Reply to the reviewer's comments on the manuscript entitled "A combination of Time-Variable Gravity Field Solutions from Multi-Satellite Datasets (1993-2024) via Constrained Collocation Model" with essd-2026-84

Reply to the Editor and Reviewers#:

First, we sincerely thank the Editor and the reviewers for their constructive comments and suggestions, which have helped us improve the manuscript. We have revised the manuscript carefully and provide a point-by-point response below. All revisions are marked in red in the revised manuscript. In the following response, C and R indicate "Comment" and "Response", respectively.

During the revision, we reconsidered the terminology and found that "Constrained Collocation Model" describes our method more accurately than "Least-Squares Collocation", because the parameter estimation is based on the criterion constrained by Tikhonov regularization rather than least-squares. We therefore revised the title from "A combination of Time-Variable Gravity Field Solutions from Multi-Satellite Datasets (1993-2024) via Least-Squares Collocation" to "A Combination of Time-Variable Gravity Field Solutions from Multi-Satellite Datasets (1993-2024)" via Constrained Collocation Model", and replaced "Least-Squares Collocation" with "Constrained Collocation" throughout the manuscript.

In addition, we have addressed the editorial office's technical checks by adding the full name of IMBIE in the summary, replacing the Zenodo link with the corresponding DOI, and revising the color schemes of the relevant figures (Figs. 3, 4, 6, 8, 9, and 11) for color-vision accessibility.

Reply to Comments of Reviewer #1

C1: This study presents gap-free monthly gravity field solutions up to degree and order 60 for January 1993-December 2024 using constrained least-squares collocation (LSC), which jointly performs solution combination and denoising without explicit filtering. The proposed product outperforms existing reconstructions based on hydrometeorological forcing and static spatial modes, yielding smaller sea-level budget misclosures, reduced basin-scale water-balance errors, and stronger agreement with IMBIE (Ice Sheet Mass Balance Inter-comparison Exercise) estimates. Overall, the work is meaningful, the methodology is robust, and the results are convincing. I therefore recommend acceptance after some revisions.

R1: Thank you for your constructive comments! We have carefully revised the manuscript according to your comments, and detailed responses are provided below. During revision, we also renamed the original Least-Squares Collocation (LSC) as the Constrained Collocation Model (CCM), as the method is based on the criterion constrained by Tikhonov regularization rather than conventional least-squares.

C1.1: Line 75: The time span "1984 to 2022" should be revised to "1984 to 2023."

R1.1: Thank you! The time span has been revised to "1984 to 2023".

C1.2: Table S2 in the supplementary file presents complex pseudocode and is not visually clear; could the authors replace it with a flowchart to improve readability and better illustrate the workflow?

R1.2: Thank you for your valuable comments! We have replaced the pseudocode in Tab. S2 of the original manuscript with a flowchart in Fig. S5 of the revised manuscript to improve readability and clarify the workflow.

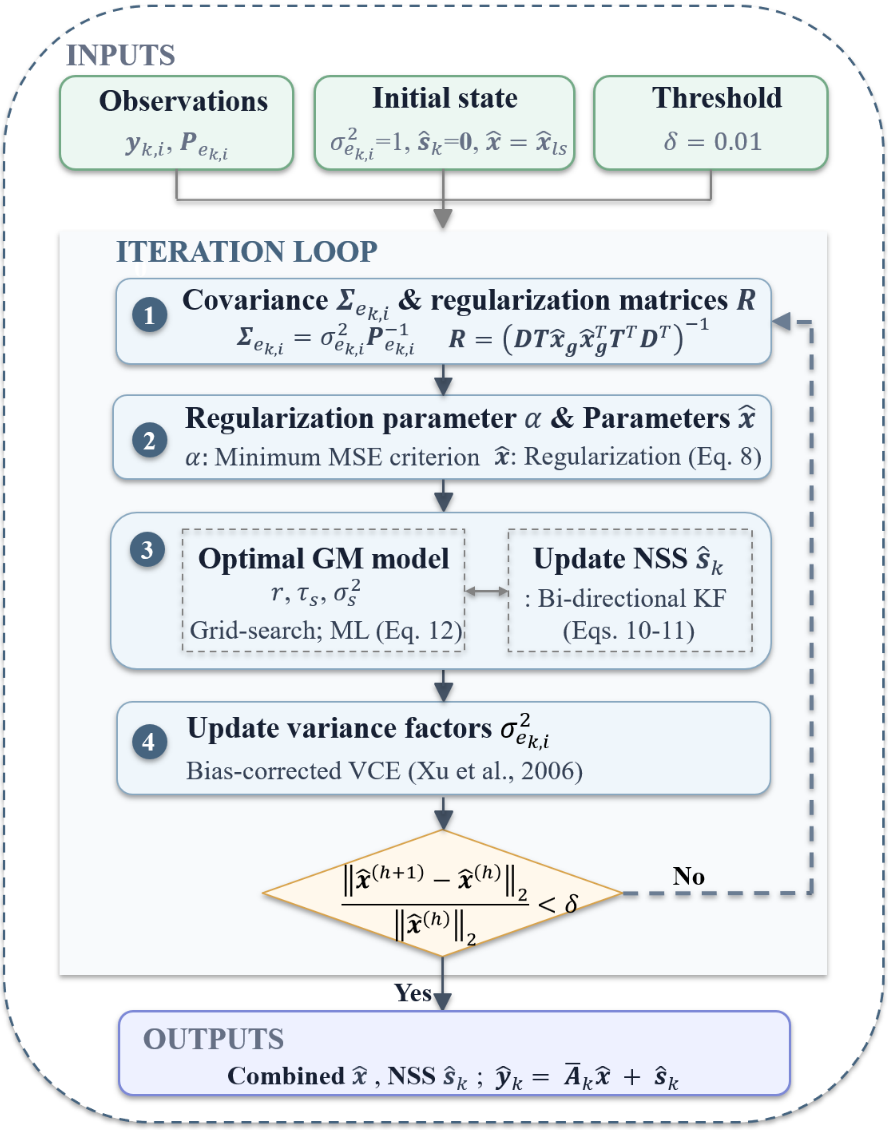

Figure S5: Flowchart for iteratively combining TVGFS (The symbols are introduced in Tab. S2).

C1.3: The labels/captions of Fig. 3(d) and Fig. 3(e) should be checked to confirm whether they are reversed.

R1.3: Thank you for your careful comment! We have checked Fig. 3(d) and Fig. 3(e) and reversed the labels.

C1.4: Notation and abbreviations should be standardized throughout the manuscript (e.g., TVGF/TVGFS, d/o, SHC/SHCs, NSS/NSSs, capitalization of defined terms) to avoid ambiguity.

R1.4: Thank you! We have standardized the notations and abbreviations throughout the manuscript.

C1.5: A short "Notation" box (or one supplementary table) listing core symbols (, , , , , , etc.) and their dimensions would improve readability.

R1.5: Thank you! We have added a short "Notation" box as Table S2 in the supplementary file to list core symbols and their dimensions for improving readability.

Table S2. Notation box for core symbols used in the Methodology.

Symbol	Description	Dimension	Initial appears in
$k,\ i$	Month index and TVGFS index.	Scalars	Eq. (1)
$l,\ m$	Spherical harmonic degree and order.	Scalars	Eq. (1)
$M_{k,i}$	Number of SHCs provided by the i-th TVGFS at month k.	1×1	Eq. (3)
M	Number of SHCs in the target combined TVGFS.	1×1	Eq. (3)
q_k	Number of individual TVGFS available at month k.	1×1	Eq. (2)
N_k	Total number of observations at month k	1×1	Eq. (2)
G	Number of global grids for one month	$G\times1$	Eq. (7)
$y_{k,i}^{lm}$	SHC from the i-th TVGFS for degree l and order m at month k.	1×1	Eq. (1)
Δt_k	Time interval between month k and the reference epoch.	1×1	Eq. (1)
$a_{0,i}^{lm},\ldots,\ a_{6,i}^{lm}$	Deterministic coefficients: constant, trend, acceleration, annual and semi-annual.	1×1	Eq. (1)
$s_{k,i}^{lm}$	Non-seasonal signal (NSS) for the corresponding SHC.	1×1	Eq. (1)
$e_{k,i}^{lm}$	Observation error corresponding to $y_{k,i}^{lm}$.	1×1	Eq. (1)
$\boldsymbol{y}_{k,i}$	Observation vector stacking SHCs from the i-th TVGFS at month k.	$M_{k,i}\times1$	Eq. (2)
$\boldsymbol{y}_k$	Combined observation vector stacking all available TVGFS at month k.	$N_k\times1$	Eq. (2)
$\boldsymbol{A}_{k,i}$	Design matrix for the i-th TVGFS harmonic model.	$M_{k,i}\ \times7M$	Eq. (4)
$\boldsymbol{A}_k$	Stacked design matrix for all available TVGFS at month k.	$N_k\times7M$	Eq. (2)
$\boldsymbol{B}_{k,i}$	Selection/extension matrix mapping target SHCs to the i-th TVGFS support.	$M_{k,i}\ \times\ M$	Eq. (3)
$\boldsymbol{B}_k$	Stacked selection matrix for all available TVGFS at month k.	$N_k\times\ M$	Eq. (2)
$\boldsymbol{x}$	Combined deterministic parameter vector for all target SHCs.	$7M\times1$	Eq. (2)
$\boldsymbol{s}_k$	Combined NSS vector at month k.	$M\times1$	Eq. (2)
$\boldsymbol{e}_k$	Combined observation-error vector at month k.	$N_k\times\ 1$	Eq. (2)
$\boldsymbol{P}_{e_k}$	Block-diagonal weight matrix of combined observation errors.	$N_k\times N_k$	Eq. (2)
$\boldsymbol{\Sigma}_{e_k}$	Covariance matrix of combined observation errors.	$N_k\times N_k$	Eq. (2)
$\sigma_{e_i}^2$	Time-invariant variance factor for the i-th TVGFS.	1×1	Eq. (2)
$\boldsymbol{\Phi}_{k,j}$	State transition matrix from epoch j to k in the GM model.	$M\times M$	Eq. (5)
r	Order of multi-order GM model	1×1	Eq. (5)
τ_s	Correlation time of the GM process.	1×1	Eq. (6)
σ_s^2	Variance component of the GM process.	1×1	Eq. (6)
$\boldsymbol{w}_k$	GM process noise at month k.	$M\times1$	Eq. (5)
$\boldsymbol{Q}_k$	Covariance matrix of GM process noise.	$M\times M$	Eq. (6)
$\boldsymbol{D}$	Diagonal degree-dependent scale matrix.	$M\times M$	Eq. (6)
$\bar{\boldsymbol{s}}_k$	Predicted NSS before the Kalman update.	$M\times1$	Eq. (9)
$\boldsymbol{\Sigma}_{\bar{s}_k}$	Predicted/updated covariance matrix of NSS at month k.	$M\times M$	Eq. (9)
α	Regularization parameter selected by the minimum-MSE criterion.	1×1	Eq. (7)
$\boldsymbol{R}$	Regularization matrix propagated from the spatial expression of $\boldsymbol{x}$.	$7M\times7M$	Eq. (7)
$\boldsymbol{T}$	Spherical harmonic analysis matrix used in constructing $\boldsymbol{R}$	$7M\times7G$	Eq. (7)
$\boldsymbol{x}_g$	Spatial expression of the deterministic parameter vector $\boldsymbol{x}$.	$7G\times1$	Eq. (7)
$\boldsymbol{K}_k$	Kalman gain matrix for updating NSS at month k.	$M\times N_k$	Eq. (11)
L	Cost function minimized to determine r, τ_s, and σ_s^2.	1×1	Eq. (12)
$\bar{\boldsymbol{v}}_k$	Prediction residual used in the GM-model cost function.	$N_k\ \times1$	Eq. (12)
$\boldsymbol{C}_k$	Covariance matrix of $\bar{\boldsymbol{v}}_k$.	$N_k\times N_k$	Eq. (12)
$\boldsymbol{M}_{\hat{x}}$	MSE matrix of the combined deterministic parameter vector $\hat{\boldsymbol{x}}$.	$7M\times7M$	Eq. (12)

C1.6: Section titles could be revised for parallel style and precision (e.g., "Function Model" → "Functional Model"; "Result Evaluations" → "Results and Evaluation").

R1.6: Thank you! We have revised the section titles to ensure a parallel style and improve precision throughout the manuscript.

C1.7: Please report the analysis period consistently in all figure captions (e.g., 1994-2020, 2002-2024), as some captions currently list only metrics without the corresponding time span.

R1.7: Thank you for your detailed suggestion! We have added the corresponding analysis periods to all figure captions to improve clarity.

C1.8: A quick language polish is recommended to fix recurrent minor issues, including subject-verb agreement (e.g., results indicates → results indicate) and number/possessive form (e.g., basin's → basin-scale or basins'). These edits do not affect the methodology but will noticeably improve the manuscript's professional quality.

R1.8: Thank you for your helpful suggestions. We have thoroughly proofread the manuscript and corrected the recurring minor language issues, including subject-verb agreement and number/possessive forms, to improve the overall professional quality of the paper.

C1.9: One concern is why IGG-SLR-HYBRID was adopted in the combination, whereas IGG-SLR-DORIS was used in the comparison. Since these two models seem to be based on similar methodologies and exhibit comparable performance over 1994–2020, the authors may wish to clarify whether it would be sufficient to include only one of them.

R1.9: Thank you for your constructive comments. The key difference between IGG-SLR-HYBRID and IGG-SLR-DORIS is that the latter additionally includes Low Earth Orbit (LEO) observations. In our combination framework, LEO information has already been incorporated through the Tongji-LEO2021 model. Therefore, to avoid reintroducing LEO information, we only used IGG-SLR-HYBRID in the combination step to isolate the SLR contribution. For the comparison experiment, we used IGG-SLR-DORIS because both IGG-SLR-DORIS and our combined model contain GRACE, SLR, and LEO information, which ensures a fair comparison. This point has been clarified in the final paragraph of Section 2.2:

"*Notably, IGG-SLR-DORIS is used only for comparison rather than combination because it additionally contains DORIS-related LEO information, which may overlap with the LEO contribution already introduced by Tongji-LEO2021 in our combination framework. Therefore, IGG-SLR-HYBRID is adopted for combination to better isolate the SLR contribution and avoid potential redundancy, while IGG-SLR-DORIS serves as a comparison product with comparable GRACE/-FO, SLR, and LEO-related information.*".

C1.10: For greater user convenience, the authors are encouraged to reorganize the combined spherical harmonic solutions into the standard monthly .gfc format adopted by the ICGEM website. Moreover, the current 1° gridded data do not include the corresponding uncertainty information or the associated latitude and longitude coordinates. It is therefore recommended that these gridded products be provided in NetCDF format. It would also be highly beneficial if the authors could provide 0.5° gridded data together with their associated uncertainty estimates, thereby offering users a wider range of choices.

R1.10: Thank you for your constructive comments. Following your suggestions, we have reorganized the combined spherical harmonic solutions, together with their uncertainties, into the standard monthly .gfc format adopted by ICGEM. We have also converted the 1° and 0.5° gridded products derived from the combined spherical harmonic solutions into NetCDF format, including latitude/longitude coordinates. These updated data products are available at https://doi.org/10.5281/zenodo.18589507.

Notably, the 0.5° grid is generated by sampling the spherical harmonic fields at a finer grid interval, rather than representing the intrinsic spatial resolution of degree/order 60 solutions. This distinction has been explained in the data description on the repository page.

Reply to Comments of Reviewer #2

C2: The authors generate a long-term TWSA dataset by combining different gravity satellite missions. This work seems unique in comparison to current widely available long-term TWSA dataset, which is generally generated using climatic inputs. It should be valuable for understanding climate change impacts on global terrestrial water storage change from independent satellite observations. I have some minor comments before it can be accepted.

R2: Thank you for your constructive comments. We have carefully addressed all comments, as detailed below. During revision, we also renamed the original Least-Squares Collocation (LSC) as the Constrained Collocation Model (CCM), as the method is based on the criterion constrained by Tikhonov regularization rather than conventional least squares.

C2.1: It is interesting to see how it differs from that based on climatic input results, for example, Mandal et al., 2025, and/or Li et al., 2021.

References:

(1) Li et al., 2021, GRL, Long-Term (1979-Present) Total Water Storage Anomalies Over the Global Land Derived by Reconstructing GRACE Data.

(2) Mandal et al., 2025, ESSD, Machine-learning-based reconstruction of long-term global terrestrial water storage anomalies from observed, satellite and land-surface model data.

R2.1: Thank you for your valuable comments. Comparisons between CCM-CS and the climate-input-forced reconstruction product developed by Mandal et al. (2025) (i.e., BNML) have already been included in the manuscript at both global and basin scales, as shown in Figs. 7-10.

As shown in Fig. 1 below, the basin-averaged RMSEs and CCs of $\mathrm{TWSC_{CCM}}$, $\mathrm{TWSC_{BNML}}$ and $\mathrm{TWSC_{Li2021}}$ relative to $\mathrm{TWSC_{WB}}$ are computed across 52 basins for the pre-GRACE (January 1994 to March 2002) and GRACE/-FO periods (April 2002 to December 2019), and the entire common period (January 1994 to December 2019). The results indicate that $\mathrm{TWSC_{CCM}}$ generally achieves lower RMSEs and higher CCs than $\mathrm{TWSC_{BNML}}$ and $\mathrm{TWSC_{Li2021}}$, especially during the GRACE/-FO period. Over the entire common period, $\mathrm{TWSC_{CCM}}$ reduces the RMSEs by 1.4% and 1.7% relative to $\mathrm{TWSC_{BNML}}$ and $\mathrm{TWSC_{Li2021}}$, respectively, and increases the corresponding CCs by 2.9% and 6.1%. In addition, $\mathrm{TWSC_{BNML}}$ shows higher consistency with $\mathrm{TWSC_{WB}}$ than $\mathrm{TWSC_{Li2021}}$ and also provides a longer temporal record. Therefore, we only select BNML as a representative climate-input-forced product in our manuscript, consistent with our comparison strategy of selecting one representative product for each category.

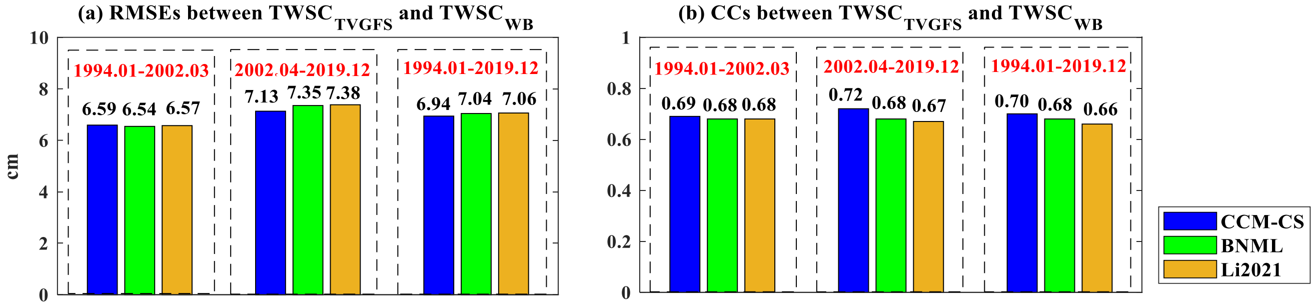

Figure 1: Average (a) RMSEs and (b) CCs of $TWSC_{CCM}$, $TWSC_{BNML}$ and $TWSC_{Li2021}$ relative to $TWSC_{WB}$ across 52 basins during pre-GRACE (January 1994 to March 2002), GRACE/-FO periods (April 2002 to December 2019), and the entire common period (January 1994 to December 2019).

C2.2: Line 24: why the reductions by LSC-CS are not significant, and is this small improvement helpful.

R2.2: Thank you for your constructive comments! The relatively small reduction in water-balance misclosure mainly results from averaging over 52 global basins with diverse hydrological conditions, data quality, and human impacts, which can smooth regionally larger improvements. As shown in Fig. 10(a-f), $TWSC_{CCM}$ clearly exhibits smaller water-balance misclosures and higher CCs than the comparison products in most basins, particularly in northern Eurasia, western North America, northern and western South America, northern Africa, and parts of South Asia.

To assess the basin-scale significance of these improvements, we added a one-sided paired Wilcoxon signed-rank test at the 95% confidence level (Wilcoxon, 1945; Conover, 1999). The results show that CCM-CS achieves significantly lower RMSEs than IGG-SLR-DORIS, RESDCAE, and BNML in 98.1%, 82.7%, and 63.5% of the 52 basins, respectively, and significantly higher CCs in 88.5%, 92.3%, and 71.2% of the basins. The related descriptions have also been added in the second paragraph of Section 4.3.2.

Accordingly, the sentence in Line 24 is revised from "Across 52 major basins, LSC-CS has the smallest water balance errors, with reductions of 4.6%, 2.6%, and 1.5%, respectively." to "*Across 52 major basins, CCM-CS achieves lower water balance errors in 98.1%, 82.7%, and 63.5% of the basins, respectively.*".

C2.3: The word evaporation is better to be changed to evapotranspiration.

R2.3: Thank you! We have changed the word "evaporation" into "evapotranspiration" in the first paragraph of Section 2.3 and the first paragraph of Section 4.3.2.

C2.4: Since the simulated runoff may differ from the truth. The in situ gauged streamflow data (for example, from GRDC) for some selected river basins can be used for water balance closure, and hopefully it may improve the validation.

R2.4: Thank you for your constructive comments! In the revised manuscript, we incorporated GRDC in-situ gauged runoff observations into the gridded simulated runoff fields using an ensemble Kalman filter (Evensen, 2003; Clark, 2008). This fusion takes advantage of the higher local accuracy of in situ observations and the spatial continuity of gridded simulations; the methodological details are provided in Section C of the Supplementary File:

C. Fusion of in-situ observations with gridded simulated runoff

For month t*, the observation equation is written as,*

$$\boldsymbol{y}_t = \boldsymbol{H}_t \boldsymbol{x}_t + \boldsymbol{\varepsilon}_t \tag{S1}$$

where $\boldsymbol{y}_t$ *is the observation vector formed from monthly GRDC runoff observations in terms of runoff depth,* $\boldsymbol{\varepsilon}_t$ *is the observation error,* $\boldsymbol{x}_t$ *is the simulated runoff state vector,* $\boldsymbol{H}_t$ *is the design matrix that maps gridded runoff to each in-situ station by aggregating grid-cell runoff over the corresponding station-controlled drainage area* [29]*.*

The fused runoff state is obtained by minimizing the cost function,

$$J(\boldsymbol{x}_t) = (\boldsymbol{x}_t - \boldsymbol{x}_t^b)^T (\boldsymbol{Q}_t^b)^{-1} (\boldsymbol{x}_t - \boldsymbol{x}_t^b) + (\boldsymbol{y}_t - \boldsymbol{H}_t \boldsymbol{x}_t)^T \boldsymbol{\Sigma}_t^{-1} (\boldsymbol{y}_t - \boldsymbol{H}_t \boldsymbol{x}_t) \tag{S2}$$

where the background state $\boldsymbol{x}_t^b = \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{x}_t^{b,(k)}$ *with* $\boldsymbol{x}_t^{b,(k)}$ *denoting the* k*th simulated runoff product among* N *ensemble members; the background error covariance* $\boldsymbol{Q}_t^b = \frac{1}{N-1}\sum_{k=1}^{N}\left(\boldsymbol{x}_t^{b,(k)} - \boldsymbol{x}_t^b\right)\left(\boldsymbol{x}_t^{b,(k)} - \boldsymbol{x}_t^b\right)^T$*; and the observation-error covariance* $\boldsymbol{\Sigma}_t = diag\left(\sigma_{1,t}^2, \ldots, \sigma_{m_t,t}^2\right)$*, with* $\sigma_{s,t}$ *representing the uncertainty of the runoff observation at* s*th station, and* m_t *is the number of valid GRDC stations at month* t*.*

The corresponding ensemble solution $\boldsymbol{x}_t^a$ *is,*

$$\boldsymbol{x}_t^a = \boldsymbol{x}_t^b + \boldsymbol{K}_t (\boldsymbol{y}_t - \boldsymbol{H}_t \boldsymbol{x}_t^b) \tag{S3}$$

where $\boldsymbol{K}_t = \boldsymbol{Q}_t^b \boldsymbol{H}_t^T (\boldsymbol{H}_t \boldsymbol{Q}_t^b \boldsymbol{H}_t^T + \boldsymbol{\Sigma}_t)^{-1}$ *is the gain matrix.*

The water-balance closures for the 52 selected basins have been recalculated using the fused runoff product, and the corresponding results have been updated in Fig. 10.

To further quantify the effect of incorporating in situ runoff, Table 1 below compares the basin-averaged RMSEs of $\mathrm{TWSC_{TVGFS}}$ relative to $\mathrm{TWSC_{WB}}$ across the 52 basins before and after the fusion. The results show that the fused runoff product improves water-balance closure for all solutions across the different evaluation periods.

Tab. 1 Basin-averaged RMSEs of $\mathrm{TWSC_{TVGFS}}$ relative to $\mathrm{TWSC_{WB}}$ across 52 basins before and after incorporating in-situ runoff

Period	Runoff	CCM-CS	IGG-SLR-DORIS	RESDCAE	BNML
Pre-GRACE	Simulate	6.60	6.90	6.58	6.56
	In-situ& Simulate	**6.59**	**6.89**	**6.57**	**6.54**
GRACE/-FO	Simulate	7.39	7.70	7.63	7.62
	In-situ& Simulate	**7.13**	**7.45**	**7.36**	**7.35**
Entire period	Simulate	7.13	7.47	7.32	7.24
	In-situ& Simulate	**6.94**	**7.29**	**7.12**	**7.04**

C2.5: A spatial distribution map showing the long-term TWSA trend can be provided for a better understanding of the result, and maybe its differences with other results.

R2.5: Thank you for your valuable comments! According to your comments, we have added the linear trends of spatial TWSA from CCM-CS and the three compared solutions in S2 of the supplementary file. To quantify the differences, we also present the absolute linear trend differences among various solutions in Fig. S3. The related analysis is added in the second and third paragraphs of Section 4.2.1:

"The spatial linear trends from CCM-CS and the reference solutions are shown in Fig. S2. The results show that the TWSA trends from CCM-CS are generally consistent with those from IGG-SLR-DORIS and RESDCAE, particularly over northern India, western North America, southern South America, the Antarctic Peninsula, and the West Antarctic margin, where pronounced decreasing trends are observed. Consistent increasing trends are also found in several regions, including parts of northern Eurasia, eastern North America, central South America, and coastal East Antarctica. In contrast, the hydrometeorological-forced BNML product shows substantially weaker trends over most regions.

To further quantify the differences among various solutions, the NSE values and linear trend differences between CCM-CS and the three reference solutions are presented in Fig. 7 and Fig. S3, where the first three rows display metrics relative to RESDCAE, IGG-SLR-DORIS, and BNML, respectively. The results demonstrate that CCM-CS exhibits superior consistency with all three reference solutions compared to their mutual cross-consistencies, as evidenced by higher NSE values and lower linear trend differences. Relative to RESDCAE, the latitude-weighted global NSE and linear trend difference, excluding Greenland and Antarctica, derived from CCM-CS are 0.53 and 0.16 cm/yr, respectively, outperforming IGG-SLR-DORIS, 0.47 and 0.27 cm/yr, and BNML, 0.51 and 0.27 cm/yr. Relative to IGG-SLR-DORIS, CCM-CS attains an NSE of 0.75 and a trend difference of 0.08 cm/yr, outperforming RESDCAE, 0.43 and 0.27 cm/yr, and BNML, 0.45 and 0.17 cm/yr. Relative to BNML, CCM-CS achieves an NSE of 0.46 and a trend difference of 0.10 cm/yr, whereas IGG-SLR-DORIS, 0.42 and 0.17 cm/yr, and RESDCAE, 0.24 and 0.27 cm/yr, exhibit lower consistency.".

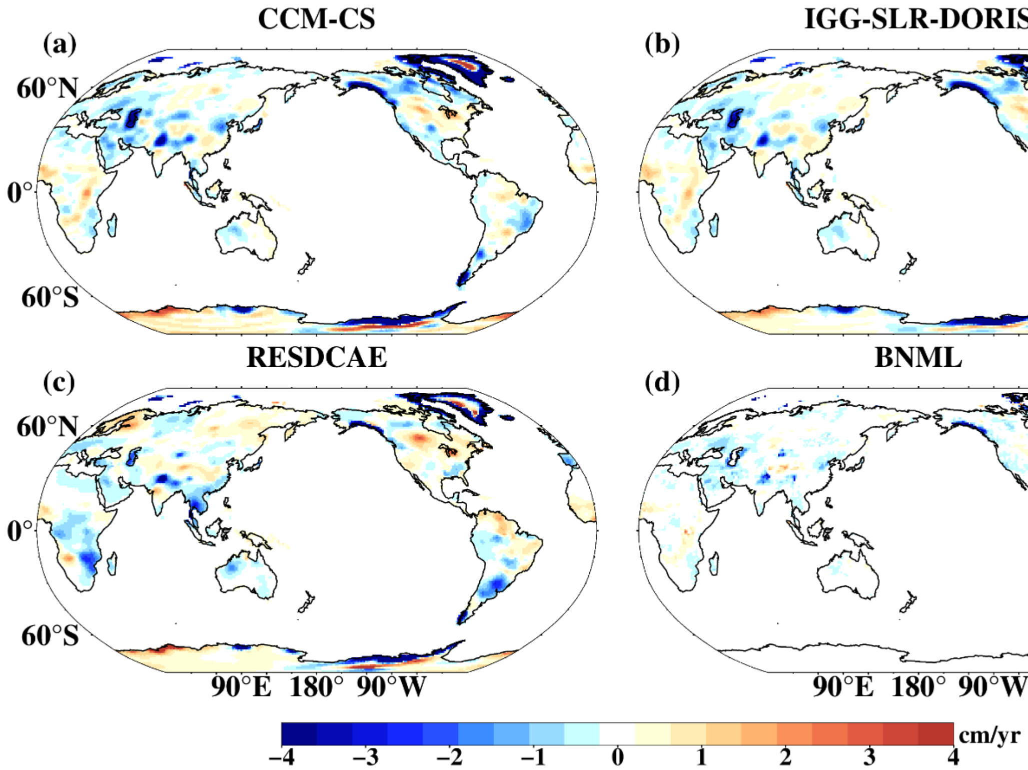

Figure S2: Linear trends of global TWSA during the common period from January 1994 to December 2020.

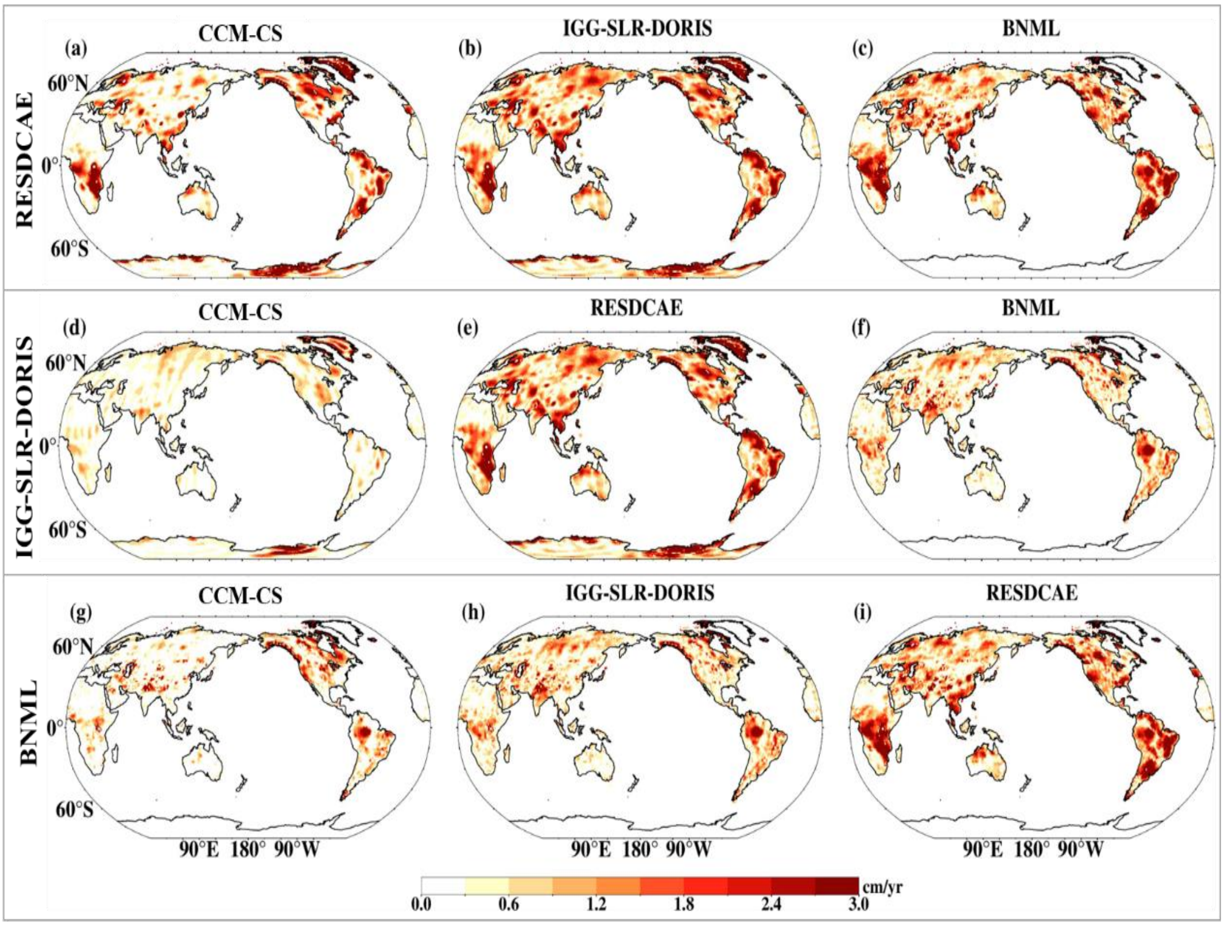

Figure S3: Absolute differences in linear trend of global TWSA from various solutions relative to RESDCAE, IGG-SLR-DORIS, and BNML corresponding to the first to third rows (January 1994 to December 2020).

References

[1] Clark, M., Rupp, D., Woods, R., and et al.: Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model, Advances in Water Resources, 31(10), 1309-1324, https://doi.org/10.1016/j.advwatres.2008.06.005, 2008.

[2] Conover, W. J.: Practical Nonparametric Statistics, 3rd ed., John Wiley & Sons, New York, 1999.

[3] Evensen, G.: The ensemble Kalman filter: Theoretical formulation and practical implementation, Ocean Dynamics, 53, 343-367, https://doi.org/10.1007/s10236-003-0036-9, 2003.

[4] Wilcoxon, F.: Individual comparisons by ranking methods, Biometrics Bulletin, 1(6), 80-83, https://doi.org/10.2307/3001968, 1945.