Supplementary Material for the article "Impact of the NGGM and MAGIC future satellite gravity missions for scientific applications and operational services"

Julia Pfeffer[1], Benoît Meyssignac[2], Rory Bingham[3], Alejandro Blazquez[2], Marie Bouih[1], Carla Braitenberg[4], Luca Brocca[5], Henryk Dobslaw[6], Ramiro Ferrari[1], Ehsan Forootan[7], Helena Gerdener[8], Muhammad Tahir Javed[4], Laura Jensen[6], Volker Klemann[6], Anna Kremer[9,8], Jürgen Kusche[8], Gilles Larnicol[1], Muhammad Usman Liaqat[5], Francesco Leopardi[5], Elisavet-Maria Mamagiannou[10], Gerardo Maurizio[4], Roland Pail[11], Isabelle Panet[9], Thomas Papanikolaou[10], Peyman Saemian[12], Ingo Sasgen[13,14], Maike Schumacher[7], Marius Schlaak[11], Linus Shihora[6], Alireza Sobouti[12], Nico Sneeuw[12], Mohamad J. Tourian[12], Dimitrios Tsoulis[10], Georgios Vergos[10], Bert Wouters[15], Fan Yang[7], Ilias Daras[16]

[1] Magellium, Ramonville - Saint - Agne, 31750, France
[2] CNES, Toulouse, 31400, France
[3] University of Bristol, Bristol, BS8 1QU, United Kingdom
[4] University of Trieste, Trieste, 34127, Italy
[5] Research Institute for Geo-Hydrological Protection, National Research Council, Perugia, 06128, Italy
[6] GFZ Helmholtz Centre for Geosciences, Potsdam, 14473, Germany
[7] University of Aalborg, Aalborg, 9220, Denmark
[8] University of Bonn, Bonn, 53115, Germany
[9] Institut de Physique du Globe de Paris, Paris, 75005, France
[10] Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece
[11] Technical University of Munich, Munich, 80333, Germany
[12] Institute of Geodesy, University of Stuttgart, Stuttgart, 70174, Germany
[13] Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Bremerhaven, 27570, Germany
[14] Institute of Geography, University of Augsburg, Augsburg, 86159, Germany.
[15] Delft University of Technology, Deft, 2628 CN, The Netherlands
[16] European Space Research and Technology Center, European Space Agency, Noordwijk, 2201 AZ, The Netherlands

Correspondence to: Julia Pfeffer (julia.pfeffer@magellium.fr)

S1. Drought monitoring and prediction

The analysis of the Rio Grande river basin revealed that the exact drought severity classes of the NGGM and MAGIC assimilation agree better with the simulated truth as compared to the GRACE-C-like assimilation. We further complement the agreement analysis by also providing how much of all areas and all months of the basin experience a disagreement in drought severity classes. For this, we further distinguish if the assimilations and the simulated truth differ by only one class or equal/more than 2 classes and if the disagreement is positive or negative. The agreement and disagreement is provided in Table S1.1. In addition to the improved agreement of NGGM and MAGIC compared to GRACE-C-like, the drought disagreement between the assimilations and the simulated truth shows that the NGGM and MAGIC assimilations show fewer disagreement in drought severity from the truth than GRACE-C-like. For example, the NGGM assimilation drought indicates 8.16% of cells and months to deviate positively severely (2 or more severity classes) from the truth, while this increases to 12.96% for the GRACE-C-like assimilation.

Table S1.1: Comparison of drought severity classes between the mission scenario assimilations and the simulated truth for the Rio Grande river basin. The statistics are computed as percentage agreement/disagreement of all grids and all months and all severity classes for that basin.

Drought severity class comparison towards truth [%]	**GRACE-C-like**	**NGGM**	**MAGIC**
Agreement	57.02	63.20	63.86
False positive (1 class higher than truth)	8.62	10.14	9.84
False negative (1 class lower than truth)	10.07	10.02	10.59
Severely false positive (2 or more classes higher than truth)	11.32	8.46	7.90
Severely false negative (2 or more classes higher than truth)	12.96	8.16	7.80

Figure S1.1 (left) shows the agreement between simulated truth and the GRACE-C-like assimilation on a global scale per basin, which for most basins shows an agreement of about 50 to 70 percent. Basins with a high agreement between truth and GRACE-C-like assimilation compared to other basins are, for example, the Mississippi, Murray-Darling and Yangtze river basins. The improvement of NGGM and MAGIC assimilation w.r.t. the GRACE-C-like assimilation in severity class agreement is shown in the middle and right panels. For example, for the Mississippi basin, Canadian basins such as the Nelson basin or the Zimbabwe basin we find a stronger improvement as compared to – among others – the Niger basin. Interestingly, the improvements are fewer along the equator and increase towards the poles.

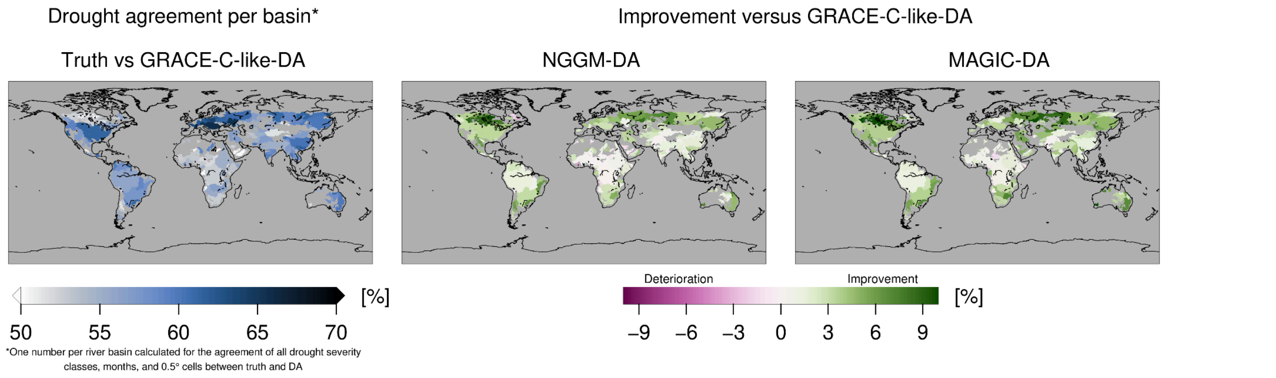


Figure S1.1: (Left) Drought agreement of GRACE-C-like-DA and the simulated truth given in percentage per basin. The agreement is computed from all DSI severity classes, months and 0.5° grids. (Middle and right) Percentage improvement of drought agreement per of NGGM-DA and MAGIC-DA compared to GRACE-C-like-DA per basin.

S2. Total Drainable Water Storage

TDWS serves as a proxy for freshwater availability by representing the drainable fraction of terrestrial water storage that can sustain runoff, ecosystems, and human water use. Figure S2.1 shows the global distribution of estimated TDWS derived from ESM2.0, expressed as equivalent water height. The spatial pattern reveals regional contrasts, including relatively shallow TDWS estimates across boreal basins. In Table S2.1, we compare the total volume of estimated drainable water across different mission scenarios and ESM2.0. MAGIC and NGGM differ from the ESM2.0 reference by about 4%, whereas GRACE-C shows a larger difference of about 11%. Overall at global scale, NGGM and MAGIC reproduce the reference TDWS more accurately than GRACE-C.

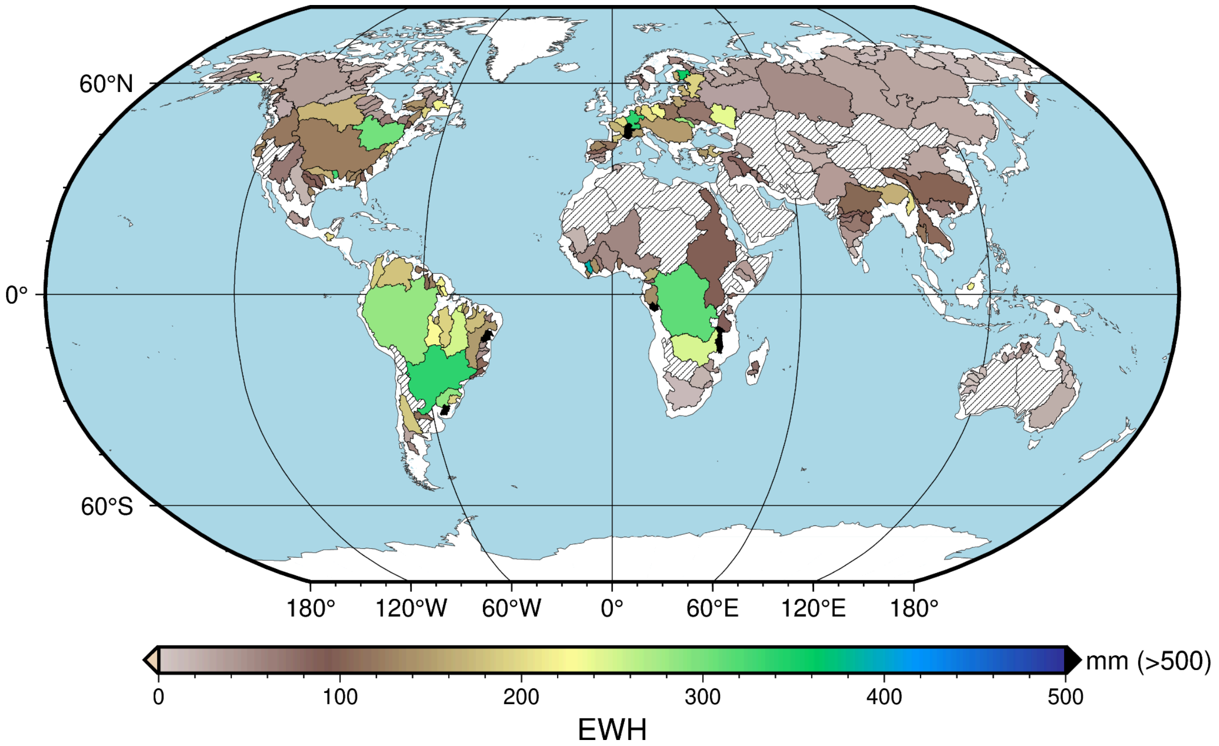


Figure S2.1: Global distribution of TDWS derived from ESM2.0-based analysis. Hatched areas represent discharge-less basins.

Model / Mission	Volume (km³)	Δ Volume %
ESM2.0	10,292,020	
MAGIC	9,863,710	4 %
NGGM	9,924,950	4 %
GRACE-C	9,128,260	11 %

Table S2.1: Comparison of total volume of estimated drainable water across mission scenarios.

S3. Ocean Science

SLB extended results at local scale

To assess the relative capacity of each mission configuration to close the SLB, simulated error metrics are normalized at local scale against the SLB residuals derived from current observational systems. Figure S3.1 presents the normalized annual cycle amplitude maps, expressing the simulated error as a percentage of the observed SLB residuals variability. Regional improvements in resolving seasonal ocean mass variations are particularly evident for next-generation constellations (NGGM and MAGIC) across the south of Pacific, Indian and Atlantic oceans. NGGM markedly improves SLB closure along the Antarctic Circumpolar Current (ACC), across the North Atlantic, and within coastal upwelling zones. MAGIC maintains these regional gains while eliminating the polar gap error amplification present in NGGM, offering optimal closure performance across the Arctic and Southern High Latitudes.

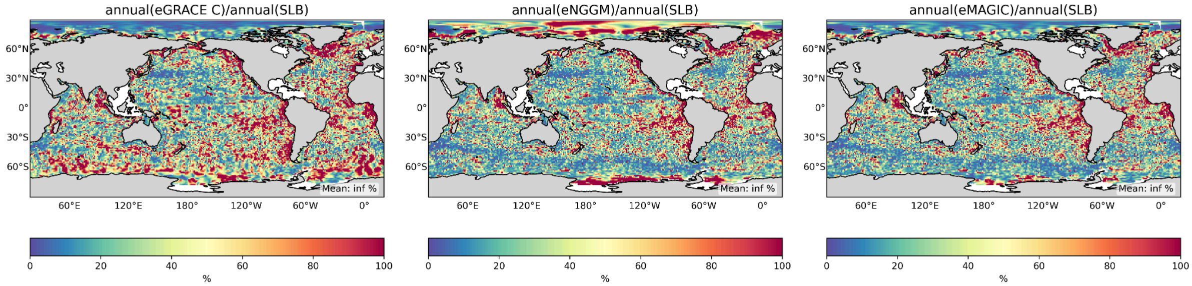


Figure S3.1: Annual cycle amplitude maps of filtered simulations errors normalized by sea level budget (SLB) residuals, and their comparisons. Top panels: Annual cycle amplitude maps of the errors from GRACE-C (left), NGGM (center), and MAGIC (right) simulations, expressed as a percentage of the sea level budget residual variability.

Similarly, Figure S3.2 displays the normalized linear trend error maps over the 12-year evaluation period. At the decadal scale, filtered solutions demonstrate relatively consistent spatial performance across constellations, with NGGM and MAGIC showing small reductions in residual noise compared to GRACE-C, except in high-latitude polar regions where NGGM is impacted by its orbit inclination gap.

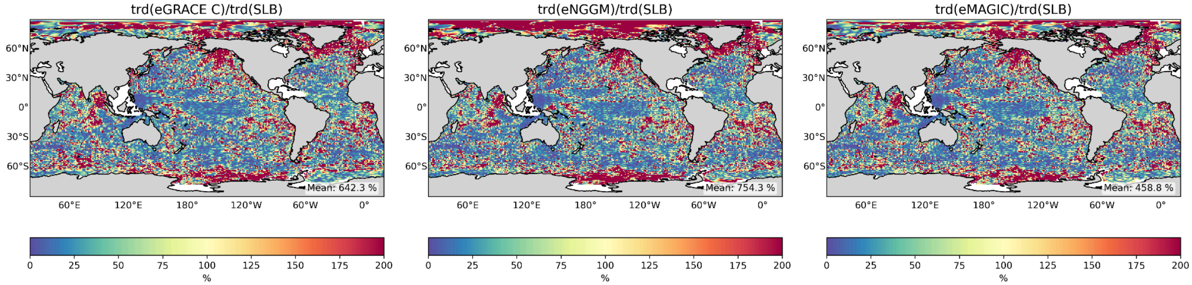


Figure S3.2: Trend maps of filtered simulations errors normalized by sea level budget (SLB) residuals, and their comparisons. Top panels: Trend maps of the errors from GRACE-C (left), NGGM (center), and MAGIC (right) simulations, expressed as a percentage of the sea level budget residual variability.

Overall, the spatial error distributions observed across decadal trends and annual cycle amplitudes align well with the patterns identified in the standard deviation analysis. However, the contrast between satellite constellations is notably less pronounced for linear trends and annual amplitudes than for short-term and interannual standard deviations.

The time-series standard deviation emerges as the most sensitive diagnostic metric for discriminating noise performance across gravity mission architectures. The reduced sensitivity observed in linear trend metrics is partially an artifact of the underlying simulation setup: the ESA ESM2.0 model does not explicitly simulate dynamic, non-barystatic ocean mass redistribution. Consequently, forward-modeled low-frequency aliasing footprint errors are artificially suppressed, attenuating the apparent performance gap between configurations when assessed on trend closure alone.

SLB results at global mean scale

To complement the regional spatial diagnostics, Table S3.1 summarizes the global mean metrics for filtered and unfiltered simulation errors across all three satellite constellations (GRACE-C, NGGM, and MAGIC), benchmarked directly against the SLB residuals obtained from historical GRACE observations.

For GRACE-C, spatial filtering is essential to suppress high-frequency noise, reducing the time-series standard deviation from 0.40 mm down to 0.24 mm. In contrast, next-generation missions achieve extremely low high-frequency noise floors even without any filtering: the unfiltered standard deviation drops to 0.04 mm for NGGM and reaches 0.01 mm for MAGIC (a 40-fold reduction compared to unfiltered GRACE-C). Interestingly, applying a DDK filter to NGGM and MAGIC slightly increases their global mean residual standard deviation (to 0.13 mm and 0.14 mm, respectively). This demonstrates that for intrinsically quiet spatial fields, spatial filtering introduces slight smoothing and leakage artifacts rather than noise suppression.

Without filtering, GRACE-C suffers from a strong annual error (7.19 mm), which requires spatial filtering to bring it down to 1.82 mm. Conversely, unfiltered solutions for NGGM and MAGIC control seasonal error propagation without filtering. When filtered, all three missions yield consistent, low annual errors, all remaining well below the observational SLB residual variability (2.99 mm).

Linear trend errors remain constrained across all configurations, ranging from -0.17 to -0.11 mm/yr for filtered solutions. Notably, unfiltered MAGIC achieves the closest proximity to a zero-error baseline with a trend error of only -0.02 mm/yr.

In a nutshell, GRACE-C remains heavily reliant on post-processing spatial filters to control noise explosion, particularly in the annual cycle. NGGM and MAGIC demonstrate that improved spatio-temporal sampling almost eliminates the need for spatial filtering. Unfiltered MAGIC solutions achieve an unprecedented precision (STD = 0.01 mm), releasing future oceanographic applications from the spatial leakage and signal attenuation traditionally caused by post-processing filters. Across all metrics, the simulated gravimetry error budgets for both NGGM and MAGIC remain at least an order of magnitude smaller than the current empirical SLB residuals, confirming that gravimetry errors will no longer be the primary limiting factor in global sea level budget closure.

Table S3.1: Values of the global mean trend and annual signal amplitude of the GRACE-C, NGGM and MAGIC errors filtered simulations and the Sea Level Budget residuals from GRACE historic data.

	Filtered GRACE-C errors unfiltered	Filtered NGGM errors unfiltered	Filtered MAGIC errors unfiltered	SLB residuals
Trend (mm/yr)	-0.17	-0.11	-0.15	0.42
	-0.15	0.10	-0.02	
Annual signal (mm)	1.82	1.56	1.55	2.99
	7.19	1.73	1.59	
STD of detrended and deseasonalized	0.24	0.13	0.14	2.20

time series (mm)	0.40	0.04	0.01	

S4. Climate sciences

Relative impact of the errors in the OHC and OHU estimates for different mission configurations.

Differences in OHC changes estimated from manometric sea level variations were assessed by comparing simulation results with those derived from ocean model estimates, after applying a low-pass filter to remove high-frequency variability with periods shorter than two years. The standard deviation of the OHC error ratios, derived from L3b GRACE-C simulations and the MOHeaCAN product, is shown in Figure S4.1. The standard deviation of OHC errors for GRACE-C reaches 20% of the OHC MOHEACAN standard deviation in some regions (e.g., East South Pacific) and exceeds 50% at high latitudes and in the Indian Ocean. The errors for L3b NGGM simulations decrease significantly compared to GRACE-C (Figure S4.1 and Figure S4.2). Beyond 30° latitude, GRACE-C simulation errors in OHC are up to 80 and 30% greater, respectively, than those derived from NGGM simulations. Similar results were found for the standard deviation of OHC errors, which shows a substantial decrease in the L3b MAGIC simulations compared to GRACE-C (not shown).

Globally, errors for GRACE-C are significantly larger than those for NGGM and MAGIC. However, for MAGIC, the standard deviation of OHC errors increases by approximately 30% in the mid-latitudes (30° and 60°) compared to NGGM. The difference between NGGM and MAGIC may stem from the application of the VADER filter, as errors from the L3a versions of both simulations exhibit similar spatial patterns and amplitudes.

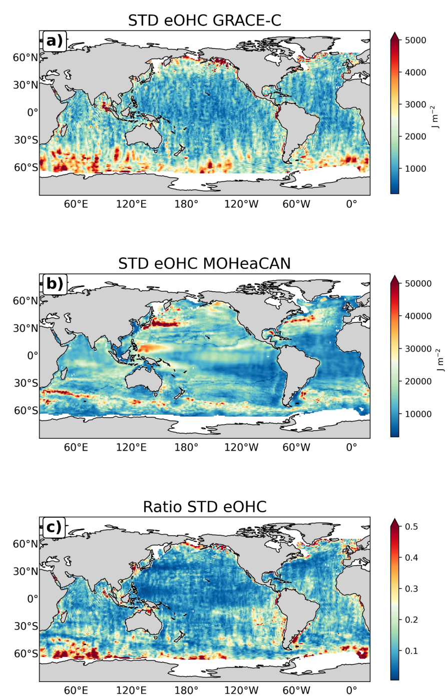


Figure S4.1: Standard deviation of OHC errors derived from filtered L3 GRACE-C simulations (a) and standard deviation of the OHC from MOHeaCAN product (b). Panel (c) shows the ratio between (a) and (b).

The mean and standard deviation of the OHU error ratios based on the MOHeaCAN space geodetic product show similar results and spatial patterns to those observed in the OHC analysis (not shown).

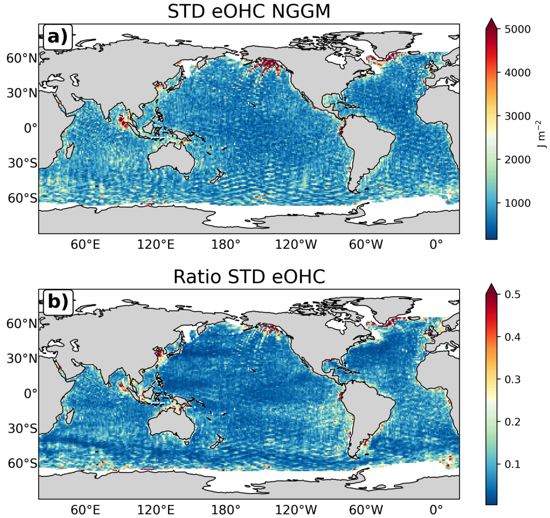


Figure S4.2: Standard deviation of OHC errors derived from filtered L3 NGGGM simulations (a). b) shows the ratio between (a) and standard deviation of the OHC from MOHeaCAN product.

Sensitivity assessment

Errors from L3a GRACE-C simulations exhibit the same spatial patterns, but with amplitudes at least two orders of magnitude greater than those from the L3b simulations. Errors from L3a NGGM simulations exhibit the same spatial patterns, but with amplitudes of one order of magnitude greater than those from the filtered L3b simulations. Furthermore, we have extended the analysis by assessing the sensitivity of OHC and OHU error estimates to temporal filtering, using an additional filter with cutoffs at 1 year. Overall, we observe the same patterns as those identified in the analysis presented in the previous section using a low-pass filter with a two-year cutoff. For GRACE-C, applying a one-year low-pass temporal filter results in an increase of approximately 10% in the standard deviation of OHC errors relative to the two-year filtered case. OHU errors exhibit a more pronounced sensitivity, with increases of up to 30% at high latitudes and in the Indian Ocean (not shown). Consistent with previous findings, while MAGIC exhibited higher standard deviations of OHC errors than NGGM in the mid-latitudes when using a low-pass filter with a two-year cutoff, similar differences are observed between the two simulations when applying a one-year low-pass filter.

Impact of the errors in the EEI estimates.

We evaluated the errors in EEI change estimates derived from manometric simulations by comparing them with estimates from a space geodetic product, following the application of low-pass filters to suppress high-frequency variability with periods shorter than two years.

Considering the 1995–2004 period, excluding the impact of the Sumatra event on the simulation errors, the mean and standard deviation of EEI errors for GRACE-C are 0.19 and 0.12 W/m2, respectively. These correspond to 27% and 30% of the mean and standard deviation of EEI estimates from the MOHeaCAN product (Table S4.1). These errors exceed those of NGGM and MAGIC by more than 30% in mean and 17% in standard deviation. Finally, when applying a one-year low-pass filter, the simulation errors increase by approximately 60% for GRACE-C and by ~10% for both NGGM and MAGIC.

Table S4.1: Mean and standard deviation of errors in EEI change estimates derived from manometric simulations in W/m2. Error ratios in EEI change estimates derived from manometric simulations relative to the MOHeaCAN product are shown in parentheses.

	GRACE-C	NGGM	MAGIC
Mean	0.19 (0.27)	0.12 (0.18)	0.13 (0.19)
Std	0.12 (0.3)	0.10 (0.25)	0.10 (0.25)

Comparison of simulated GRACE-C OHC and OHU errors with real manometric uncertainties.

Errors in OHC and OHU derived from GRACE-C simulations were compared with uncertainty estimates derived from historical GRACE solutions. To ensure methodological consistency historical GRACE data and GRACE-C simulations were filtered using the same approach. The historical GRACE solutions and unfiltered GRACE-C simulations have been reprocessed with the DDK3 filter. GRACE uncertainty estimates represent the variability across the five processing centres included in the ensemble. They capture only the variability arising from centre-specific processing strategies. This metric therefore provides an indication of the realism of the GRACE-C simulated error fields. OHC and OHU errors were computed following the methodology described in S1.5.

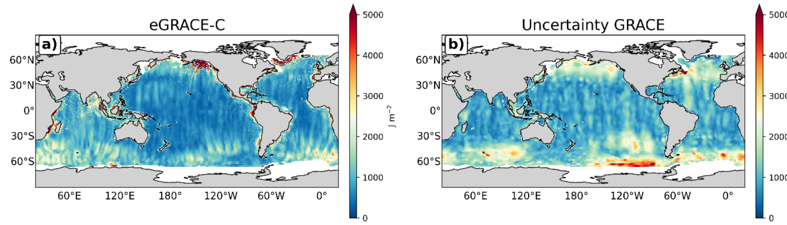


Figure S4.3: Comparison between OHC GRACE-C's simulated errors and uncertainties from GRACE historical data. Standard deviation of GRACE-C simulation errors (a), of the GRACE's uncertainties (b), and Standard deviation of GRACE-C simulation errors normalized by GRACE'S uncertainties (c).

Figure S4.3 shows that the standard deviation of the OHC derived from GRACE uncertainty and the simulated GRACE-C error are relatively close to each other. GRACE-C error explains nearly 70 % of the estimated uncertainty over large portions of the global ocean. The representation of the GRACE-C simulated errors as a percentage of the GRACE uncertainty suggests a slight underestimation of the simulated errors relative to the actual uncertainty. Analyses conducted using the standard deviation of OHU yielded similar results. The comparisons between the mean OHU derived from GRACE uncertainties and the simulated GRACE-C errors reveal a general agreement, except for the stripe pattern. Over a large fraction of the global ocean, the representation of the GRACE-C simulated errors as a percentage of the GRACE uncertainty suggests that the simulated errors are underestimated compared to the true uncertainties.

We also evaluated the errors in EEI change estimates derived from GRACE-C simulations by comparing them with estimates from GRACE'S uncertainties, following the application of low-pass filters to suppress high-frequency variability with periods shorter than two years. For the 1995–2004 period, the mean and standard deviation of EEI errors derived from GRACE-C are 0.06 and 0.05 W/m², respectively. These values underestimate the corresponding EEI derived from GRACE uncertainties by ~60%, which were evaluated over a different period (2003–2015) (Table S4.2).

Table S4.2: Mean and standard deviation of errors in EEI change estimates derived from GRACE-C simulated errors and and from historical GRACE uncertainties in W/m2.

	GRACE-C	GRACE
Mean	0.06	0.08
Std	0.05	0.07

S5. Solid Earth Sciences

We show the performance of the inversion of the earthquake slip in terms of rake and amplitude factor with respect to a synthetic purely strike-slip and thrust field. The result for the unfiltered and VADER filtered gravity solution is analyzed for the three missions for moment magnitude Mw 8.0 and 8.5 (Table S5.1, Table S5.2). The inverted gravity observation tests are equal to the field generated by an intermediate-deep earthquake (80 km depth of hypocenter), with a dip of 30°, a rake of 30°, and moment magnitude (Mw) of 8.5, downscaled after the first test to a moment magnitude (Mw) of 8.0. To this signal, the retrieval error of the three missions has been added, using a solution for a 5-day sampling interval. For the VADER filtered simulation, the same filter was also applied to the earthquake signal. In both cases, signal and noise are synthesized up to d/o 70. We invert 36 realizations of the observation with noise and calculate the average and the standard deviation of the slip inversion.

Table S5.1 Inversion Results, case M8.5. Average results and standard deviations (in parentheses) for 36 solutions after the time of the simulated earthquake are shown. S1, S2: inverted factors for strike slip and thrust (adimensional). Rake in degrees, amp is the adimensional seismic moment amplitude factor relative to the input earthquake field.

		MAGIC				NGGM				GRACE-C			
M	Theoretical Moment scaling factor respect to M8.5	S1	S2	rake	amp	S1	S2	rake	amp	S1	S2	rake	amp
VADER Filtered solution. maximum d/o N=70. 5-daily													
8.5	1	00.84 (00.09)	0.48 (0.10)	29.90 (5.58)	0.97 (0.09)	0.85 (0.09)	0.49 (0.10)	29.79 (5.51)	0.98 (0.09)	00.89 (00.14)	0.49 (0.16)	28.74 (8.77)	1.01 (0.14)
Unfiltered solution. maximum d/o N=70. Gentle cut signal 70-90.5-daily													
8.5	1	0.86 (0.18)	0.51 (0.18)	30.50 (10.09)	1.00 (0.18)	0.87 (0.17)	0.50 (0.17)	29.93 (9.71)	1.00 (0.17)	0.77 (4.17)	1.30 (5.45)	59.29 (171.79)	1.51 (5.15)

Below, the same inversion results for a downscaled moment magnitude Mw 8.0 earthquake.

Table S5.2 Inversion Results, case M8.0. Average results and standard deviations (in parentheses) for 36 solutions after the time of the simulated earthquake are shown. S1, S2: inverted factors for strike slip and thrust (adimensional). rake in degrees, amp is the adimensional seismic moment amplitude factor relative to the input earthquake field.

		MAGIC				NGGM				GRACE-C			
M	Theoretical Moment scaling factor respect to M8.5	S1	S2	rake	amp	S1	S2	rake	amp	S1	S2	rake	amp
VADER Filtered solution. maximum d/o N=70. 5-daily.													

		MAGIC				NGGM				GRACE-C			
8.0	0.178	0.72 (0.49)	0.41 (0.55)	29.69 (36.59)	0.83 (0.50)	0.75 (0.51)	0.42 (0.54)	29.05 (35.16)	0.86 (0.51)	0.96 (0.78)	0.43 (0.89)	24.06 (47.83)	1.05 (0.80)
Unfiltered solution. maximum d/o N=70. Gentle cut signal 70-90. 5-daily													
8.0	0.178	0.84 (0.99)	0.54 (0.99)	32.78 (56.86)	1.00 (0.99)	0.88 (0.98)	0.50 (0.94)	29.57 (54.17)	1.01 (0.97)	0.35 (23.5)	5.01 (30.6)	86.05 (268.3)	5.02 (30.6)

S6. Geodesy

Evaluation of Mission Contributions to IHRF Realization

To quantitatively illustrate the geodetic performance of the simulated mission scenarios (GRACE-C, NGGM, and MAGIC) in realizing the IHRF, we present a detailed time-series analysis for a representative core station: NRIL (Norilsk, RU; Station ID: 12364M001). Located at geodetic coordinates φ=69.3618° N and λ=88.3598° E, the NRIL station is situated in the Siberian region of Arctic Russia, offering an excellent testbed for evaluating gravity field recovery at higher latitudes, given the NGGM foreseen contribution up to 70oN. The evaluation specifically focuses on the derived point-specific gravity potential anomalies over the 12-year simulation period. To quantify the retrieval accuracy and the added value of the forthcoming gravity constellations, the Level-4 gravity potential anomalies derived from the GRACE-C, NGGM, and MAGIC missions are subjected to a detailed, three-step examination: first, we present the temporal evolution of the time series for each mission, then, we provide the pairwise mission differences to assess their inter-mission consistency, and finally, the retrieved mission signals are directly evaluated against the Earth System Model (ESM2.0), which serves as the absolute ground truth. Focusing on the detailed analysis of the NRIL reference station,Figure S6.1 displays the temporal evolution of the gravity potential anomalies for the three missions along with their pairwise differences, fulfilling the first two steps of our analysis, while Table S6.1 presents their corresponding statistical values.

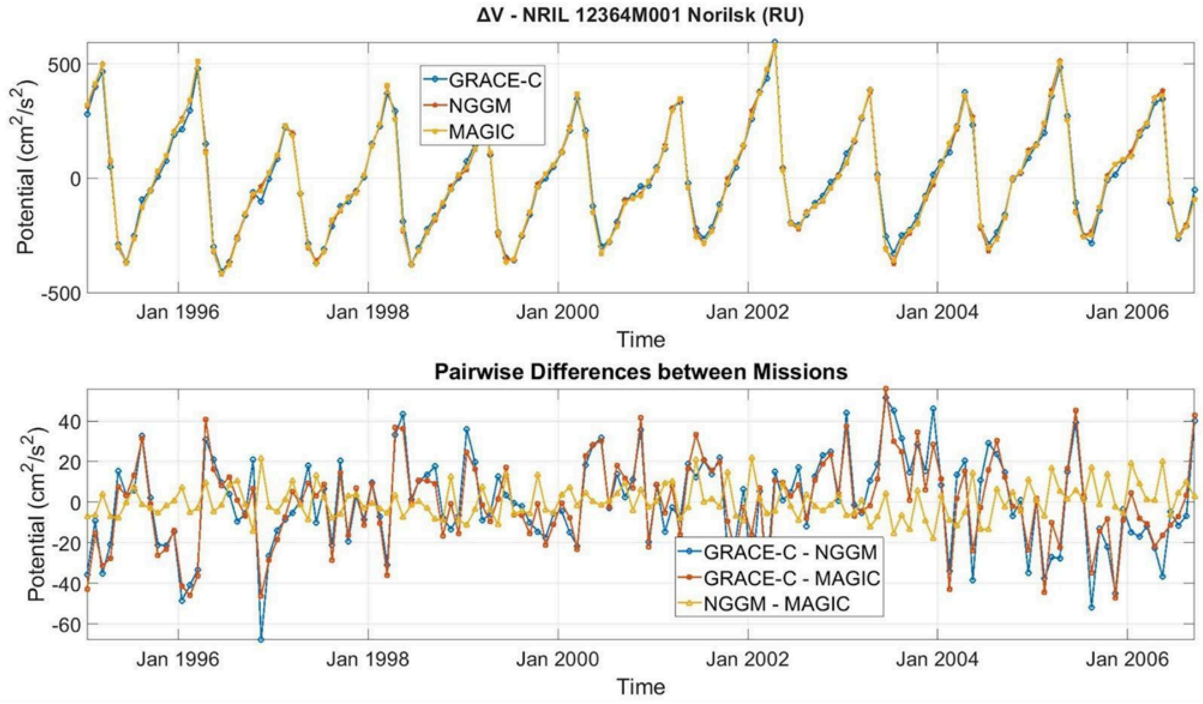


Figure S6.1: Time series of the Level-4 gravity potential anomalies (ΔV) for the GRACE-C, NGGM, and MAGIC simulated missions (top), along with their pairwise differences (bottom), at the NRIL IHRF reference station.

Table S6.1: Statistics (mean and standard deviation) of the gravity potential anomalies (ΔV) and their pairwise inter-mission differences at the NRIL IHRF reference station (Unit: cm²/s²).

Dataset	Mean	Std
GRACE-C	-1.133	228.771
NGGM	-0.935	238.052
MAGIC	-1.129	232.191
[GRACE-C] - [NGGM]	-0.197	22.756

[GRACE-C] - [MAGIC]	-0.004	21.267
[NGGM] - [MAGIC]	0.193	8.2083

As observed from Figure S6.1 and Table S6.1, the overall temporal evolution of ΔV exhibits comparable magnitude across the three scenarios. However, the pairwise mission differences highlight a strong internal consistency between the NGGM and MAGIC (with a standard deviation of just 8.208 cm^2/s^2), whereas GRACE-C shows notably larger discrepancies when compared to either NGGM or MAGIC (standard deviations of ~21–22 cm^2/s^2). For the final evaluation step, Figure S6.2 depicts the residuals between each mission dataset and the ESM2.0 ground truth, with Table S6.2 presenting the corresponding statistics. This residual analysis clearly demonstrates that both NGGM and MAGIC significantly outperform the GRACE-C baseline. As observed from Figure S6.2 and Table S6.2, the standard deviation of the differences to ESM2.0 drops from 22.266 cm^2/s^2 for GRACE-C to just 9.675 cm^2/s^2 for NGGM and 9.582 cm^2/s^2 for MAGIC. Consequently, the MAGIC constellation achieves the highest accuracy, delivering a relative performance improvement of approximately 56.9% compared to the GRACE-C baseline, with NGGM providing a comparable improvement to 56.6%. In both cases, the mean residual values remain practically zero across all missions, confirming the absence of systematic biases.

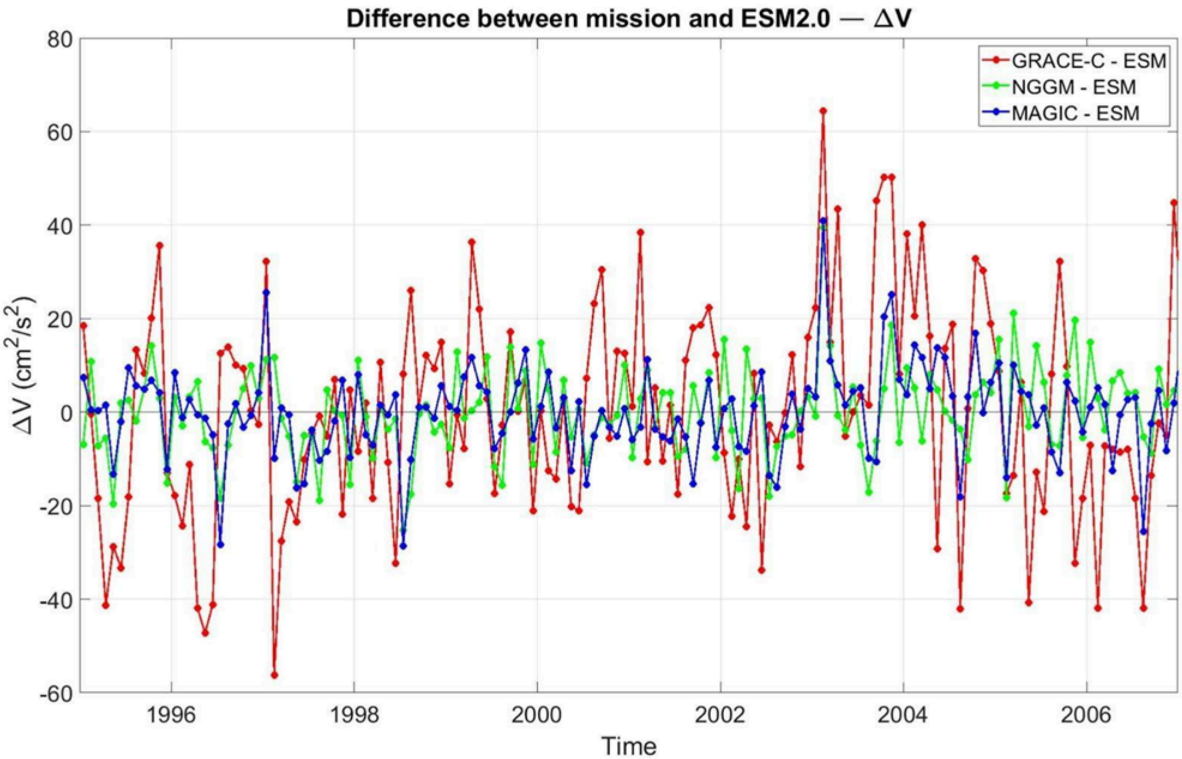


Figure S6.2: Residual differences in the Level-4 gravity potential anomalies (ΔV) between the mission datasets (GRACE-C, NGGM, MAGIC) and the ESM2.0 ground truth model at the NRIL IHRF reference station.

Table S6.2: Statistics (mean and standard deviation) of the residual differences in gravity potential anomalies (ΔV) between the simulated mission datasets and the ESM2.0 ground truth model at the NRIL IHRF reference station (Unit: cm^2/s^2).

Dataset	Mean	Std
[GRACE-C] – [ESM2.0]	-1.495×10^{-9}	22.266
[NGGM]- [ESM2.0]	9.468×10^{-9}	9.675
[MAGIC]- [ESM2.0]	2.591×10^{-8}	9.582

To demonstrate the broader impact of the next-generation missions, Table 14presents the consolidated statistical results, specifically the standard deviations, for the residual differences in gravity potential anomalies relative to the ESM2.0 ground truth across all 12 examined core IHRF stations, as initially presented in Appendix J. From Table S6.3 it is clear that the future mission concepts enhance significantly the accuracy of the IHRF realization globally. For almost all examined sites, the NGGM/MAGIC yield substantial performance gains, generally ranging between 30% and 57% relative to the GRACE-C baseline. These results prove that the NGGM and MAGIC missions are highly capable of supporting the strict accuracy requirements for the long-term sustainability and precise temporal update of the IHRF. The three stations where GRACE-C outperforms NGGM, namely EUR2, OHI3 and SYOG are located either at the maximum latitudes in the polar regions that NGGM contributes or above that. Hence, it is expected that in these regions, NGGM will not be able to improve the performance of GRACE-C.

Table S6.3: Consolidated statistics (standard deviations) of the residual differences in gravity potential anomalies (ΔV) relative to the ESM2.0 ground truth across all 12 examined core IHRF stations. Percentage values in parentheses indicate the relative performance improvement of the respective mission compared to the GRACE-C baseline. Values highlighted in red denote stations where no improvement was observed. (Unit: cm^2/s^2).

IHRF sites	ΔV: GRACE-C	ΔV: NGGM	ΔV: MAGIC
ADIS	**30.474**	**19.129 (37.2%)**	**18.569 (39.1%)**
AUT1	**33.201**	**15.736 (52.6%)**	**16.052 (51.7%)**
CUIB	**50.225**	**33.905 (32.5%)**	**32.472 (35.4%)**

EUR2	**16.851**	**43.755**	**22.217**
MKEA	**19.15**	**14.918 (22.1%)**	**10.663 (44.3%)**
NRIL	**22.266**	**9.675 (56.6%)**	**9.582 (57.0%)**
OHI3	**22.951**	**40.488**	**35.895**
SG27_ngs	**30.14**	**28.969 (3.9%)**	**16.014 (46.9%)**
SHAO	**38.187**	**33.056 (13.4%)**	**31.302 (18.0%)**
SCH2	**22.171**	**19.603 (11.6%)**	**18.661 (15.8%)**
STR2	**20.324**	**12.652 (37.8%)**	**10.267 (49.8%)**
SYOG	**24.199**	**28.068**	**12.705**

Evaluation of Mission Contributions to POD

The performance of the simulated MAGIC, NGGM and GRACE-C models data within POD is demonstrated through orbit determination and intersatellite ranging data analysis of the GRACE Follow-On and GRACE missions. The POD results are based on a pure dynamic orbit determination approach by estimating a minimum number of orbital parameters, avoiding introducing empirical forces in order to avoid capturing part of the gravity signal.

The GRACE-FO orbit residuals are shown in Figure S6.3. The inter-satellite ranging data analysis is shown in Figure S6.4, providing residuals of the Laser Ranging Interferometer (LRI) range and range-rate data. In particular, the LRI range-rate data have been used as observations within the orbit determination and parameter estimation algorithm.

The LRI analysis focuses on variations at the level lower than 10-6 m/sec2, showing a close performance between MAGIC and NGGM mission scenarios while revealing minor discrepancies for the GRACE-C mission.

In the case of the GRACE mission, POD has also been performed for the period of December 2004. The K-Band inter-satellite ranging data residuals are shown in Figure S6.5 for the models of MAGIC, NGGM and GRACE-C. The range-rate data have been applied as observations within the combined parameter estimation algorithm for orbit determination. The KBR range-rate analysis shows that MAGIC and NGGM models provide close fitting residuals, while the GRACE-C models lead to wider variations.

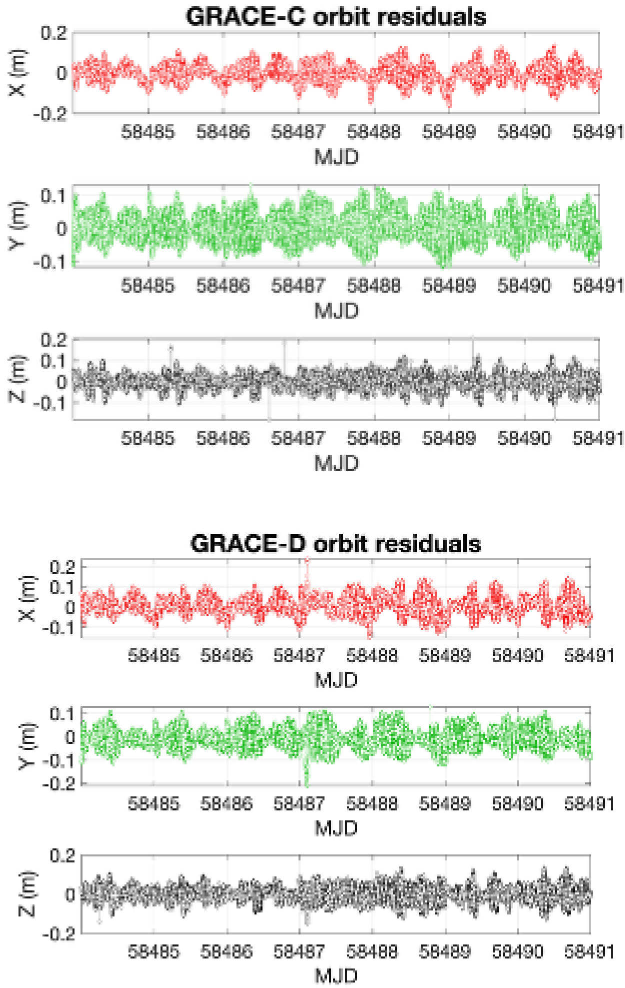
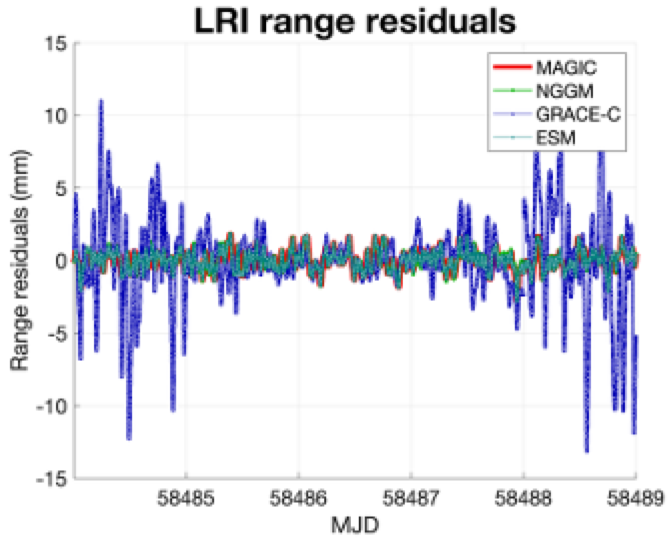


Figure S6.3: Precise Orbit determination application to the satellites of the GRACE-FO mission. Orbit residuals per epoch refer to results of 7 daily orbit arcs with overall RMS values equal to 7.7 cm and 8.1 cm for GRACE-C and GRACE-D satellites, respectively. The gravity model applied is based on the MAGIC-based harmonics coefficients (January 2006) and the EIGEN-6C4 gravity model as the background model.



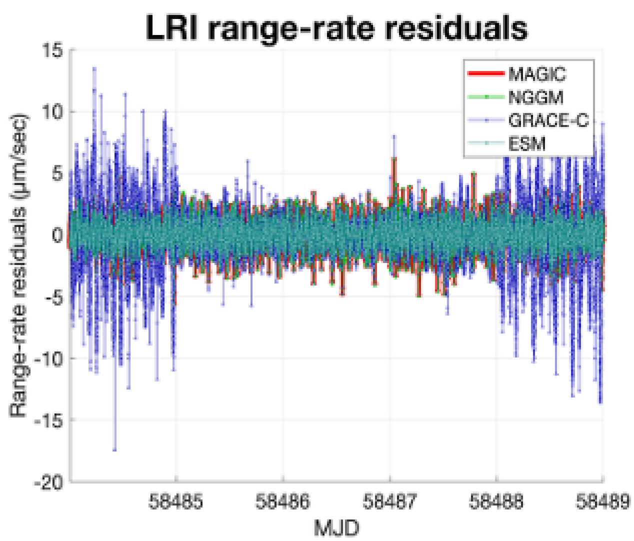


Figure S6.4: GRACE-FO Laser Range Interferometer (LRI) intersatellite range and range-rate data residuals per epoch based on the 5-day L2a models. RMS values of the range residuals are 0.706, 0.699, 3.121 and 0.712 (mm) for MAGIC, NGGM and GRACE-C missions and ESM model, respectively. RMS of the rage-rate residuals are 0.926, 0.940, 2.623 and 0.926 µm/sec for the models, respectively.

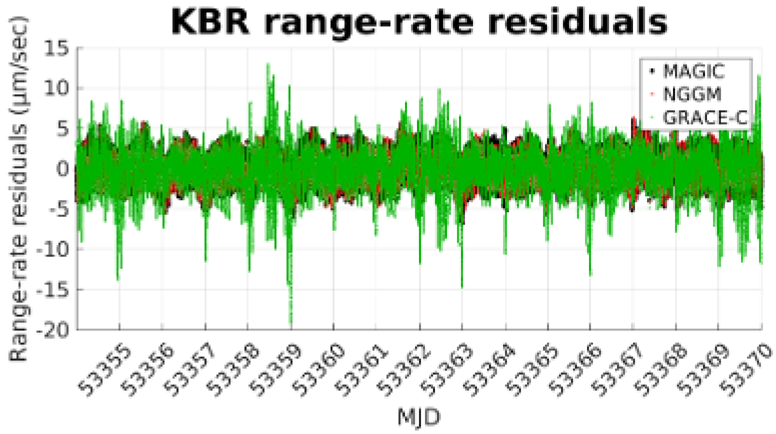


Figure S6.5: GRACE K-Band Ranging (KBR) intersatellite range-rate data residuals per epoch for a 20-day period. RMS values of the KBR range-rate residuals are 4.01, 3.95 and 4.74 µm/sec for MAGIC, NGGM and GRACE-C missions, respectively.