of the mean depth with water surface area and the average slope within a 100 m buffer around the water body (Table S2). Messager et al. (2016) have validated the predicted lake depth and volume derived from the geo-statistical model against observed data. The symmetric mean absolute percent error (Eq. (R1)) and correlation between predicted and observed lake depth are 47.4% and 0.71, respectively (Messager et al., 2016). Furthermore, the SMAPE and correlation between predicted and reference volume are 48.8% and 0.95, respectively (Messager et al., 2016).

$$SMAPE = 100 \times \frac{1}{N} \sum \frac{|\text{observed value} - \text{predicted value}|}{(\text{observed value} + \text{predicted value})/2} \quad \textbf{(R1)}$$

We used this statistical model to estimate reservoir depth and volume dynamics from 1984-2015. We modified the sentences in L154-158 to clarify that the lake depth was predicted by the geo-statistical model in the revised manuscript:

> *"Messager et al. (2016) proposed a geo-statistical model that provides the empirical relationship of the mean lake or reservoir depth with water surface area and the average slope within a 100 m buffer around the water body. The main assumption of this model is that lake bathymetry can be extrapolated from surrounding topography using slopes. Four empirical equations to predict depth from area and slope were developed by Messager et al. (2016) for different lake size classes (i.e., 0.1–1, 1–10, 10–100 and 100–500 km^2) (Table S2)."*

In addition, also responding to comments R1C2 and R1C13, we included the absolute error (SMAPE) in Fig.2 and Fig.3 and listed the SMAPE and correlation metrics for individual reservoirs in supplementary material. We modified the corresponding paragraph in L235-248:

> *"Monthly storage data with at least 20-year time series of 65 reservoirs via the US Army Corps of Engineers and Australian Bureau of Meteorology were used for error assessment. Comparison of observed and estimated volumes showed R>0.9 for 69% of the 65 reservoirs, and R>0.7 for 89% of them (Table S3). Messager et al. (2016) reported that the symmetric mean absolute percent error (SMAPE) of the geo-statistical model is 48.8% globally. The average SMAPE between predicted and reference volumes was 27.8%, lower mainly because we adjusted reservoir storage estimates by reported reservoir capacity. Some cases are shown in Fig. 2. Annual average water levels for Lake Aswan, one of the largest reservoirs in the world, were published as a graph only (El Gammal et al. 2010); comparison showed strong agreement between the satellite-derived storage and in situ measurements (R=0.97, Fig. S1). Assuming the estimation method for Group A is more accurate than that for Group B, the latter can also be evaluated against the former. The results show that 25 of the total 33 overlapping estimated reservoirs show strong agreement (R≥0.9) between the two methods, and the average SMAPE between them is 13.1%. This implies good consistency of reservoir storage estimates from Group A and B. Some cross-validation examples are shown in Fig. 3. The average Pearson correlation between our Landsat-derived water volumes and published MODIS-derived estimates (Tortini et al. 2020; 1992–2015) for 100 reservoirs was R=0.87, and R values did not vary as a function of reservoir size."*

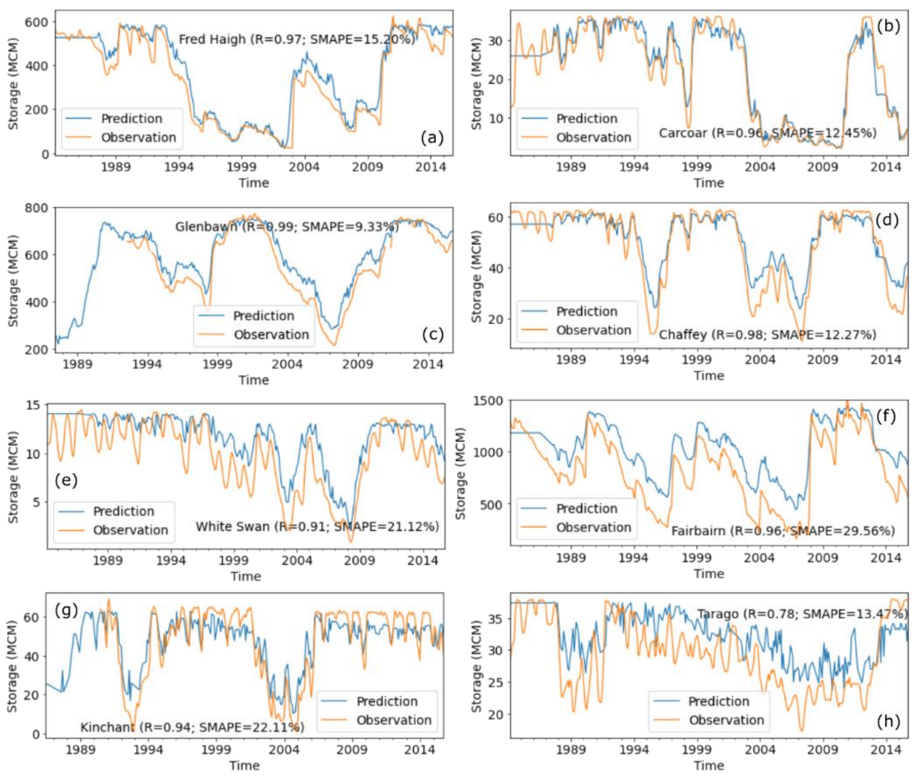

Figure 2 Validation examples of monthly reservoir storage time series reconstruction against in situ storage data.

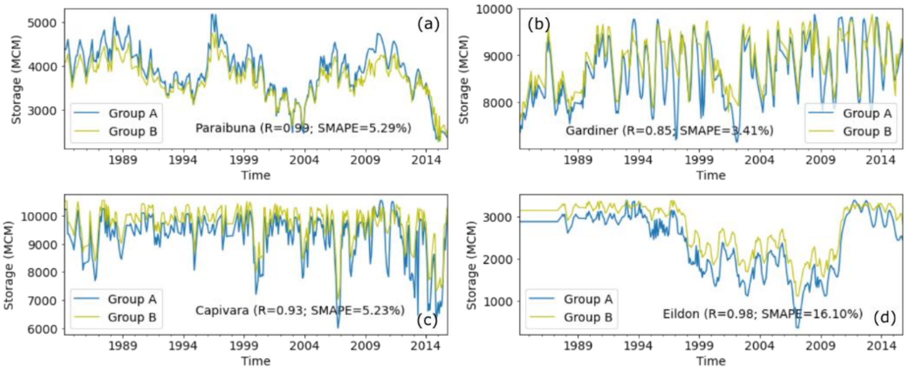

Figure 3 Validation examples of monthly reservoir storage time series reconstruction for Group B against results obtained using the method for Group A.

Table S3 The SMAPE and Pearson correlation between predicted and reference volumes for 65 reservoirs

Grand ID	Dam Name	Latitude	Longitude	Capacity (MCM)	R	SMAPE (%)
307	Fort Peck Dam	48.00	-106.41	23560	0.98	28.6
597	Glen Canyon	36.94	-111.49	25070	0.99	39.1
753	Garrison Dam	47.51	-101.43	30220	0.97	31.0
870	Oahe Dam	44.46	-100.40	29110	0.97	30.5
6199	Darwin River Dam	-12.83	130.97	265	0.90	8.2
6579	Tinaroo Falls	-17.16	145.55	407	0.91	11.0
6581	Paluma	-18.95	146.15	12.3	0.77	18.3
6582	Copperfield River Gorge	-19.04	144.12	20.6	0.79	14.5
6583	Ross River	-19.41	146.74	417	0.95	59.5
6586	Peter Faust	-20.37	148.38	500	0.94	28.7
6588	Burdekin Falls	-20.65	147.14	1860	0.89	14.7
6592	Eungella	-21.14	148.39	131	0.94	27.1
6593	Kinchant	-21.21	148.90	62.8	0.94	22.1
6594	Fairbairn	-23.65	148.07	1440	0.96	29.6
6595	E.J. Beardmore	-27.91	148.65	101	0.84	30.5
6600	Windamere Dam	-32.73	149.77	368	0.96	26.9
6603	Carcoar Dam	-33.62	149.18	35.8	0.96	12.5
6605	Wyangala	-33.97	148.95	1220	0.97	22.2
6613	Burrinjuck	-35.00	148.60	1026	0.92	39.0
6618	Blowering	-35.40	148.24	1628	0.92	42.0
6619	Googong	-35.42	149.26	124.5	0.97	7.5
6620	Bendora	-35.45	148.83	11.1	0.37	12.2
6621	Corin	-35.54	148.84	75	0.72	20.0
6629	Eucumbene	-36.13	148.61	4800	0.99	34.6
6652	Malmsbury	-37.21	144.37	18	0.92	43.7
6655	Lauriston	-37.27	144.39	20	0.80	14.0
6656	Upper Coliban	-37.29	144.39	37.5	0.80	57.9
6657	Rosslynne	-37.47	144.57	24.5	0.93	75.8
6658	White Swan	-37.52	143.92	14.1	0.91	23.1

6659	Yan Yean	-37.55	145.13	32.7	0.93	43.4
6662	Greenvale	-37.63	144.90	27.5	0.83	11.9
6663	Maroondah	-37.64	145.56	28.4	0.65	46.4
6664	Upper Yarra	-37.67	145.90	207.2	0.58	38.9
6667	Silvan	-37.84	145.42	40.2	0.27	8.2
6668	Glenmaggie	-37.91	146.80	190	0.84	39.7
6669	Cardinia	-37.97	145.39	288.9	0.90	19.1
6670	Tarago	-38.02	145.94	37.5	0.78	13.5
6673	Devilbend	-38.29	145.11	14.5	0.93	9.8
6676	West Barwon	-38.53	143.72	21.7	0.74	61.6
6701	Awoonga High	-24.07	151.31	300	0.96	40.0
6702	Callide	-24.37	150.62	127	0.96	49.1
6703	Cania	-24.65	150.98	89	0.93	60.3
6704	Fred Haigh	-24.87	151.85	586	0.97	15.2
6706	Glebe Weir	-25.46	150.03	17.3	0.69	35.4
6707	Boondooma	-26.10	151.43	212	0.93	11.6
6708	Bjelke-Petersen	-26.30	151.98	125	0.98	13.9
6709	Borumba	-26.51	152.58	42.6	0.91	13.6
6715	Cressbrook	-27.26	152.20	83	0.98	29.4
6717	Perseverance Creek	-27.30	152.12	30.9	0.95	27.4
6723	Moogerah	-28.04	152.54	92.5	0.92	45.6
6725	Maroon	-28.19	152.65	38.4	0.95	21.5
6726	Leslie	-28.22	151.92	108	0.98	26.9
6728	Coolmunda	-28.44	151.22	75.2	0.93	18.1
6731	Glenlyon	-28.98	151.46	254	0.91	23.2
6731	Glenlyon	-28.97	151.45	254	0.92	20.5
6733	Copeton	-29.90	150.92	1364	0.81	42.6
6735	Split Rock Dam	-30.58	150.70	372	0.95	39.5
6736	Keepit Dam	-30.88	150.49	423	0.93	23.4
6737	Chaffey	-31.35	151.14	61.8	0.98	12.3
6738	Glenbawn	-32.10	150.99	750	0.99	9.3
6739	Chichester	-32.24	151.69	17.7	0.47	15.8
6740	Lostock	-32.33	151.46	20	0.63	9.9
6741	Glennies Creek	-32.36	151.25	283	0.97	18.7
6742	Grahamstown	-32.77	151.79	152.6	0.84	18.3
6743	Mangrove Creek	-33.22	151.13	170	0.93	49.2

Table S4 The SMAPE and Pearson correlation of predicted volumes between Group A and B for 33 reservoirs

Grand ID	Dam Name	Latitude	Longitude	Capacity (MCM)	R	SMAPE (%)
250	Mica	52.08	-118.57	25000	0.90	15.2
253	Gardiner	51.27	-106.86	9870	0.84	3.4
297	Libby	48.41	-115.32	7434.2	0.89	16.0
310	Grand Coulee	47.95	-118.98	6395.6	0.92	13.2
370	Cascade	44.52	-116.05	805.5	0.98	15.4
597	Glen Canyon	36.94	-111.49	25070	0.99	22.4
1275	Sam Rayburn Dam And Reservoir	31.07	-94.11	7815.6	0.94	5.4
1320	International Falcon Lake Dam	26.56	-99.17	3920	0.96	13.7
1863	Buford	34.16	-84.07	3150.3	0.93	9.1
2376	Itumbiara	-18.41	-49.10	17000	0.96	7.8
2377	Emborcacao	-18.45	-47.99	17590	0.97	6.4
2388	Mascarenhas de Moraes	-20.28	-47.06	4040	0.91	3.3
2405	Capivara	-22.66	-51.36	10540	0.93	5.2
2416	Paraibuna	-23.36	-45.66	4732	0.99	5.3
2447	Passo Fundo	-27.55	-52.74	1570	0.97	6.5
2467	Araras	-4.21	-40.45	1000	0.98	12.1
2490	Boa Esperanca	-6.75	-43.57	5060	0.94	9.7
3014	Bagre	11.47	-0.55	1700	0.95	5.8
3670	Mape	6.04	11.30	3300	0.94	8.3
4212	Sterkfontein	-28.39	29.02	2620	0.99	27.5
4431	Karakaya	38.23	39.14	9580	0.87	7.7
4500	Nyumba ya Mungu	-3.82	37.47	1135	0.89	12.4
4501	Mtera	-7.14	35.98	3200	0.98	13.7
4686	Kayrakkum	40.28	69.82	4160	0.98	4.6
4702	Tarbela	34.09	72.69	13940	0.74	30.0
4715	Kajakai	32.32	65.12	2680	0.86	12.2
4739	Ukai	21.26	73.60	8510	0.80	16.2
4943	Upper Indrawati	19.28	82.83	2300	0.99	41.1
5150	Lam Pao	16.60	103.45	1430	0.97	9.9
5796	Sirindhorn	15.21	105.43	1966	0.97	8.4
5902	Shuifeng	40.46	124.97	14700	0.86	13.8
6606	Lake Victoria	-34.04	141.28	680	0.98	33.6
6653	Eildon	-37.22	145.93	3390	0.98	16.1

R1C3) The presented attribution on reservoir storage changes seems to be so simplified that I have many concerns. First, the authors simply compared the directions of the trend in reservoir storage versus that in potential drivers but the analysis only produces coincidence rather than causation. Second, the authors conclude that the evaporation did not significantly impact the reservoir storage but the calculation for the evaporation is too cheap. The authors may need to use more advanced approaches (e.g., Zhao and Gao) in order to draw a confident conclusion. Third, reservoirs, particularly large ones as documented in GranD dataset, are highly regulated by humans. The authors depend on the outputs of global models on estimating human water release from reservoirs. Are the data really reliable for producing trend in human release for each reservoir? In sum, the authors need to pay careful attention to these limitations that potentially affect their conclusions.

“Zhao, G. & Gao, H. Estimating reservoir evaporation losses for the United States: Fusing remote sensing and modeling approaches. Remote Sens. Environ. 226, 109–124 (2019)”

We thank the reviewer for this comment, but in fact we explicitly considered the difference between coincidence and causation in our study to the extent possible. In a first step, we indeed looked at the coincidence of trends per se. We identified that the spatial distribution of trends of storage and in situ river flow show very similar global patterns (Fig. 4). We could not relate each individual reservoir to a corresponding river gauge because the limited number of gauging stations upstream of reservoirs globally. Instead, in a second step, we performed trend analysis using modeled river flow (validated against in situ river flow in Fig.8b) at the basin scale, given total basin water storage can be expected to respond to a change in overall precipitation and streamflow. We confirmed the same directions of trends between precipitation, streamflow and reservoir storage in most basins, though not all (Fig. 7). Third, we focused more on attribution by calculating Pearson correlations among the different variables, which provides evidence if not proof for a causative relationship. Thus, we showed that there are reasonably strong correlations among linear trends in precipitation, streamflow and reservoir storage (Fig. 8a and c). Furthermore, positive relationships between annual time series of storage change and reservoir inflow and between reservoir inflow and precipitation were found in a majority of basins globally (Fig. 9). We modified the last paragraph (L79-94) in the Introduction section to clarify the logic of our analysis as early as possible in the revised manuscript.

> *“In this study, we combined Landsat-derived surface water extents, satellite altimetry, and geo-statistical models to reconstruct monthly reservoir storage globally for 1984-2015, and examined long-term trends of global reservoir water storage and changes in reservoir resilience and vulnerability over the past three decades. Part of our objective was to determine the extent to which climate variability and human activity each affected global reservoir dynamics over the past three decades. It is currently impossible to analyse the influence of human activity at global scale directly: there are very few in situ reservoir water release records available publicly, and no hydrological models that can provide reliable estimates. Instead, we consider all climate terms in the reservoir water balance and infer the influence of the remaining unknown term, water releases. First, we investigated trends in precipitation, streamflow and storage at both the reservoir and basin level. If the trends between these variables show similar spatial patterns globally, then this increases the likelihood that climate variability commonly explains storage changes. Second, we examined the temporal correlation between precipitation, reservoir inflow and storage change to further understand potential causative relationships. Third, beyond reservoir releases, net evaporation is the only other potential loss term, and we examined what fraction of observed*

trends in storage was attributable to net evaporation. Using the combined insights, we deduced the role of human activity on reservoir storage change, noting that a direct attribution would require in situ records of reservoir water releases. To support our inference, we analysed the trends of global water withdrawal to discuss whether it could be a significant factor to lead reservoir storage change."

With regards to reservoir evaporation, our estimates are robust. Various hydrological variables estimated by the W3 model have been evaluated in previous studies. Therefore, we argue that the E_0 derived from the W3 model are entirely appropriate to analyze linear trends in net evaporation. However, we are also able to provide more direct evidence. Zhao and Gao (2019) estimated evaporation losses for 721 reservoirs in the contiguous United States using three different meteorological datasets, including TerraClimate, North American Land Data Assimilation System phase 2 (NLDAS-2) forcing and Global Land Data Assimilation System Version 2 and Version 2.1 (GLDAS-2 and GLDAS-2.1). We used their monthly reservoir evaporation (1000 m^3/month) estimates to compare trends in net evaporation with those in the W3 model estimates we used (Van Dijk et al. 2018) for 721 reservoirs. The results show strong agreement in derived linear trends, especially with regards to the more detailed TerraClimate dataset, which would be expected the most accurate among the three (Fig.R2).

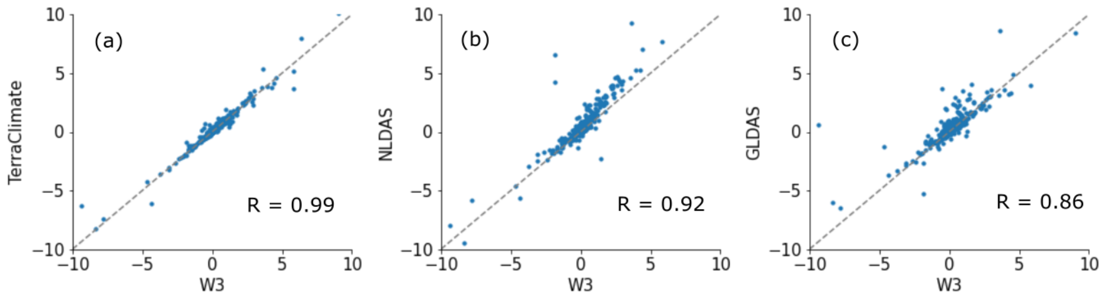

Figure R2 The comparison (dash grey line: 1: 1 line) of the linear trends (%) in net evaporation using evaporation losses derived from the W3 model, TerraClimate, NLDA and GLDAS for 721 reservoirs.

No eartH2Observe models includes the impacts of reservoirs on river flows, and there is currently no global hydrological model that is capable of estimating the impact of historical operational water management at the reservoir or basin level with any acceptable degree of accuracy – hence the need for our approach. We focused on all climate terms of the reservoir water balance and inferred the influence of the remaining unknown variable (i.e. reservoir releases). Specifically, we analysed the interaction between precipitation, streamflow, evaporation and reservoir volume and inferred the influence of human activity given that the water volume dynamics in a reservoir is the net balance of inflow (streamflow, affected by precipitation), net evaporation (i.e., evaporation minus direct precipitation) and reservoir releases (L349-363). We modified the last paragraph (L79-94) in the Introduction section to better explain the logic of our approach to infer the role of human activity in reservoir storage

long-term changes early on in the revised manuscript. (Please see the revised text in the first paragraph)

Specific comments:

R1C4) Line 25: "The majority of …particularly in South America, Southeast Asia and Africa". The authors may consider add more relevant references here.

"Wang, J. et al. GeoDAR: Georeferenced global dam and reservoir database for bridging attributes and geolocations. Earth Syst. Sci. Data 0–52 (2021)"

"Mulligan, M., van Soesbergen, A. & Sáenz, L. GOODD, a global dataset of more than 38,000 georeferenced dams. Sci. Data 7, 1–8 (2020)."

Thank you. We included these two references in L33-34 in the revised manuscript:

> *"The majority of these are in developing countries, particularly in South America, Southeast Asia and Africa (Bonnema et al. 2016; Mulligan et al. 2020; Wang et al. 2021; Zarfl et al. 2015)."*

R1C5) Line 65: Schwatke et al. 2019 is another study on estimating long-term lake area changes.

Schwatke, C., Scherer, D. & Dettmering, D. Automated Extraction of Consistent Time-Variable Water Surfaces of Lakes and Reservoirs Based on Landsat and Sentinel-2. Remote Sens. 11, 1010 (2019)

We included this paper in L69-71 in the revised manuscript:

> *"Several regional and global time series of reservoir water extent have been produced based on MODIS, Landsat or Sentinel-2 imagery (Khandelwal et al. 2017; Ogilvie et al. 2018; Schwatke et al. 2019; Yao et al. 2019; Zhao and Gao 2018)."*

R1C6) Line 89: It is hard to understand "coefficient of determination" here. Could you define or explain it?

We used Pearson correlation throughout this paper. We converted coefficient of determination (R^2) to Pearson correlation (R) in L104-105 in the revised manuscript:

> *"The average Pearson correlation (R) between satellite-derived extent and observed elevation or volumes increased from 0.66 to 0.92 using the algorithm developed by Zhao and Gao (2018)."*

R1C7) Line 120: I do not quite understand what's the purpose showing the correlation between A0 and calculated V0 (based on A0). It makes more sense to me to show the correlation between A0 and h0 as these two are independent estimates. The authors may only need to consider a pearson' R greater than 0.8 (or R2 higher than 0.6) as correlation between A-L or A-V should be pretty high, otherwise it indicates substantial uncertainty in the data sets.

Thank you for this suggestion. We increased correlation (R) thresholds between A-L and between A-V for reservoir storage estimation. Please see our response to R1C2 for full details.

R1C8) Line 135: the equation does not make sense to me. The authors need to show more details about the rationale.

We apologize for the confusion. We modified the sentences in L160-166 to clarify the rationale of this equation:

"In line with Eq. (1), we assumed maximum observed surface water extent (A_{max}) as the area at capacity. Water depth (D_c in m) at capacity was calculated as the ratio of volume (V_c) and area (A_c=A_{max}) at capacity:

$$D_c = \frac{V_c}{A_c} \qquad (2)$$

A bias-corrected water depth (D in m) was calculated by solving D based on the ratio of water depth (D_c in m) at capacity and maximum observed depth (D_{max} in m):*

$$D^* = D \times \frac{D_c}{D_{\max}} \qquad (3)\text{"}$$

R1C9) Line 150: "Only 132 reservoirs with both area and level observations….". Do you conclude based on the 132 reservoirs or all reservoirs, majority of which do not have both observations?

We performed a long-term trend analysis for 4,573 reservoirs, including both Group A (area and level observations available) and Group B (only area) (L189-190). This sentence aimed to introduce how many reservoirs and how much capacity were estimated in Group A and B. We modified the sentences in L179-183 to clarify this information:

"There were only 58 reservoirs for which storage dynamics could be estimated most directly, by a combination of satellite extent and water level observations (Group A), but together they already account for 25.5% of combined global reservoir capacity (Fig. 1). The total capacity of the 193 reservoirs not measured constitutes 36.4% of global capacity. There were 6,611 reservoirs in Group B for which by the geo-statistical approach could be applied, and these contribute 41.1% to total global capacity."

R1C10) 164: It seems the MSWEP v1.1 may not be the latest version of the dataset.

Although there is a latest version of MSWEP now, we carried out our study using MSWEP 1.1 two years ago. We considered updating the results with the latest MSWEP product, but among us are authors of the MSWEP product and with knowledge of the changes between successive versions we are confident there are no meaningful differences for the type of long-term analysis done here.

R1C11) 192: The authors only validated on 1% of the studied reservoirs and the validation samples are located in U.S. only, which could be a concern.

Please see our response to R1C2 for full details: we provide validation of our calculated storage volumes for 65 reservoirs with publicly available storage data from US and Australia and cross-

validation between two volume estimation methods for 33 reservoirs globally. We note that the availability of *in situ* data is severely limited, which was the primary reason for us to develop a remote sensing-based methodology. However, to further evaluate our storage estimates, we also compared our estimates with MODIS-derived water storage dynamics for 1992–2018 for another 100 reservoirs by Tortini et al. (2020) (L246-248).

R1C12) 194: What do you mean by "published"? The authors use pearson's R (correlation) for doing validation, which does not give insights on the accuracy of estimated values.

We changed "published" to "observed". We think Pearson correlation is the more relevant metric for this study as we focus on trend analysis, which depends on temporal patterns rather than absolute values. However, we included the absolute error (SMAPE) valuation and compared this validation to Messager et al. (2016) in the revised manuscript. Please see our response to R1C2 for full details.

R1C13) Figure 2: it would be more clear to show global-scale evaluation and move the evaluations on individual cases into the supplementary.

Thank you for this suggestion. We added the validation results for individual reservoirs in the supplementary material. Please see our response to R1C2 for full details.

R1C14) Line 214: "a positive trend in combined global reservoir storage of 3.1 km3 per yr". This rate seems to less than 10% of earlier estimates on global reservoir storage rates (e.g., Chao et al.). Could you provide an uncertainty estimate for this rate?

"Chao, B. F., Wu, Y. H. & Li, Y. S. Impact of artificial reservoir water impoundment on global sea level. Science (80-.). 320, 212–214 (2008)."

We explicitly performed trend analysis for the reservoirs constructed before 1984 only, in order to remove the influence of new reservoir water impoundments after that year. This was done deliberately, to provide a clearer understanding on the interaction between precipitation, streamflow, evaporation and reservoir volume. Our study therefore differs in key aspects from Chao et al. (2008), who focused on the change in cumulative storage by increased water impoundment. We clarified this information in L186-188 in the revised manuscript:

> *"Our focus was on interactions between precipitation, streamflow, evaporation and storage in existing reservoirs, rather than the consequences of new impoundments. Therefore, we excluded from consideration all reservoirs that were destroyed, modified, planned, replaced, removed, subsumed or constructed after 1984."*

R1C15) Line 215: "this was almost entirely explained by positive trends for the two largest reservoirs in the world, Lake Kariba (+0.8 km3 yr-1) on the Zambezi River and Lake Aswan". This statement is confusing. I know some completed projections of megadams in China and Brazil, such as the three gorges dams.

Please refer to R1C14. Most mega-dams in China and Brazil were constructed after 1984 and hence were excluded from our long-term analysis.

R1C16) Line 219: "while 948 reservoirs showed increasing trends, distributed in northern North America and southern Africa". The reported hotspots of increasing reservoir storage are inconsistent with the patterns of recent dam booms.

Please refer to R1C14.

R1C17) Figure 4: This map is confusing to me. For example, China may be the global lead in dam constructions during the study period. Why its reservoir storage decreased? Is the data correctly shown in this map?

Please refer to R1C14.

R1C18) Line 245: "We summed storage for individual reservoirs to calculate combined storage in 134 river basins worldwide". Do reservoirs show a similar pattern of storage change in the same river basin? Is it more meaningful to analyze each of them individually?

The majority of reservoirs showed the same trends in each basin where a significant trend in total storage was found (Fig. 11). We show the trend analysis at both the individual (Fig.4a) and basin scale (Fig.7c). Due to the lack of accurate data on the contributing area for each individual reservoir, we performed our climate analysis at the basin scale. The strong spatial correlation in all variables involved meant that combined basin reservoir water storage can be related to changes in basin-average precipitation and streamflow. However, we do find and explicitly discuss cases where those 'average' relationships appeared to break down.

R1C19) Line 268: "In summary, we did not find evidence for widespread reductions in reservoir water storage due to increased releases". Reservoir storage increase could be a result of increased impoundments. Did you consider that?

Please refer R1C14. We have removed the effect of the increased reservoir water impoundments from 1984-2015. Therefore, reservoir storage increase cannot be a result of increased impoundments.

R1C20) Line 339: As Zhao and Gao used contaminated Landsat imagery to increase the monthly coverage of reservoir areas by 80%, do the estimates from poor-quality images affect your storage analysis? I know some studies (e.g., Busker et al) only adopted good-quality images due to this issue.

"Busker, T. et al. A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry. Hydrol. Earth Syst. Sci. 23, 669–690 (2019)."

We think the consistency of observations from 1984-2015 is more important for long-term analysis. This is the reason why we used the reservoir area product developed by Zhao and Gao (2018). In addition, Zhao and Gao (2018) demonstrated that the correlations between observed and estimated reservoir areas were improved from 0.66 to 0.92 by 'repairing' contaminated Landsat images.

However, we did include a reference to the study of Busker et al., as it is certainly relevant to the topic.

Reviewer #2 Comments:

General Comments

This study demonstrates an integrated remote sensing framework for improving the understanding of long-term reservoir storage dynamics at the global scale. The methods of this study highlight a combination of well-established quantitative approaches and publicly available data sets and have the potential to benefit studies across water resources management and satellite remote sensing. The manuscript is well written and organized, but further explanation or clarification might be needed on the hydrology part, particularly for some components of trend analysis and associated conclusions.

Specific Comments

R2C1) My major concern is that the trend analysis didn't include reservoir outflow and water use at the reservoir or basin level. The authors did attempt to explain the lack of data behind their decision, but this may not be sufficient to justify an incomplete analysis of the reservoir water balance. Without a reasonable estimation of the dynamics of outflow and water use, it is not convincible that the trend in precipitation/stemflows alone can effectively explain the trend in reservoir storage, particularly for those reservoirs where the trends in precipitation/streamflow and storage are not consistent. Therefore, some of the conclusions on the influence of water use are not robust, e.g., lines 17-18, 221-223, 248-249, 267-268, 362-365, and 376-377.

We thank the review for this comment. Attributing the causes of reservoir storage change is at the same time important and challenging. There are no water demand and supply or dam operation data available globally (and even very hard to come by locally), and so we are not able to assess the influence of human activities on reservoirs directly using such data. The underlying principle of our study is that the water volume dynamics in a reservoir are the result of a relatively simple water balance involving inflow (streamflow, driven by precipitation), net evaporation (i.e., evaporation from minus direct precipitation onto the reservoir) and reservoir releases. Following that logic, we analysed the individual terms inflow (temporal correlation) and net evaporation (trend ratio in volume) and then, where possible, deduced the role of dam water releases as the residual. This indirect method was the only approach open to us, given the lack of water release data, but thanks to the small number of terms in the reservoir water balance, applying this logic we were still able to derive useful insights.

Thus, for the majority of the 65 basins with significant storage changes, trends were of the same sign for storage, runoff and precipitation (Fig.7). If rainfall and runoff trends show the same directions as reservoir storage, then it is plausible that rainfall variations dominate reservoir storage trends. On the other hand, if rainfall/runoff and reservoir storage show *opposite* trends, that strongly suggests that direct evaporation or water releases (or both) are the driving process.

We were able to exclude the former as a driving process. We propose that this logical framework is robust but welcome arguments as to why it might not be.

Incidentally, there are other recent studies that come to similar conclusions about the limited impact of water releases. We included these studies to support our findings in L359-363 in the revised manuscript:

> *"Evidence that the impact of human activity is less than that of climate variability is also found in other recent studies. For example, Wang et al. (2017) found that climate variability was the dominant driver of the decreasing trend in lake area across China's Yangtze Plain; human activities only accounted 10-20% of trends despite construction of the Three Gorges Dam upstream. Furthermore, Gudmundsson et al. (2021) demonstrated that climate change dominates changes in river flow from 1971-2010 worldwide, rather than water and land management."*

Overall, however, we agree with the reviewer that we do not have direct evidence on reservoir releases (or downstream water use) and that, although our interpretation is coherent and logical, some of the associated conclusions are not as robust as we might have preferred. Hence, on one hand, we tempered the relevant statements to acknowledge the indirect nature of our evidence, for example in L18-24:

> "*Many of the observed reservoir changes could be explained by changes in precipitation and river inflows, emphasising the importance of multi-decadal precipitation changes for reservoir water storage. Our results also show that there is generally little impact from changes in net evaporation on storage trends. Based on reservoir water balance, we deduce it is unlikely that water release trends dominate global trends in reservoir storage dynamics. This inference is further supported by different spatial patterns in water withdrawal and storage trends globally. A more definitive conclusion about the impact of changes in water releases at global or local scale would require data that unfortunately are not publicly available for the vast majority of reservoirs globally*."

in L287-292:

> *"If precipitation and runoff trends show the same direction as reservoir storage trends, then it is plausible that climate variations play an important role in reservoir storage trends. On the other hand, if rainfall and runoff show opposite trends to those in reservoir storage, then that could suggest a dominant influence from either net evaporation or water releases. For the majority of the 65 basins, the trends were of equal sign for storage, runoff and precipitation, suggesting that precipitation changes are commonly the most likely explanation for observed trends (Fig. 7a and b)."*

in L428-433:

> *"Both lakes and reservoirs are influenced by changing inflow and net evaporation in response to climate variability. Although human regulation has more influence on reservoirs than on natural lakes, our results suggest that for the majority of basins natural influences dominate human impacts, although human impacts on the hydrological regime still exist, of course. For example, Cooley et*

al. (2021) found that human interventions have resulted in larger seasonal variability in reservoirs than that in lakes globally."

in L443-448:

"Reservoir storage dynamics are the net result of river inflows, net evaporation and dam water releases. We found a reasonably strong relationship between changes in river flow and reservoir storage, while changes in net evaporation do not seem to have affected storage trends significantly. We infer that reservoir water releases are unlikely to be the dominant driver of the three-decadal trends in reservoir storage for the majority of basins. However, we acknowledge that this particular conclusion is based on deductive rather than observational evidence, and would benefit from corroboration for any individual reservoir using actual release records, which often exist but not publicly available."

On the other hand, although we cannot access global reservoir water release data, we carried out human impact analysis using global water withdrawal estimates (Huang et al., 2018) as an approximation of water releases. We included this analysis in L365-392 in the Discussion to support our conclusion on the impact of human activity on reservoir storage trends:

"There is currently no global hydrological model capable of estimating the impact of historical operational water management at the reservoir or basin level with meaningful accuracy. However, to get an indication of the potential impact of human activity and associated reservoir water releases on reservoir storage changes, we analysed the global water withdrawal estimates produced by Huang et al. (2018). The gridded monthly withdrawal time series for 1971-2010 were spatially and temporal downscaled from 5-year temporal resolution estimates from FAO AQUASTAT and USGS, which were based on national assessments and surveys (Huang et al, 2018). Their estimates provide separate water withdrawal estimates for irrigation, hydroelectricity, domestic, livestock, manufacturing and mining, respectively. The withdrawals are from reservoirs, rivers and groundwater, and as such cannot be compared directly to reservoir water release, but may provide useful context. We calculated total withdrawals from the six sectors combined and examined trends from 1984-2010 at basin scale. The results show that significant increasing trends in withdrawals in 78 basins, mainly in South America, Africa and Asia, and significant decreasing trends in 29 basins in Europe, Australia, and parts of Northern America, but noting that the magnitude of withdrawals varied widely compared to, for example, total river inflows or reservoir capacity (Fig. 12; Fig. S6). The global pattern in water withdrawals trends is different from the spatial patterns in precipitation, inflow, and storage (Fig. 7). We calculate that (either significant or non-significant) water withdrawal trends are associated with about equal numbers of increasing and decreasing water storage trends (Table 2). By contrast, rainfall and inflow trends lead to a change in storage in the same direction for around 80% of basins (Table 2). These observations further support the notion that climate trends rather than water withdrawals are primarily responsible for the observed trends in reservoir storage. Nonetheless, there are basins where storage trends may have been influenced by water withdrawals. For example, inflows increased by 43% in northern Venezuela

while total reservoir storage deceased by 15%, conceivably because water withdrawals tripled from 1984-2010 (Fig. 7 and 12). A comparable scenario also occurred in coastal basins in Angola, Mozambique, Tanzania and Kenya. Storages in Iran, Turkmenistan and northern India decreased by an average 33%, which may be attributed to an unknown combination of reduced inflows (-6% - -21%) and increased withdrawals (+42% - +50%), although it is noted that a large fraction of withdrawals is from groundwater in some of these basins."

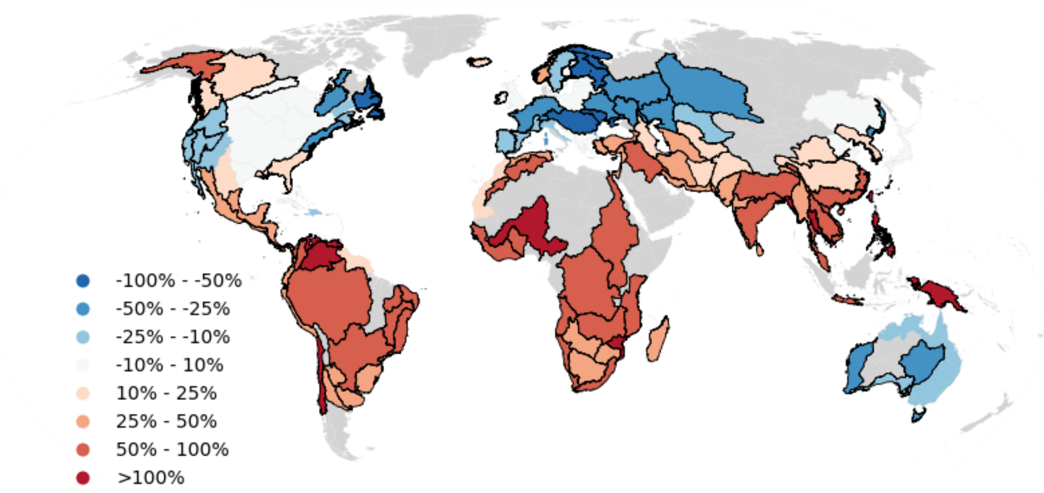

Figure 12 Linear trends in annual water withdrawal between 1984–2010 (grey shade: no reservoir data; black outlines: trend significant at p<0.05). Note that the magnitude of withdrawals varies strongly between basins.

Table 2 Comparison of trends in reservoir storage reconstruction against climate variability and human activities.

Drivers	Reservoir Storage				
	Trend (number of basins)	Significant increase	Significant decrease	Increase	Decrease
Water withdrawal	Significant increase (78)	18	9	36	42
	Significant decrease (29)	6	6	12	17
	Increase (93)	21	10	43	50
	Decrease (41)	10	10	19	23
Climate (precipitation and inflow)	Significant increase (23)	11	0	21	2
	Significant decrease (14)	2	4	5	9
	Increase (70)	31	6	54	16
	Decrease (64)	6	21	11	53

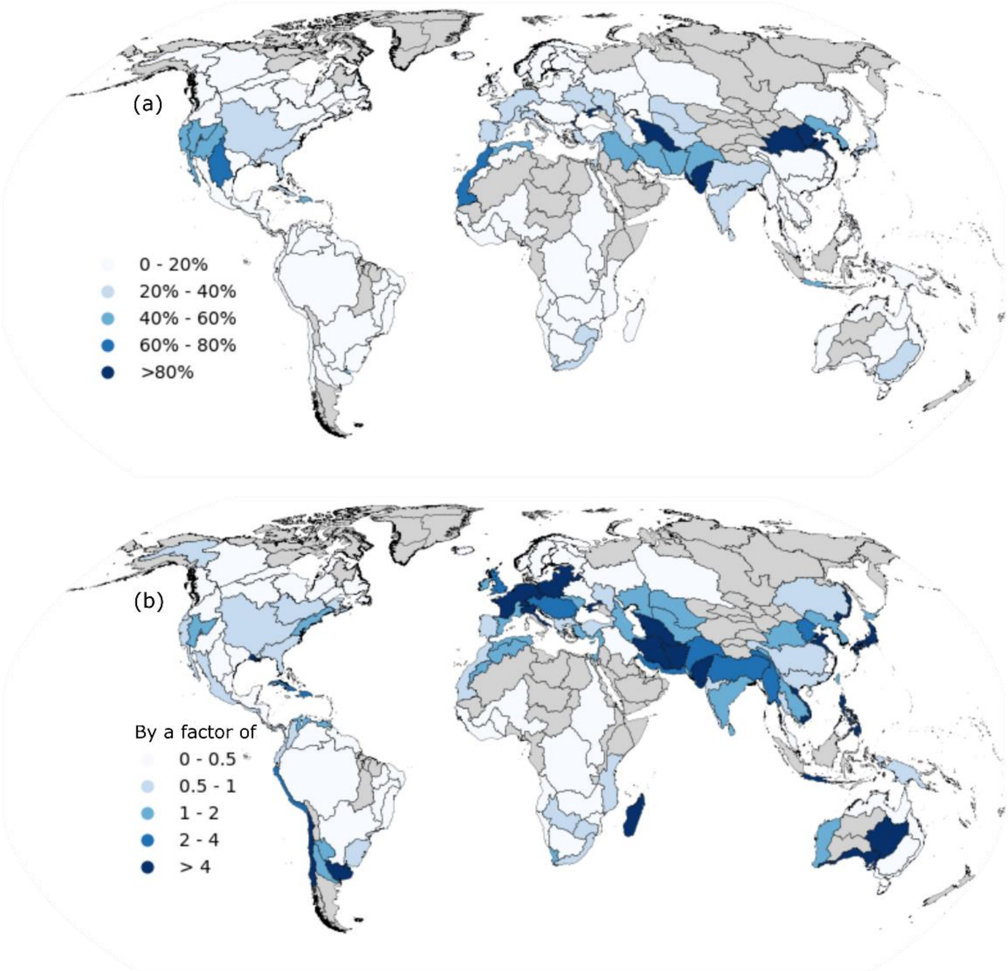

Figure S6 The ratio of the average annual total water withdrawal and total river inflow (a) and total water withdrawal and total reservoir capacity (b) in each basin. Annual withdrawal is a small or modest fraction of annual inflow in a majority of basins and water withdrawal is probably mainly from non-storage sources (i.e. groundwater, direct river extraction and reuse) in some basins such as in central Eurasia.

[2] Wang, J., Sheng, Y., & Wada, Y. (2017). Little impact of the Three Gorges Dam on recent decadal lake decline across China's Yangtze Plain. Water Resources Research, 53(5), 3854-3877.

[3] Gudmundsson, L., Boulange, J., Do, H.X., Gosling, S.N., Grillakis, M.G., Koutroulis, A.G., Leonard, M., Liu, J., Schmied, H.M., Papadimitriou, L. & Pokhrel, Y. (2021). Globally observed trends in mean and extreme river flow attributed to climate change. Science, 371(6534), 1159-1162.

[4] Cooley, S. W., Ryan, J. C., & Smith, L. C. (2021). Human alteration of global surface water storage variability. Nature, 591(7848), 78-81.

[5] Huang, Z., Hejazi, M., Li, X., Tang, Q., Vernon, C., Leng, G., Liu, Y., Döll, P., Eisner, S., Gerten, D. & Hanasaki, N. (2018). Reconstruction of global gridded monthly sectoral water withdrawals for 1971–2010 and analysis of their spatiotemporal patterns. Hydrology and Earth System Sciences, 22(4), 2117-2133.

R2C2) The analysis of reservoir reliability, resilience, and vulnerability (lines 172-189) is a good extension to the estimated reservoir storage dynamics. The concepts and calculations in this part could be better introduced by using a real reservoir as an example, perhaps a well-known reservoir with good data availability. Also, how did the authors determine the time length of failure events (line 178) determined? How does the value of this factor vary among different reservoirs or basins? What is the unit of resilience (line 185)?

We thank the reviewer for this suggestion. The time length of failure event is defined as the number of continuous months when the storage level drops below 10% lowest value (see "Duration Time (month)" in Table 3). The time length of failure event is converted to the resilience index using Eq.6, following previous studies (Hashimoto et al. 1982; Kjeldsen and Rosbjerg 2004). The resilience index has a unit of month^{-1} and ranges from 0 to 1 in this study. A lower index value indicates a slower recovery rate (weakened resilience). We included a worked example for an actual reservoir to introduce reliability, resilience, and vulnerability (Fig. S5 and Table S5 in the supplementary material). We also added some text to describe this example in L225-229 in the revised manuscript:

> *"A worked example is shown for the Toledo Bend Reservoir (Texas, USA) (Fig. S5 and Table S5). Four failure events occurred during 1984–2000 and three during 2000–2015. Before 2000, it took an average of three months to recover from failure, with an average deficit volume of 357 GL. After 2000, it took an average of 10.5 months with a larger average deficit volume of 498 GL (Fig. S1). It follows that resilience was reduced (resilience index 0.12 vs. 0.33) and vulnerability increased (deficit volume 498 vs. 357 GL) when compared to the years before 2000 (Table S5)."*

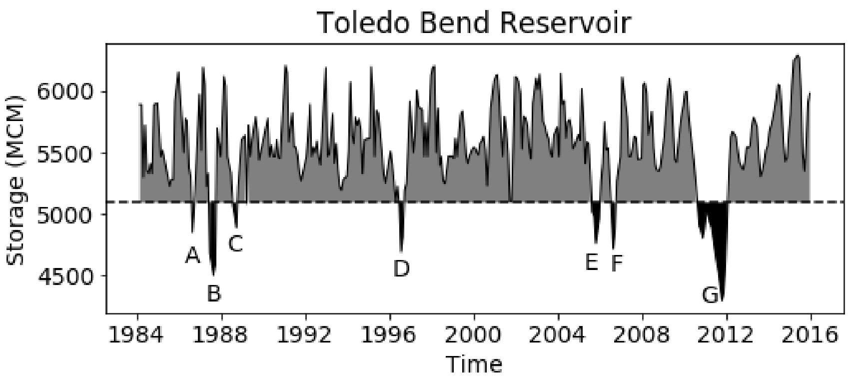

Figure S5 Example storage time series showing the definition of resilience and vulnerability (black shaded: unsatisfactory state; grey shaded: satisfactory state, dashed line: 10% threshold; letters: labeled failure events).

Table S5 The statistics of resilience and vulnerability for the reservoir in Fig. S5.

Period	1984-2000				2000-2015		
Failure Event	A	B	C	D	E	F	G
Duration Time (month)	2	4	3	3	5	3	18
Resilience (1/month)	0.33				0.12		
Deficit Volume (GL)	239	589	202	399	329	373	792
Vulnerability (average deficit volume)	357				498		

R2C3) Field observations and modeling studies have shown that evaporative loss from reservoir surface can be quite significant, especially for reservoirs in arid and semi-arid regions. This seems to be contradictory to some conclusions from this study (lines 265-266, 307-308 and 311).

Thank you for this comment. We also found evidence that evaporative losses from reservoirs are large in arid and semi-arid regions, but they did not explain long-term trends. Large evaporative losses tend to affect seasonal storage dynamics but this does not necessarily mean that *trends* in evaporation explain a long-term *trend* in reservoir storage. As an aside, evaporative losses also tend to be less significant in large reservoirs due to their greater depth (Mady et al., 2020).

We included this statement in L351-354 in the revised manuscript:

> *"Mady et al. (2020) and various other authors found that evaporative losses can account for much of the loss of water from small reservoirs (e.g., <0.1 km^2) in semi-arid regions. However, this does not necessarily mean that trends in evaporation can explain trends long-term trends in storage, especially for the mostly larger (and deeper) reservoirs considered here."*

In addition, we show the validations of trend analyses of net evaporation against Zhao and Gao (2019), referring to our response to R1C3 for full details.

[6] Mady, B., Lehmann, P., Gorelick, S. M., & Or, D. (2020). Distribution of small seasonal reservoirs in semi-arid regions and associated evaporative losses. Environmental Research Communications, 2(6), 061002.

Technical Corrections

R2C4) Figures 2-3. No need to use the second y-axis.

We changed it to use the same vertical scale.

R2C5) Line 171. Remove the comma.

We removed the comma in this sentence.

Reviewer #3 Comments:

This study presents a multi-satellite remote sensing approach to understand long term storage changes in over six thousand reservoirs around the world. The authors combine well-established remote sensing based reservoir monitoring techniques to build monthly time series of storage variations. These variations are then synthesized with streamflow to provide insight into long term trends. This is an important study that pushes the boundaries of our understanding of global reservoir storage variations and explores possible drivers of the observed changes. However, I have two major concerns and several minor concerns that should be addressed before publication.

R3C1) First, I am unsure of the value of using long term trends to characterize reservoir storage as increasing or decreasing between 1985 and 2015 (as in lines 210-240). Figure 2 suggests that reservoirs of these sizes can go through shorter, but still multi-year periods of increased and deceased storage throughout this time period. For example, Fort Peck and Fairbairn Reservoirs show ~10 year long oscillations in storage that are not easily characterized by simply increasing or decreasing trends.

We agree. For our long-term analysis, we not only calculated whether there were increasing or decreasing trends from 1984-2015, but also tested whether these trends were significant or not, using the Mann-Kendall trend test ($p<0.05$). The red and blue points in Fig.4 and the basins with black outlines in Fig.7 showed significant increasing or decreasing trends in storage. The Fort Peck and Fairbairn Reservoirs show non-significant trends according to the Mann-Kendall trend test. In contrast, for example, the change in total storage is significant in Colorado River Basin (including 76 reservoirs) of southwestern USA, although there are decadal oscillations in storage (Fig.R4).

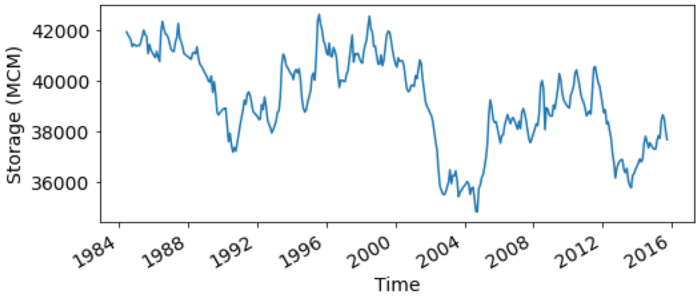

Figure R4 Total monthly storage dynamics with significant decreasing trend in one basin of southwestern USA.

R3C2) Second, I am unconvinced of the conclusion that human intervention is an insignificant contribution to storage variability. According to equation 8, changes in storage are related to Qin and Qout (assuming small E). One could argue that any change in storage is due to human alteration of Qout, because without modification of Qout (relative to Qin) there would be no storage variation at all. Without some quantification of the drivers of Qout (hydropower demand, irrigation needs, etc.) I find it hard to make an argument for Qin to be the dominant driver with only what has been quantified in this study. Perhaps an alternate way to frame your findings is that Qin can be used as a good predictor of positive or negative reservoir storage variations.

We thank the reviewer for this constructive suggestion. We agree that the conclusions on the influence of human intervention are not as robust as we might like, given the lack of publicly available records on releases. However, we believe our logic to deduce the role of human activity is sound and our conclusions sufficiently cautious, especially in the revised manuscript. Besides, we did further analysis on human impacts using global water withdrawal estimates (Huang et al., 2018) as an approximation of water releases. This analysis provides additional evidence on the conclusion that climate trends rather than water withdrawals are primarily responsible for the observed trends in reservoir storage. Please refer to our response to R2C1 for full details.

Line comments:

R3C3) Lines 65-79: The limitations of past efforts and techniques are summarized well here, but how this study overcomes these limitations and provides something new should also be given a sentence or two here.

Thank you, we modified the last paragraph (L79-94) in the introduction to highlight the advancement of this study over previous ones. Please refer to our response to R1C3 for full details.

R3C4) Line 125: This figure could use a legend describing what the colors and inner and outer rings area.

We added a legend to this figure.

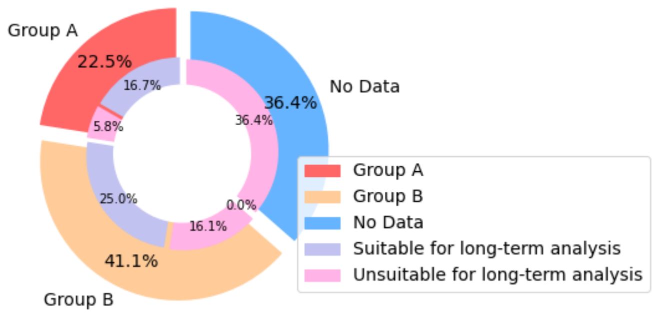

Figure 1 The total storage capacity in Group A (red) and B (brown) and left unaccounted (blue) and the combined capacity of reservoirs for which the data were suitable (teal) or unsuitable (pink) for long-term analysis.

R3C5) Line 150: Would reservoirs constructed during the study period have an impact on the quantified Qin for older reservoirs?

Thank you for this question. The hydrological models do not simulate the effect of reservoirs and river operations on flows and therefore would not reflect any such changes.

R3C6) Line 171-190: I was confused by the methods for calculating reliability, resilience, and vulnerability. How does assuming 90% reliability simplify the calculations? Why is this a reasonable assumption?

We apologize for the confusion. We included an actual reservoir as an example to introduce reliability, resilience, and vulnerability. Please see our response to R2C2 for full details.

R3C7) Line 205: The two vertical axis on Figure 2 and 3 need to be equal for each subplot. As it is now, only correlation is apparent, but it would be much more realistic to plot the observed and predicted on the same vertical scale to get a realistic sense of the errors.

Thank you for this suggestion. We changed it to use the same vertical scale in Fig. 2 and 3.

R3C8) Line 342-350: This paragraph felt a little out of place here. Maybe consider moving the content to the methods section.

Agree. We moved this paragraph to the method section in L144-150 in the revised manuscript.