Authors' reply to Anonymous Referee #1

We thank the referee for his/her helpful comments to and kind words about our manuscript. We have incorporated most of the comments into the revised manuscript, as detailed below. Referee comments appear in italics, our reply in normal font, and changes to the manuscript in blue.

Review of: Intercomparison of freshwater fluxes over ocean and investigations into water budget closure By Gutenstein et al. This paper presents an inter-comparison of five recent satellite-based and one re-analysis E-P data sets. The different data-sets and the assumptions behind them are described. The different components of the hydrological cycle are presented separately. This is a well written paper, which presents a valuable contribution to the climate community. I have little to add to this paper, which in my opinion is almost ready for publication in its current form.
Thank you!

The few and very minor comments I have are:
In the introduction I missed a section motivating the study from a climate change perspective like you added to the "Final Comments" section. Monitoring trends in the hydrological cycle is of great importance under uncertain changing climate conditions. In that aspect I would like to point the authors to some recent papers on the topic [1-3].
The referee is right. We added the following paragraph to the introduction:
„At long temporal and/or large spatial scales, the increases in E and P with rising global temperature are relatively small (2-3%K^{-1}) and are constrained by the energy budget. At smaller scales (less than approximately 4000 km and/or 10 years) these changes can be much larger (or smaller) due to dynamical contributions (Dagan et al., 2019; Yin and Porporato, 2019; Allan et al., 2020). The nature and extent of these changes, which affect the livelihoods of many millions of people, are difficult to model due to various counteracting influences such as forcing by clouds and aerosols, or land use change (Allan et al., 2020). Close monitoring of E and P by (satellite) observations thus yields an important contribution to a better understanding of impacts of climate change at regional and local scales."

The inter-comparison presented here is a very nice and useful framework also for comparing with climate models [4-6] such as CMIP6. I suggest to propose it in the "Final Comments" section (or elsewhere) for future work. It could enlarge the connection of this work to climate change research.
That is also a good suggestion. We added the following statement to the Final Comments section:
„The presented framework is based on co-variation of water cycle components and global water budget constraints. We applied it to the inter-comparison of satellite observations, but it can also be used for climate model assessments such as the Coupled Model Intercomparison Project CMIP (see, e.g., Held and Soden, 2006; Liepert and Previdi, 2012; Knutti and Sedlacek, 2013; Allan et al., 2020)."

L96: Could you please elaborate on how wind speed is calculated based on temperature (BT) measurements?
Wind speed cannot be directly derived from passive satellite observations. However, wind effects alter the roughness and emissivity of the ocean surface, particularly affecting the 19 and 37 GHz channels (see, e.g., Meissner and Wentz, 2012). Just like heat fluxes (and evaporation rate), near-surface wind speed cannot be determined from passive microwave observations in scenes with heavy precipitation. Scatterometers, active microwave instruments, are capable of retrieving wind speed under rainy conditions and their data are used in conjunction with passive wind observations by several of the retrieval algorithms presented in the manuscript (see Section 2). For more details of wind speed data used in the various E (LHF) retrievals, we refer the referee to the literature cited in Section 2.1.
[Meissner, T. and Wentz, F.J.: The Emissivity of the Ocean Surface Between 6 and 90 GHz Over a Large Range of Wind Speeds and Earth Incidence Angles, IEEE Trans. Geosci. Rem. Sens., 50(8), doi: 10.1109/TGRS.2011.2179662, 2012.]

In addition, If E estimates are based on BT measurements, which are more accurate in clear sky conditions than in cloudy sky conditions, wouldn't that cause a bias? How is it calculated in cloudy (and rainy) conditions? If it is only calculated in clear sky conditions, wouldn't the E estimations be biased high (as in cloudy and rainy conditions the evaporation is lower)?
The retrieval of E, LHF, and wind speed from passive microwave observations is not possible in scenes with heavy precipitation. This is, indeed, expected to lead to a positive bias, but it has not been quantified. Interestingly, global mean E from ERA5 exceeds all satellite-based estimates, despite the fact that all sky conditions are included. We inserted the following statement after line 527 (Section 5): "It is interesting to note that satellite-based E are very likely biased high by the removal of scenes with strong precipitation (where the retrieval of WS, LHF, and E is not possible). In this light, the difference in E between ERA5 and the satellite-based retrievals should actually be larger than observed in Fig.3, as monthly mean E is determined from all sky conditions in reanalysis. As the OAFlux and SEAFLUX blend satellite estimates with continuous background fields (Sect. 3), these algorithms should be less impacted by such sampling biases."

L368: is the largest deviation in E estimations in the tropics due to the large (and optically thick) cloud cover?
Passive microwave instruments "observe" humidity in the total column with only limited information on the vertical structure. Retrievals of near-surface humidity are most accurate under typical conditions of moisture stratification. Biases arise when the vertical stratification of moisture departs strongly from these typical conditions. Thus, estimates of near-surface humidity and, subsequently, E often vary between products (Roberts et al., 2019). Note that the most recent J-OFURO and SEAFLUX products include additional a priori information on moisture stratification within the retrieval algorithms to mitigate these issues. Accounting for this improves the consistency of retrieval results appreciably compared to *in situ* measurements (Roberts et al., 2019).
We added this information to the manuscript by modifying line 368 to:
"(...) while the largest deviations appear mainly in the tropics. This is due to the frequent occurrence of weather conditions in which the moisture stratification departs substantially from typical conditions to which the retrieval algorithms of near-surface moisture are tuned. Accounting for this dependence on moisture stratification, as in the SEAFLUX and J-OFURO algorithms, improves retrieval results appreciably compared to *in situ* measurements (Roberts et al., 2019)."

L437: I think that the correlation does not decrease when Delat Qocean is not considered because there is basically no correlation even when it is considered. So, it can't get any lower than that. Is that correct?
In principle, the referee is correct. However, the statement in the manuscript is not quite accurate, as there is appreciable improvement in R^2 for J-OFURO-G and IFREMER-G for yearly means and monthly anomalies. With the updated SEAFLUX3 statistics, Table 3 reads:

Data set	Monthly mean	Yearly mean	Monthly anomaly
HOAPS-4.0	0.03	0.00*	0.06
J-OFURO3 - GPCP-1DD	0.16	0.31	0.22
IFREMER4.1 - GPCP-1DD	0.13	0.23	0.20
OAFlux3 - GPCP-1DD	0.14	0.01*	0.11
SEAFLUX3 - GPCP-1DD	0.17	0.02*	0.12
ERA5	0.86	0.86	0.83

Ignoring the contribution of $\nabla(vq)$ yields:

Data set	Monthly mean	Yearly mean	Monthly anomaly
HOAPS-4.0	0.01*	0.00*	0.01*
J-OFURO3 - GPCP-1DD	0.12	0.19*	0.08
IFREMER4.1 - GPCP-1DD	0.12	0.17*	0.08

OAFlux3 - GPCP-1DD	0.11	0.00*	0.01*
SEAFLUX3 - GPCP-1DD	0.11	0.00*	0.00*
ERA5	0.42	0.57	0.55

Hence, including the $\nabla(vq)$ contribution not only improves correlations of yearly means and monthly anomalies, but also yields more cases where the correlation is significant (not marked by an asterisk). We corrected the statement, while also changing the notation of ∇Q to $\nabla(vq)$, as recommended by referee #2.

"Including the contribution of ∇(vq) improves the correlation appreciably for ERA5, as mentioned above. For satellite data the correlation also improves, particularly for yearly means and monthly anomalies of IFREMER-G and J-OFURO-G (not shown)."

Technical comments:
L208: us–>use
Corrected.

You alter between italic and non-italic in P, E and E-P. I think it should all be italic.
In accordance with convention, all symbols are written in italic. The fact that the abbreviations chosen for evaporation, precipitation, and freshwater flux are the same as their respective symbols (see Table 2) causes E, P, and E-P to be written in italic in some cases and non-italic in others. This should not lead to confusion, so we would like to keep the notation this way in the paper.

References

[1] Allan, R. P. et al. Advances in understanding large-scale responses of the water cycle to climate change. Annals of the New York Academy of Sciences (2020).
[2] Dagan, G., Stier, P. & Watson-Parris, D. Analysis of the atmospheric water budget for elucidating the spatial scale of precipitation changes under climate change. Geophysical Research Letters (2019).
[3] Yin, J. & Porporato, A. Looking up or looking down? Hydrologic and atmospheric perspectives on precipitation and evaporation variability. Geophysical Research Letters 46, 11968-11971 (2019).
[4] Liepert, B.G. & Previdi, M. Inter-model variability and biases of the global water cycle in CMIP3 coupled climate models. Environmental Research Letters 7, 014006 (2012).
[5] Knutti, R. & Sedláček, J. Robustness and uncertainties in the new CMIP5 climate model projections. Nature Climate Change 3, 369 (2013).
[6] Held, I. M. & Soden, B. J. Robust responses of the hydrological cycle to global warming. Journal of Climate 19, 5686-5699 (2006).

We thank the referee for pointing us to these papers and have incorporated them into the revised manuscript. In particular, we added the global total fluxes estimated in [1] to Table 4. Please note that [6] is already cited in the first version of the manuscript.

The updated Table 4:

Table 4. Estimates of ocean total E and P, land and ocean total $E-P$, net transport of water vapor, and continental runoff given in 10^3 km^3 yr^{-1}. The upper three rows contain results from this study, the lower four those from earlier investigations. ERA5 estimates are calculated from ensemble mean data, the standard deviation (std) is derived from ensemble statistics. The satellite-based data sets used in our study were averaged to obtain the mean and std of observed (Obs.) E_{ocean} and P_{ocean}, and the range is given in the third row. Net water vapor flux divergence over land (∇Q_{land}) and ocean (∇Q_{ocean}) and continental runoff R are given in the last three columns. The estimates from the study by Rodell et al. (2015) are separated into observations (obs.) and model-optimized observations (opt.), see the text for details.

	E_{ocean}	P_{ocean}	$(E-P)_{ocean}$	$(E-P)_{land}$	∇Q_{land}	∇Q_{ocean}	R
ERA5	467 ± 1	426 ± 2	43 ± 2	-44 ± 0.4	-43 ± 0.2	31 ± 0.2	42.1
Obs. mean $\pm$ std	425 ± 20	360 ± 25	52 ± 13	–	–	–	–
Obs. range	397–453	335–384	35–65	–	–	–	–
Oki and Kanae (2006)	436.5	391	45.5	-45.5	-45.5	45.5	45.5
Trenberth and Asrar (2014)	413	373	40	-40	-40	40	40
Rodell et al. (2015) obs.	410 ± 36	385 ± 39	25 ± 53	-45 ± 9	-43 ± 8	47 ± 19	50 ± 7
Rodell et al. (2015) opt.	450 ± 22	403 ± 22	47 ± 31	-46 ± 7	-46 ± 4	46 ± 2	46 ± 4
Allan et al. (2020)	480 ± 48	434 ± 43	46 ± 65	-46 ± 14	-46 ± 5	46 ± 5	51 ± 3

And we added the following lines to Sect. 4.6:

"The global total fluxes estimated by Allan et al. (2020) derive from Rodell et al. (2015), but following the recommendation by Stephens et al. (2012), E_{ocean} and P_{ocean} were both increased by 30.10^3 km^3 yr^{-1} to improve the agreement with energy constraints, yet keeping land-ocean fluxes constant. These increases are larger than the $\pm 22 \cdot 10^3$ km^3 yr^{-1} uncertainty on E_{ocean} and P_{ocean} estimated by Rodell et al. (2015) based on the optimized method and so a more modest increase of $20 \cdot 10^3$ may be appropriate. This would produce fluxes of $E_{ocean} = 470 \cdot 10^3$ km^3 yr^{-1} and $P_{ocean} = 424 \cdot 10^3$ km^3 yr^{-1} that are quite close to ERA5 estimates (R. Allan, personal communication, Oct. 2020)."