Geophysical Journal International

Geophys. J. Int. (2025) **242,** 1–22
Advance Access publication 2025 May 27
Research paper

https://doi.org/10.1093/gji/ggaf195

Benefit of MAGIC and multipair quantum satellite gravity missions in Earth science applications

Jürgen Kusche,[1] Christina Strohmenger,[1] Helena Gerdener,[1] Bernd Uebbing,[1] Anne Springer,[1] Yorck Ewerdwalbesloh,[1] Annette Eicker,[2] Carla Braitenberg,[3] Alberto Pastorutti,[3] Roland Pail,[4] Philipp Zingerle,[4] Marius Schlaak,[4] Mirko Reguzzoni,[5] Lorenzo Rossi,[5] Federica Migliaccio[5] and Ilias Daras[6]

[1] *University of Bonn, Nußallee* 17, *D*-53115 *Bonn, Germany. E-mail: kusche@uni-bonn.de*
[2] *HafenCity University, Hamburg, Germany*
[3] *University of Trieste, Trieste, Italy*
[4] *TU Munich, Munich, Germany*
[5] *Politecnico di Milano, Milan, Italy*
[6] *ESA/ESTEC, Noordwijk, The Netherlands*

Accepted 2025 May 12. Received 2025 March 27; in original form 2024 December 20

SUMMARY

We evaluate simulations for single-, double- and multiple-pair satellite gravimetry missions with respect to applications in hydrology, sea level budgeting and solid Earth science. We begin with the retrieval of weekly spherical harmonic solutions from GRACE-FO and MAGIC-like intersatellite laser tracking in the presence of realistic aliasing, as well as from more distant scenarios that would involve flying quantum accelerometers on satellite pairs in various orbital planes of different inclination. To account for realistic applications, we simulate the impact of such data products in basin-averaged total water storage recovery, in the retrieval of water storages via assimilation into global and regional models, in global and regional ocean mass estimation also in combination with radar altimetry and in the monitoring of Earthquakes and submarine volcano growth. While we find that the MAGIC simulation provides the largest improvement step with respect to our GRACE-FO simulation, the more advanced scenarios add sensitivity in particular in applications where gravity and mass change data can be directly equated to observable phenomena. It is more challenging to judge the benefit of advanced missions with scientific applications that involve combination with model ensembles and additional remote sensing data, as their uncertainties may determine the noise floor and will need to be projected into the future, which we did not attempt at here.

Key words: Satellite gravity; sea level change; hydrology; Earthquake dynamics.

1 INTRODUCTION AND RESEARCH QUESTION

The GRACE and GRACE-FO satellite missions (hereafter GRACE/-FO; Tapley *et al.* 2019) have provided unprecedented data for understanding the Earth system at scales down to a few hundreds of kilometres. Monthly GRACE/-FO gravity models, after conversion to total water storage anomaly (TWSA) maps, have become a mainstay in global hydrological modelling (e.g. Rodell & Reager 2023), evaluation of Earth System Models (Jensen *et al.* 2019, 2024), ocean mass change retrieval (Chambers *et al.* 2010) and the assessment of large earthquakes (Han *et al.* 2024). When assimilated into hydrological and land surface models, at basin scale (Zaitchik *et al.* 2008) or grid level (Eicker *et al.* 2014), they improve the spatiotemporal resolution and realism of hydrological modelling (e.g. Li *et al.* 2019; Gerdener *et al.* 2023). Further combination of GRACE-/FO data with radar altimetry enables one to resolve the sea level budget (Rietbroek *et al.* 2016; WCRP Global Sea Level Budget Group 2018), and retrieve estimates of the Earth's Energy Imbalance independent of radiometric sensors (Marti *et al.* 2022).

GRACE was successfully continued in May 2018 by the GRACE-FO mission (Landerer *et al.* 2020), with a gap of about eleven months. In the current GRACE-FO 'wide deadband' attitude control mode, it is estimated that its lifetime could be potentially further extended beyond its five-year design life (Flechtner *et al.* 2024a). However, both GRACE and GRACE-FO data are limited in terms of spatiotemporal resolution and data product latency. New user requirements, including the more widespread integration of data in operational services, have been translated into mission requirements in Pail *et al.* (2015) and Wiese *et al.* (2022). A recent expert survey

in the framework of the ESA QSG4EMT study (Eicker *et al.* 2024) has suggested that applications in various fields would benefit from improvements in both temporal and spatial resolution.

Mass-change And Geosciences International Constellation (MAGIC) is the ESA-NASA jointly developed concept for collaboration on future satellite gravity observations that addresses needs of the international user community and stakeholders (Haagmans *et al.* 2020; Daras 2023). MAGIC consists of a staggered deployment of the NASA/DLR-funded GRACE-C mission and the ESA-funded Next-Generation Gravity Mission (NGGM). GRACE-C, the first pair of MAGIC, is expected to be launched in 2028 in a near-polar orbit at around 500 km orbital altitude (Flechtner *et al.* 2024b) and continue the GRACE, GRACE-FO programme of record. NGGM, the second pair of MAGIC, is expected to be launched in 2032 in an inclined controlled repeat orbit at around 400 km orbital altitude (Daras 2023; Pail *et al.* 2024).

MAGIC should thus provide for at least four years data from combined operations, with the main innovation being the much reduced (when compared to GRACE-FO) effects of temporal aliasing, and data thus being much less reliant on de-aliasing models and post-processing. MAGIC will provide mass-change products at higher spatial resolution, sub-weekly temporal sampling and with shorter latency and higher accuracy (Daras *et al.* 2024b). Simulations have shown that, without the need for post-processing, smaller hydrological and ocean basins will be resolved, emergence times of climate signals (e.g. how long it takes to identify AMOC weakening) are reduced and the threshold for detecting large Earthquakes will be lowered considerably (Cambiotti *et al.* 2020; Daras *et al.* 2024b).

However, various studies have already demonstrated that for major breakthroughs in resolution beyond MAGIC either more satellite pairs need to be implemented (Purkhauser & Pail 2020; Yan *et al.* 2023) or unconventional constellations (Elsaka *et al.* 2014), which may necessitate the development of new drag-free and pointing techniques, potentially combined with new instruments based on quantum technology that alleviate some of the current noise sources.

Carraz *et al.* (2014) proposed a one-axis cold-atom interferometer (CAI) gradiometer that would use two magneto-optical traps and Bose–Einstein Condensate (BEC) cooling stages, and would generate two counterpropagating atom clouds, allowing one to measure the radial gradient and the rotation rate along an axis perpendicular to the radial direction. Due to the physical principle, CAI accelerometers and gradiometers promise flat noise down to very low frequencies and thus where the time-variable gravity signals are dominant (as an accelerometer or in gradiometer common-mode, the CAI would provide the necessary measurement of non-gravitational forces). Through closed-loop simulations, Douch *et al.* (2018), Migliaccio *et al.* (2019), Reguzzoni *et al.* (2021) and Trimeche *et al.* (2019) showed that this concept could significantly outperform the GOCE gradiometer and measure temporal variations of the gravity field without intersatellite ranging, however at the expense of technically challenging requirements on the satellite attitude and orbit control system and/or extending to a 2- or 3-axis instrument. The CARIOQA Pathfinder Mission was proposed to carry a demonstrator 1-axis CAI accelerometer (Lévèque *et al.* 2019), which would measure the non-gravitational forces acting on the platform. Subsequently, Lévèque *et al.* (2023) suggested a GRACE-FO like constellation where a laser link would measure the distance between the two satellites and thus couple two CAI instruments in order to produce a correlated differential acceleration (i.e. single-arm gradiometric) measurement.

In Rossi *et al.* (2023) these simulations were extended to multiple pairs of satellites, including Bender configurations, and equipped with ultraprecise clocks, in the framework of the MOCAST+ study. Along a similar line of thought, Bender (2022) suggests replacing the conventional accelerometers in GRACE-FO type missions by the LISA gravitational reference sensor. An overview with respect to the programmatics and road-map initiatives for developing cold atom technologies in space has been provided in Alonso (2022).

In summary, various concepts for quantum gravity missions have been proposed with the objective to deliver unprecedented data on key Earth processes, to enhance climate monitoring and cope with impacts on water cycle, for example, ground- and soil-water changes (e.g. droughts, crop yield), extreme events like flash floods and sea level rise. In the framework of the ESA Quantum Space Gravimetry for Earth Mass Transport (QSG4EMT) study, mission architectures and the development of user requirements are currently investigated (Zingerle *et al.* 2024), with the focus on low-low intersatellite tracking mission constellations, quantum gradiometry and combined concepts. Here, we report about findings from QSG4EMT with respect to specific applications that include time-series analysis in hydrological basins, assimilation into hydrological and land surface models, combination with radar altimetry and the identification of very localized land and ocean signals originating from Earthquakes and submarine volcano growth.

Several simulation studies have pointed out that errors in the tidal and non-tidal background models prevent the GRACE/-FO data products from exploiting the full potential of current instruments (Flechtner *et al.* 2016). Background model errors were typically assessed from differences across a set of 'independent' model data sets (Dobslaw *et al.* 2015, 2016), inevitably neglecting the possibility of common errors in these models due to using similar limitations in the physics or the input data. Strategies have been developed for mitigating effects due to background model errors to some extent, for example, through extended gravity field parametrizations or weighting background model uncertainty in the least-squares procedure (Wiese *et al.* 2011; Abrykosov *et al.* 2022).

Progress in ocean and atmosphere models, which are required to generate non-tidal dealiasing data sets, may be achieved through exascale computing, increasing spatial resolution and more explicit process representation such as eddy transport and mesoscale internal ocean variability (Shihora *et al.* 2024) and non-hydrostatic atmosphere physics (Springer *et al.* 2024), more widespread use of machine learning (Dueben & Bauer 2018) and the integration of new data sets such as SAR altimetry and in particular SWOT for coastal ocean tides (Monahan *et al.* 2024) and for example, Sentinel-3 Next-Generation Topography (NGT) in the future. It is however very difficult to predict how the uncertainty of ocean tide models and of numerical Earth system models will actually evolve over the next decade. Background model error bars from the 'AOe07' data set (Shihora *et al.* 2024) appear lower by a factor between 1.5 and 3, depending on region, when compared to the earlier 'AOerr' (Dobslaw *et al.* 2016), see fig. 8 in Shihora *et al.* (2024). When we simply assume that modelling errors would drop at the same rate in future, this could suggest that at the time of MAGIC model errors have dropped by another factor between 1.5 and 3 when compared to now, and by a factor of 2 to 10 at the time of launching a quantum mission. For this study, we decided therefore to stick to the error assumptions provided by Shihora *et al.* (2024), which are conservative but in line with the approach described in Daras *et al.* (2024b).

Our research question here is thus: Given the conservative assumptions that (1) modelling efforts over the next two decades will not significantly improve the atmosphere-ocean dealiasing (AOD)

process, and (2) no mission-specific post-processing or smoothing of Level-2 products, beyond what is already applied with GRACE/GRACE-FO data, will be introduced, what level of improvement in scientific applications can realistically be expected from a low-low intersatellite tracking multipair quantum constellation?

In contrast to previous studies, we include hydrology applications that involve assimilation of mass change data into models, as well as combination with radar-altimetric data for sea level science. The latter necessitates that we explicitly pay attention to error correlations at grid scale, which is often overlooked in studies that identify minimum resolvable basin area only. We are aware that both assumptions (1) and (2) may offset our assessment in terms of the maximum achievable science for any given mission scenario to the conservative side. However, it is difficult to predict the scenario-specific optimal post-processing, as this depends on background model improvement and other factors and will evolve gradually in the real world. We foresee that the level of post-processing optimization will also depend on constraints defined by specific applications, for example, long-term consistency with earlier data sets.

Within the framework of the QSG4EMT study, we picked a few high-profile science applications that would be fundamental for developing upgraded or new operational services for hazard monitoring and/or early warning, for example, for droughts and floods, ocean heat content anomalies and submarine volcano surveillance.

This paper is organized as follows: We will first describe the study logic, that is, our general mission simulation and gravity retrieval approach, error scenarios and our general assumptions with respect to application scenarios. We will then discuss selected applications that we chose to be representative for the mission objectives, together with results for four mission scenarios that include a GRACE-FO simulation as a reference. Finally, we synthesize our findings in the conclusions (Section 4).

2 STUDY LOGIC

2.1 Mission and sensor simulations

The hypothesis underlying the QSG4EMT study is that multiple pairs of GRACE-FO- or MAGIC-like satellites, equipped with advanced laser ranging instruments (LRI) and quantum cold-atom interferometers (CAI) for measuring non-conservative forces and possibly gravity gradients, would provide the most promising design for retrieving temporal variations of the geopotential. Through higher accuracy and stability as compared to today, these constellations will enable significant progress in applications as mentioned above. In order to test this hypothesis, we had to carry out simulations that account for instrument-specific noise and realistic background model errors, mimic advanced gravity retrieval methods and emulate the subsequent ingestion of level-2 data into application-specific analysis frameworks. We decided to include both GRACE-FO-like and 'MAGIC-like' simulations to provide a reference for our simulation of quantum pairs, at the same level of simulation detail.

In the course of the QSG4EMT study, it became clear that the time-variable gravity field retrieval performance will not be primarily limited by the performance of future instruments (Zingerle *et al.* 2024) if their performance is equal or better than what is assumed for the ESA NGGM mission (Daras *et al.* 2024b). The main limiting factor is then temporal aliasing caused by an insufficient spatiotemporal observation coverage and a non-fitting parameter model. Regarding the simulations, the main scientific focus has thus been laid on mitigating temporal aliasing. In an attempt to keep a conservative approach, we decided to concentrate on the impact of increasing the spatiotemporal coverage by increasing the constellation size and by still retaining the typical static parametrization approach. For the choice of instruments, this means that we tried to avoid any additional impact of this error source by assuming the most optimistic future noise model for our simulations. As part of the QSG4EMT project results, Encarnação *et al.* (2024) provide an in-depth evaluation of a range of possible future noise models for the low-low satellite-to-satellite tracking concept (LL-SST, similar to GRACE-FO), where we here selected the most promising one. This model consists mainly of an improved LRI instrument with a projection of the performance developments to the year 2040 and CAI accelerometers, which have been designed in a manner that the combined performance is not significantly affected by them. For simulating the GRACE-FO-like and MAGIC-like reference missions, we used the appropriate noise models as given in Massotti *et al.* (2021) and Encarnação *et al.* (2024). In Fig. 1, we summarize the amplitude spectral densities of the overall observation noise (i.e. the sum of all individual noise contributors) for the three instrument noise models; it becomes obvious that the noise of a CAI mission is assumed as 2 to 3 times below GRACE-FO or MAGIC over the entire spectrum.

Mission simulations in the QSG4EMT study were conducted in a numerical closed-loop environment, where we employed two different retrieval strategies. The first strategy was developed by Technical University Munich (TUM) and the second one by Politecnico di Milano (PoliMI).

In the TUM simulations, a linear acceleration approach as detailed in previous studies (Murböck *et al.* 2014; Schlaak *et al.* 2023) was used. In a linear least-square adjustment, weekly gravity field solutions were retrieved as spherical harmonic coefficients for each assumed constellation (Section 2.2) to resolve the Hydrology (H), Ice (I) and solid Earth (S) mass transport signals. The LL-SST observations are simulated as acceleration differences caused by the gravity differences at a given orbit position in time, considering the corresponding constellation's sensor noise models. The gravity differences are modelled by the ESAESM time-variable mass transport model with a 6-hr resolution (Dobslaw *et al.* 2015). To represent the observations realistically, the Atmosphere (A) and Ocean (O) components of the full Atmosphere-Ocean-Hydrology-Ice-Solid Earth (AOHIS) signal in the ESM are reduced by a de-aliasing model (Dobslaw *et al.* 2016), taking de-aliasing errors into account (Shihora *et al.* 2024). In addition, Ocean Tide (OT) model errors are considered as typical aliasing sources by including model differences using EOT11a (Savcenko *et al.* 2012) and GOT4.7 (Ray 2008) OT models. Furthermore, the Solid–Earth (S) component of the ESAESM only represents a single earthquake event. Therefore, the S-component has been updated (S-upd) to include nine earthquake events with their respective co- and post-seismic signatures over the ESAESM period to enable this study to investigate the recoverability of earthquake events. In summary, the target signal for this study is HIS-upd. The simulation results retrieve this signal with a weekly resolution, considering de-aliasing model errors, OT model errors and instrument noise.

In the PoliMI simulations, the same LL-SST observations were simulated, considering the same background models. Again, a linear least-square adjustment was implemented to estimate weekly gravity field solutions as spherical harmonic coefficients. To model the temporal aliasing contribution, the formal error covariance matrix

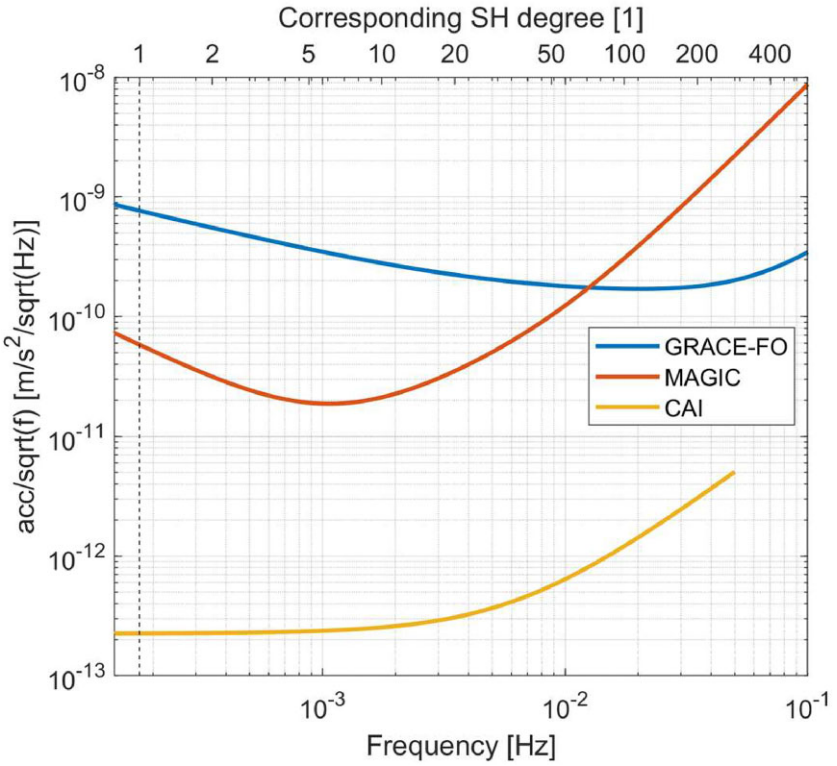

Figure 1. Amplitude spectral densities of the overall observation noise for the different scenarios in terms of acceleration errors in line-of-sight direction. Blue (top) line: noise specification similar to GRACE-FO. Orange (middle) line: noise model according to the ESA NGGM mission threshold requirements. Yellow (bottom) line: future instrument noise model using CAI accelerometers, which has been assumed for the QSG4EMT study.

was rescaled on empirical variances computed from a simultaneous Monte Carlo simulation (Migliaccio *et al.* 2009), where noise samples were generated from instrumental error power spectral densities. The main difference with the TUM simulations is a second-step refinement (Migliaccio *et al.* 2006; Rossi *et al.* 2023), computing an intermediate grid of some functionals of the anomalous potential, for example, total water storage expressed as equivalent water height (Wahr *et al.* 1998). As in a standard remove-restore procedure observations were reduced by the least-square solution, and residual values on the grid were computed by collocation. More specifically, collocation was applied over local data patches in which the global data set was previously subdivided (Reguzzoni & Tselfes 2009). An isotropic signal covariance function was derived from the degree variances of the rescaled least-square error covariance matrix and localized for the gridding of each data patch by adapting the variance only (Reguzzoni *et al.* 2014). Corrections to the least-square spherical harmonic coefficients were computed by applying a discretized quadrature formula (Colombo 1981) to the estimated global grid. However, these corrections are small for the considered simulations.

As a consequence, we find that TUM and PoliMI retrievals for the four scenarios provide very close solutions (i.e. within error bars at TWSA or gravity anomaly grids), albeit with some expected differences in the error representation. Given that this study intentionally does not focus on optimizing the stochastic model of the specific TWSA data product in, for example, data assimilation but rather on how these data perform in existing frameworks that were developed for GRACE, we decided to continue only with the TUM solutions for now. We believe the above shows that our results are robust towards the choice of the specific retrieval approach. However, the results of PoliMI simulations were used to compute regional solutions consisting of local grids of TWSA, where the signal covariance function was tailored on the specific area under study (see Section 3.5).

2.2 Error scenarios for this study

The retrieved weekly gravity solutions are computed for four mission constellation scenarios with one, two, three and six satellite pairs (Fig. 2). For the single-pair mission a polar orbit (inclination 89°) and GRACE-FO-like sensor noise is assumed. Its performance is comparable to the current GRACE-FO mission. The double-pair constellation resembles the MAGIC mission's coverage and noise assumptions with a polar pair and second pair at an inclination of 70°. For future missions, a three-pair (CAI3) and a six-pair (CAI6) scenario are used, assuming the improved CAI noise and the additional satellites at lower inclinations (see Table 1). The orbital planes are chosen to optimize the spatial sampling geometry while maintaining a homogeneous enough sampling to keep a stable adjustment system.

All orbits are integrated over four weeks and split into 7-d orbits with a 10 s sampling rate (limited by CAI sampling rate). The entire ESM timeframe of 12 yr is simulated in this study. To avoid accumulated integration errors over the 12-yr timespan, the integrated 4-week orbits are repeated every month. To avoid resonances with the input signal, the orbits are also shifted by 17° in the right ascension of the ascending node for each month. The resulting unconstrained weekly gravity field solutions are resolved, in terms of spherical harmonics, up to degree and order, d/o, 120 (corresponding to roughly 166 km spatial resolution). Due to an insufficient ground-track coverage, the single pair is only resolved up to d/o 100 (200 km).

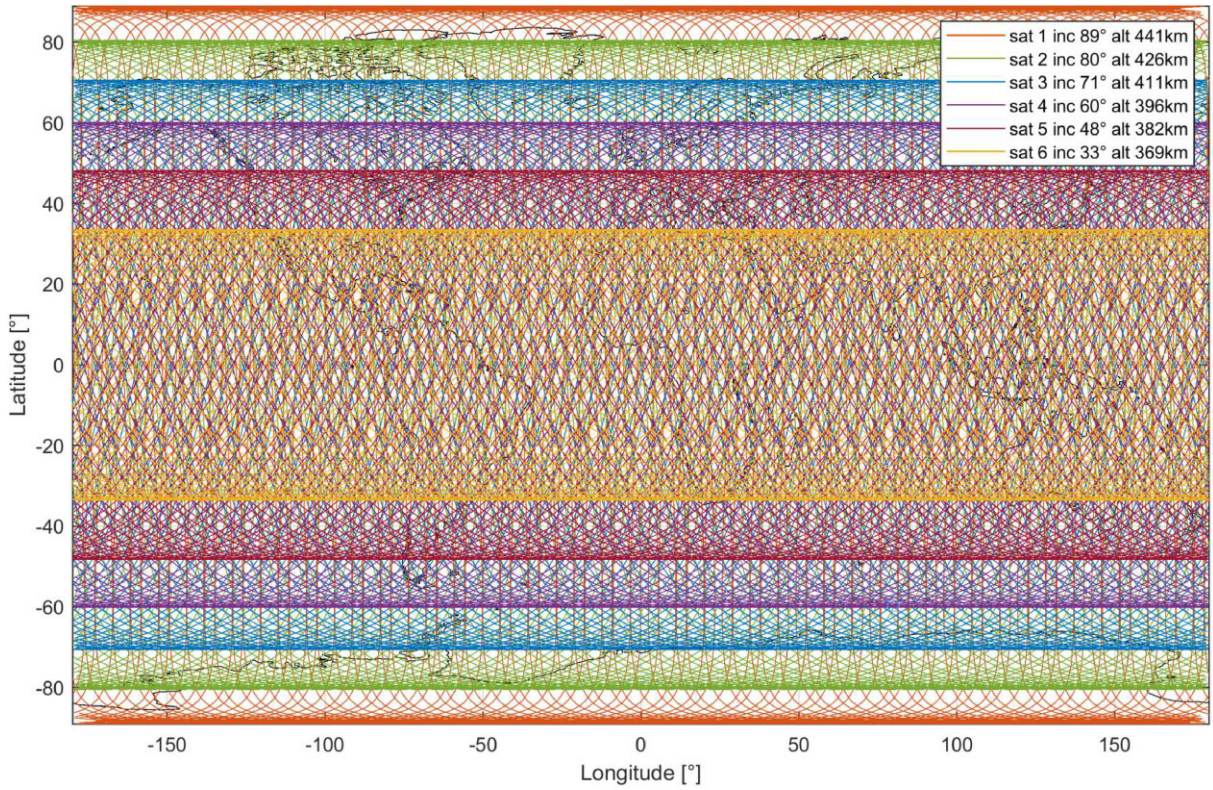

Figure 2. Illustration of the ground track pattern of the future 6-pair CAI scenario (CAI6) after a period of 5 d corresponding to the 6-pair constellation described in Zingerle *et al.*(2024).

Table 1. Settings used for assessing mission scenarios in the QSG4EMT study

Scenario	Satellite pairs	Instruments	Remarks
GRACE-FO 1-pair	89°	GRACE-FO	'GRACE-FO like'
MAGIC 2-pair	89°, 70°	MAGIC	'MAGIC like'
C9AI3 3-pair	89°, 70°, 40°	CAI	'future small'
CAI6 6-pair	89°, 80°, 71°, 60°, 48°, 33°	CAI	'future large'

2.3 Errors that result for the application case studies

We intend to simulate applications of future data products from a user-centric perspective, so it is helpful to briefly reflect about current and potential new user groups. Users of level-3 data, that is, TWSA, ocean bottom pressure or gravity maps, may not necessarily be acquainted with the intricacies of gravity mission processing. New applications may be first driven by scientific curiosity and later merge into semi- or fully operational work flows. Users may be categorized as 'expert users' e.g. as being familiar to GRACE/-FO data and open to adapting their work flows to obtain 'best' results e.g. in terms of resolution, or they may be 'standard users' in that they do not wish to work with e.g. spatially correlated error representations, or simply want to continue with a processing established for GRACE/-FO, or they may be 'bulk data users' that integrate various data sets within a single framework. For example, for expert users certain post-processing choices such as spatial smoothing, rescaling and dealing with spatial correlation and anisotropic error patterns would be equally meaningful, while accounting for omission errors due to limited spatial resolution and for signal leakage would clearly differ for these groups, while all this may be less important to bulk data users as long as a gap-less low-latency data stream is provided. It is also important to note that in data assimilation and data fusion applications, space-gravimetric data are optimally merged with model simulations or other remote sensing, typically under the least-squares paradigm where one has to implement error assumptions for data and model. This means that impact studies such as ours will always depend on the respective error models, which for data assimilation are driven by model initialization, parameter and forecast errors, and which may include error models for other relevant data sets. In this study, we mainly target the broader group of non-expert users and therefore avoid to individually tailor the post-processing of scenario output data.

The weekly gravity solutions that we retrieve from simulated orbit and instrument data can be directly compared, in the spherical harmonics domain, with the spherical harmonic coefficient (SHC) sets obtained from the modified ESAESM model, which still refers to the 1995-2006 time frame, and this approach provides 'true' errors for each simulation week. Average maps of error standard deviation are shown in Fig. 3 below. Note that we show errors for unfiltered (d/o 60) solutions to emphasize the differences across scenarios here, whereas in the applications all solutions are DDK-filtered. As expected, aliasing errors introduce a prominent land-ocean signature with significant spatial correlations, which is best visible here in the MAGIC plot, due to the colourscale.

Data assimilation and data fusion applications require an error representation which can be expressed through a positive definite

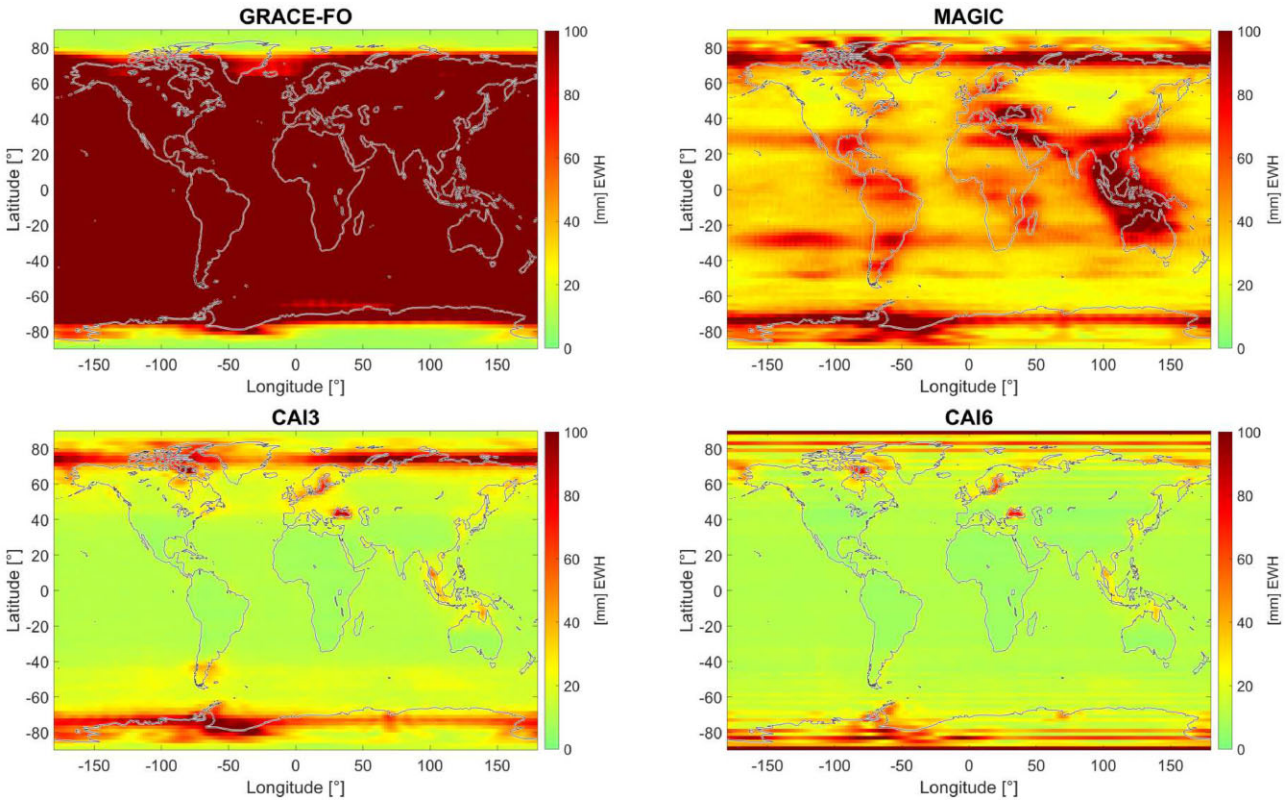

Figure 3. TWSA sigma resp. standard deviation maps (d/o 60, unfiltered) for the four scenarios (GRACE-FO, MAGIC, CAI3 and CAI6), from ESAESM. Note that in the development of data analysis (e.g. assimilation) schemes simplified error patterns are used.

error covariance matrix, and this poses challenges when spatial correlations are to be taken into account. Since the sample variance covariance matrices (empirical error covariances as in Daras *et al.* 2024a) are necessarily rank-defect, we decided to circumvent this problem in the following way: (1) we accumulate weekly to monthly normal matrices, (2) these monthly formal normal matrices are then scaled degree-wise to match the degree variances of the sample covariance matrix and (3) these scaled normal matrices are subsequently inverted to fully populated SHC variance covariance matrices and projected to the TWSA grid. This approach has its limitations as it, for example, is not able to correctly map error and correlation discontinuities at continental boundaries, yet it allows us to take average error magnitudes and spatial correlations into account that result from orbital patterns, and matches to the magnitude found from empirical simulations for comparable wavelengths. We emphasize here that our approach differs from Daras *et al.* (2024b) in the spatial patterns of errors. However, it is known that the sample (empirical) variance covariance matrix, when based on a small number of simulations only, represents a poor estimator in particular for the off-diagonal terms and should not be relied on for assessing spatial correlation. We believe that our approach here is on the conservative side.

In operational data assimilation frameworks, it has been often impossible or impractical to specify spatial (or temporal) observation error correlation (Miyoshi *et al.* 2013). While this is typically not a problem for assimilating, for example, pixellated soil moisture data, TWSA data expressed in spherical harmonics can in principle be projected onto any grid, and disregarding real correlations will thus lead to an overweighting of these data as compared to model simulations or other remotely sensed data. A way out is then selecting the grid spacing just such that it decorrelates; that is, corresponding to the real, 'effective' data resolution. As an alternative, researchers have developed *ad-hoc* approaches to derive spatial correlation for TWSA maps (Forman & Reichle 2013).

Generating maps of TWSA based on GRACE/-FO data generally requires some spatial smoothing (filtering). While this may not be required for generating standard data products with MAGIC and/or future quantum missions, we expect that for the sake of consistency of time-series many climate-related applications will simply continue with the post-processing that was optimized for GRACE and GRACE-FO. For this study, we decided to avoid any ambiguity introduced by applying different or mission-specific filters and use the same [DDK-type; Kusche (2007)] approach. We note here that our conservative filtering approach will diminish the performance differences between scenarios to some extent.

3 APPLICATIONS AND RESULTS

3.1 TWSA retrieval in hydrological basins

The analysis of time-series of water storage variations in hydrological units such as river basins and aquifers have been among the most common applications of satellite gravimetry. Understanding the dynamics in river basins provides crucial information on the availability of water resources, builds the basis for climate predictions and quantifies pre-conditions of hydrological extremes such as flood conditions (Reager *et al.* 2014). When combined with data on hydrometeorological fluxes, water storage change can additionally provide the basis for the closure of the terrestrial water balance enabling a comprehensive understanding of hydrological systems in response to climatic, environmental and anthropogenic changes (Springer *et al.* 2014). However, the relatively low spatial

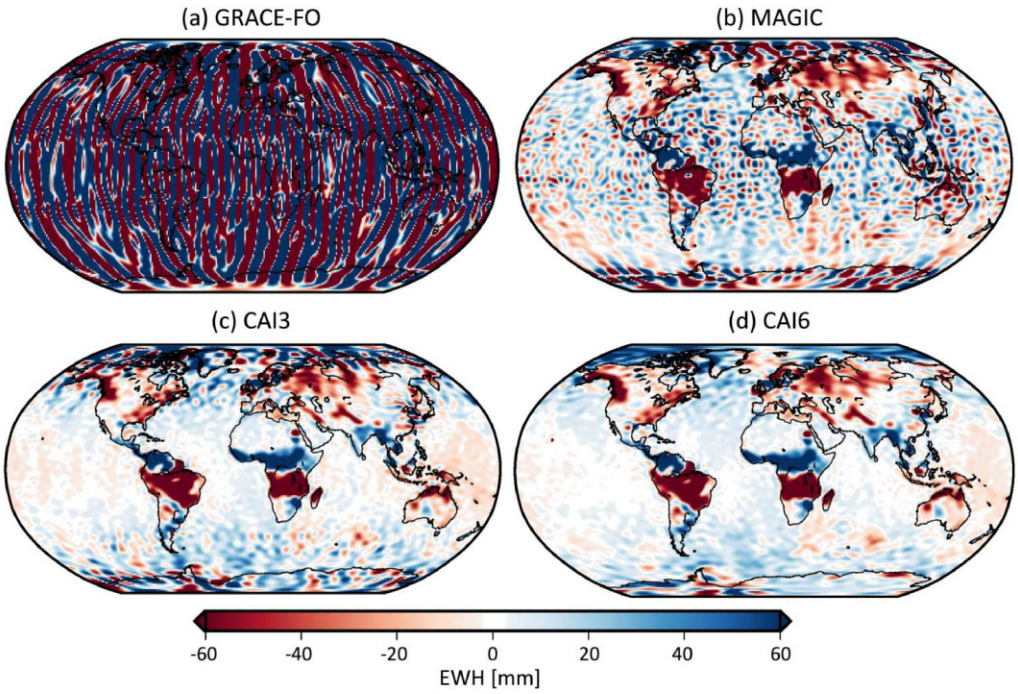

Figure 4. Gridded EWH for a single exemplary 7-d solution, DDK5 filtered.

and temporal resolution (as compared to grid width and time-step of hydrological models) limits the use of the mass change data by a wider community in water cycle studies and water resources assessments. Water management or drought risk assessment often operate at country or administrative levels and for them the value of water storage information tends to increase with higher spatial resolution (Meza *et al.* 2020). In this study, we follow Daras *et al.* (2024a) and investigate the retrieval of basin-average water storage time-series for more than 400 largest individual river basins worldwide under the above defined mission scenarios. The temporal root-mean-square deviation (RMSD) between the simulation output and the reference signal computed for each river basin will be used to assess the accuracy of the simulation results.

As described earlier, we focus on an assessment of the four different scenarios for retrieving basin-averaged total water storage anomalies, when the same spatial resolution (d/o 90 here) is to be resolved and the same post-processing applied to all simulation outputs. Fig. 4 shows global maps of equivalent water heights for an exemplary 7-d solution (beginning of January 2002) after applying a DDK5 filter, and this reveals the pronounced difference in signal-to-noise ratio between scenarios. For the simulated GRACE-FO scenario it is not possible to see any signal, as (differing from the common processing of the real GRACE-FO data) we assume weekly retrievals here, and the DDK5 method applies less smoothing than in most real-data studies. Also the MAGIC solution appears noisy but the large-scale hydrological signals emerge. In contrast, it appears that the three-pair quantum CAI3 scenario would already show strong improvement, while adding additional pairs in the CAI6 constellation would further reduce the noise.

We then derive uncertainties, estimated empirically as root mean square difference of the retrievals with respect to the ESAESM 'truth' model, of basin-averaged time-series for more than 400 hydrological basins (according to GRDC definitions) in Fig. 5. We again use DDK5 filtered solutions until d/o 90 and and compare to the unfiltered ESAESM reference up to the same d/o, thus including the filter omission error. To assess which accuracies can be achieved for the majority of the river basins, thresholds were computed for which the RMSD of 70 per cent of the river basins are smaller. These thresholds are indicated by horizontal lines in Fig. 5 and amount to 19.5 cm (GRACE-FO scenario), 2.3 cm (MAGIC scenario), 1.5 cm (CAI3 scenario) and 1.0 cm (CAI3 scenario). We thus suggest that accuracies of approximately 1.5 cm (CAI3) and 1 cm (CAI6) for 220 km spatial resolution is realistic for hydrological applications at weekly resolution; better results might be obtained by tailored post-processing.

In other words, when aiming for cm-accuracy at shorter timescales, CAI6 could resolve hydrological units from around 3000 to 5000 km^2, that is, at least one order of magnitude smaller than the minimum basin-averaging area required with MAGIC. We suggest that while with MAGIC investigations into, for example, groundwater storage change would be limited to only the very large aquifers as defined by van der Gun (2022), while CAI6 would allow to assess aquifers down to medium size. Examples of relevant aquifers for which storage decline and recharge would be difficult to isolate with MAGIC but could be a science target for CAI6 are the San Joaquin aquifer (26 000 km^2; Ojha *et al.* 2019) with critical importance for water supply in California, or the Chu aquifer with geopolitical relevance as a trans-boundary groundwater supply for Kazakhstan (7500 km^2 corresponding to 40 per cent of the aquifer area; Liu *et al.* 2020) and Kyrgyzstan.

3.2 Assimilation for global hydrological modelling

Data assimilation (DA) provides a way to integrate satellite-derived TWSA with hydrological model simulations in a statistically optimal manner (Kalnay 2002; Zaitchik *et al.* 2008). It allows to correct model and forcing data deficiencies at scales where satellite gravimetry provides credible information, while preserving high-resolution information from model forcing and boundary (i.e. topography, land cover, soil) data. DA also enables one to disaggregate TWSA into compartmental storage anomalies, and combine with

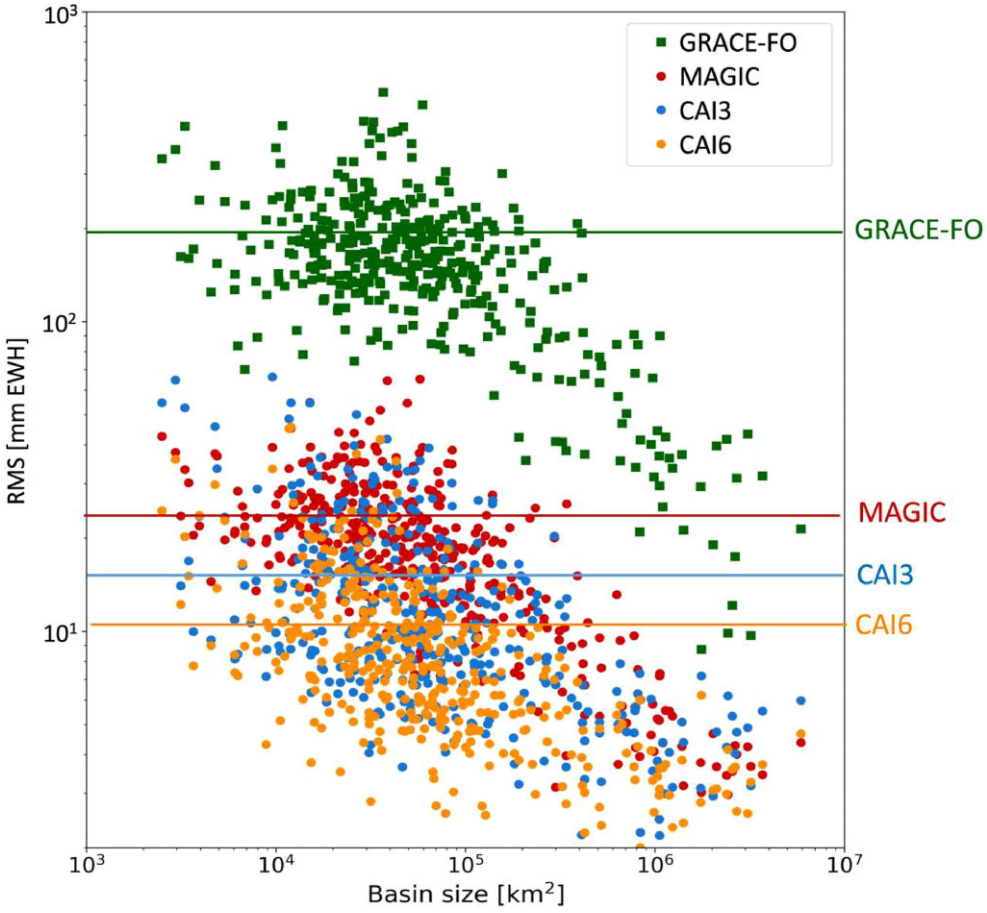

Figure 5. Scatter plot of RMSD of basin average time series versus size of the river basin for different mission scenarios, for simulated DDK5 filtered weekly solutions truncated at d/o 90, including filter omission error. Horizontal lines: 70 per cent of basins below threshold.

remote sensing data such as snow cover maps (Wang *et al.* 2021). A number of teams have implemented assimilation with GRACE/-FO data, for example, to provide high-resolution water storage data, investigate groundwater and droughts (Li *et al.* 2019) and assess the response of the terrestrial biosphere to precipitation (Gerdener *et al.* 2022).

In this study we follow Gerdener *et al.* (2023) and investigate the potential benefit of the above discussed scenarios in an Ensemble Kalman filter (EnKF) assimilation framework, with and without modifying the current DA setup that was developed for integrating GRACE/-FO observations with the WaterGAP (Müller Schmied *et al.* 2021) model. Different from Gerdener *et al.* (2023), we find that for MAGIC and quantum scenarios we can assimilate TWSA maps indeed at the native 0.5° resolution of the model, due to improved resolution of these missions over GRACE-FO. We thus assimilate monthly TWSA simulated with WaterGAP and perturbed with errors that correspond to the four scenarios as discussed above. We implement the DA via the PDAF framework, which had been coupled with WaterGAP in the online mode (Nerger *et al.* 2020) for numerical efficiency. Data uncertainty representation in the DA is simplified via a single scaled SHC normals matrix as above, propagated to the grid while applying the DDK5 filter (Kusche 2007), while the uncertainty of the hydrological model (i.e. climate forcing and parameter uncertainties) which is required to be specified in DA is kept as in Gerdener *et al.* (2023). With this setup, we indeed assume the simulations of water storages to represent the 'truth'; thus we expect that for more advanced scenarios the assimilated TWSA should be closer to the unperturbed model run (whereas the disaggregation of TWSA into individual storages will be still limited by the assumed model uncertainty). Due to the expensive computational costs, the study is run here for South America; however results have been found as transferable to other continents.

As mentioned above, the WaterGAP model has to serve both as the 'truth' in this experiment for TWSA but also individual model storages, and as the input basis for data assimilation and thus the retrieval of simulated level-4 data sets. It is important to understand that as a consequence of this, in an ideal scenario with vanishing gravity mission errors the DA should indeed replicate this truth (for TWSA), whereas in the other end-member situation with mission errors larger as compared to errors in the simulated hydrology model, the DA would reproduce the 'open-loop ensemble mean' which is not exactly identical to the simulated truth.

Since the GRACE/-FO data do not resolve the common 0.5° grid spacing of global hydrology models, TWSA maps are often aggregated to a more realistic spatial resolution before DA; in Gerdener *et al.* (2023) to 4° grids. Fortunately, we find that for the simulated MAGIC and quantum scenarios no such aggregation appears necessary, and therefore in the following we were able to assimilate 0.5° maps directly. However, there is a trade-off with the representation of spatial correlation at smaller scales; we noted that sampling observation errors within limited ensembles from the dense covariance matrices would generate spurious long-range correlations, and as a result we had to suppress spatial correlations here.

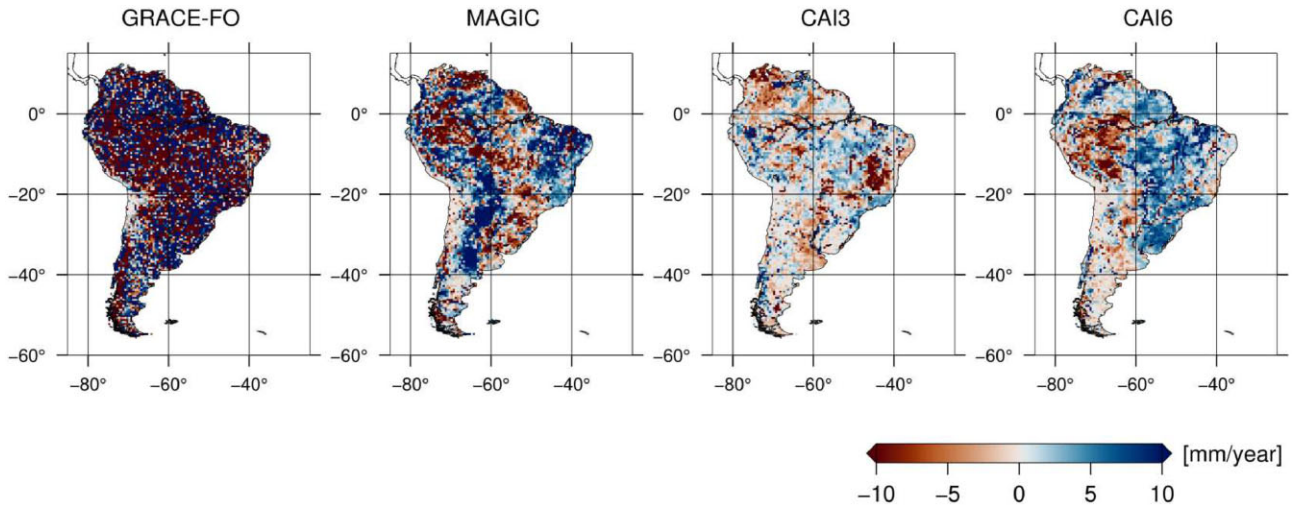

Figure 6. Differences of TWSA trends between truth and outcome of the WaterGAP assimilation when assuming empirical (i.e. formal errors scaled to empirical ones) errors according to the four scenarios considered here. Trends were estimated using multilinear regression for the period 2003 to 2019.

Fig. 6 shows the differences between TWSA trends between 'truth' (WaterGAP) and 'retrieved' for our four scenarios after assimilation, shown here for South America and estimated from monthly DA covering the years 2003 to 2019. As mentioned before, input and output TWSA maps are given on an 0.5° grid. The GRACE-FO scenario here shows significantly higher trend errors (differences) as compared to the other scenarios, this is related to the lack of spatial resolution and if we would assimilate simulated GRACE-FO data accumulated to 4°, these differences on the 0.5° model grid would be lower by a factor of about four. Differences drop dramatically for the 3-pair scenario, with a slight degradation here observed for the 6-pair quantum mission.

Similar conclusions can be drawn when comparing the spatial patterns of the retrieved annual TWSA amplitude differences (Fig. 7) for the GRACE-FO and MAGIC scenarios against the simulated truth. For the three-pair scenario differences with respect to the truth are much smaller, while we hardly observe further improvements with the 6-pair scenario. In summary, all three scenarios MAGIC, CAI3 and CAI6 show very clear improvements compared to the GRACE-FO scenario on the trend, annual and monthly timescales, as expected.

In our data assimilation experiment, we are able to verify how well TWSA maps are partitioned into the major water storages, here groundwater (GWSA), surface water (SWSA) and root zone (0.1 to 4 m) soil moisture (SMSA) storage anomalies, where SWSA refers to the aggregation of river, wetland, lakes and reservoir storage anomalies. Spatial mean RMSD for comparing these storages inferred from the DA to the truth WaterGAP simulation are provided in Table 2, for linear trends, annual amplitudes and annual phases at the 0.5° pixel scale. Similar as with TWSA, we find an improved performance indeed when advancing from the GRACE-FO-like simulation to MAGIC or the quantum scenarios also for groundwater storage change map retrievals, which is expected since groundwater variability dominates the total water storage signal in the region. For example, the spatial mean RMSD of the assimilation result for the annual amplitudes for South America drops by half with the MAGIC scenario, and another about 20 per cent for the CAI3 and CAI6 scenarios. However, this is not always the case for the smaller contributors surface water and soil moisture variability, which we ascribe to the uncertainty of the underlying model simulation.

These errors will obviously be smaller when one moves from a 0.5° grid representation to spatial averages, either at basin scale or for coarser grids. We provide RMSD values for aggregated 1.0° and 3.0° grids in brackets in the table, and this allows us to compare across spatial scales. We find that TWSA trend errors (1 cm yr^{-1}) for MAGIC (and 3-/6-pair quantum scenarios) at 0.5° correspond approximately to TWSA trend errors for GRACE-FO at 3.0° scales after assimilation. GWSA trend retrieval errors for the advanced scenarios at 0.5° would be lower or similar to GWSA trend errors for GRACE-FO at the 300 km scale.

In other words, it is demonstrated here that via data assimilation, groundwater storage trends can be derived at the 1 cm yr^{-1} uncertainty level for areas as small as 2500 km^2 for CAI3 and 6, and for regions of the order of 10 000 km^2 for MAGIC. This suggests that MAGIC assimilation products could, for example, resolve about 44 per cent of the NUTS2 (Nomenclature des Unités territoriales statistiques–level 2, European Commission 2003) regions—the administrative unit where the EU collects water and land use information. CAI assimilation products would allow us to resolve even about 88 per cent of NUTS2 regions, and this could serve, for example, for monitoring agricultural or industrial water use.

TWSA amplitude errors for MAGIC or the quantum scenarios at 100 km scale would be smaller than TWSA amplitude errors for the simulated one-pair mission at 300 km scale, and surface water anomaly errors for the advanced scenarios at 100 km appear lower than for GRACE-FO at the 300 km scale. However, we need to remind the reader again that these numbers are affected by the WaterGAP model uncertainty that we specify currently in our real-data analyses with GRACE and GRACE-FO, and this may not reflect future knowledge in Earth system modelling.

3.3 Assimilation for regional land surface modelling

Land surface models (LSMs) represent the coupled cycling of water, energy and other biogeophysical matter such as carbon within vegetation and soils (Fisher & Koven 2020). They simulate land-atmosphere interactions and provide thus a key component of climate models. In LSMs, assimilation of satellite-derived TWSA enables one to inform the entire vertical profile of soil moisture, in

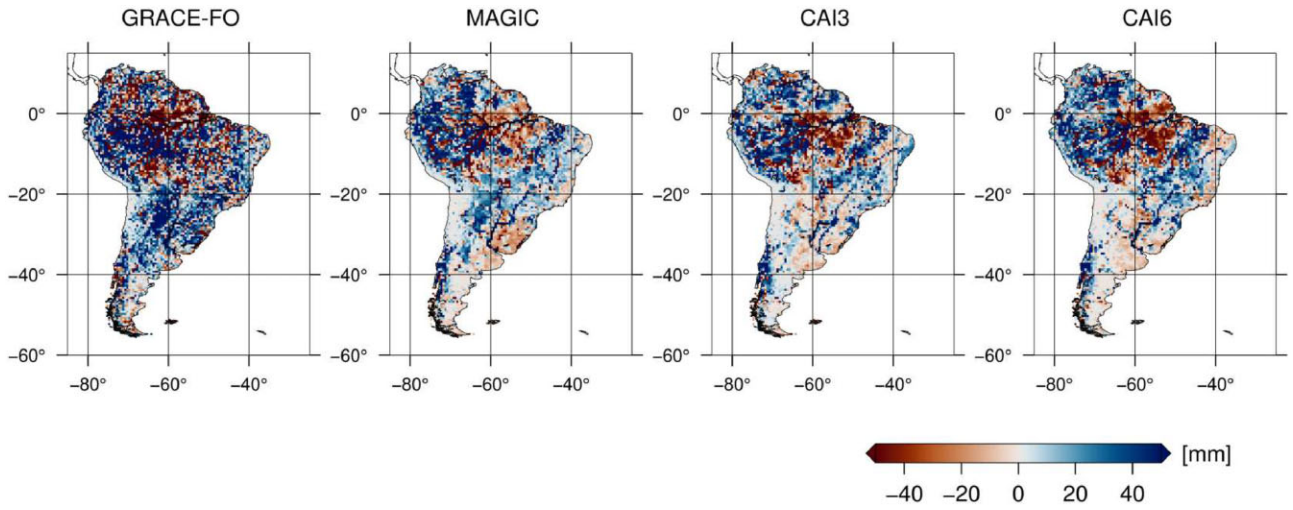

Figure 7. Differences of annual TWSA amplitudes between truth and outcome of the WaterGAP assimilation when assuming empirical (i.e. formal errors scaled to empirical ones) errors according to the four scenarios considered here. Amplitudes were estimated using multilinear regression for the period 2003 to 2019.

Table 2. RMSD between gridded maps of water storage, 0.5° resp. 1° and 3° (in brackets) from the WaterGAP assimilation and the simulated truth. Trend differences are in mm yr^{-1}, differences of annual amplitude in mm and phase differences in months. All RMSD are derived for 2003 to 2019, spatially averaged for South America, and shown here for TWSA, groundwater (GWSA), surface water (SWSA) and soil moisture (SMSA) storage anomalies

		RMSD			
Signature	**Scen.**	**TWSA**	**GWSA**	**SWSA**	**SMSA**
Linear trends (mm yr^{-1})	GRACE-FO	44.5 (27.4, 13.3)	40.4 (25.2, 12.6)	14.5 (7.9, 3.9)	1.8 (1.6, 1.3)
	MAGIC	16.2 (10.6, 7.1)	14.3 (9.8, 6.7)	10.4 (5.7, 3.3)	2.3 (2.0, 1.6)
	CAI3	12.7 (8.9, 4.3)	11.5 (8.3, 3.8)	11.9 (4.2, 2.0)	3.2 (2.9, 2.0)
	CAI6	13.8 (8.9, 4.8)	13.1 (8.8, 5.4)	22.2 (12.8, 3.3)	2.6 (2.2, 3.5)
Annual ampli. (mm)	GRACE-FO	122.1 (79.1, 45.5)	46.9 (37.5, 29.9)	114.4 (75.6, 49.3)	32.3 (29.8, 25.7)
	MAGIC	65.5 (42.1, 23.2)	18.7 (16.7, 14.1)	89.0 (54.5, 29.2)	41.0 (40.3, 40.7)
	CAI3	53.8 (35.0, 20.9)	23.7 (21.2, 18.7)	81.9 (44.8, 19.5)	55.6 (51.9, 37.8)
	CAI6	54.5 (36.2, 20.7)	25.1 (23.4, 20.7)	156.6 (44.8, 19.1)	49.1 (45.5, 34.3)
Annual phases (mo)	GRACE-FO	3.4 (2.5, 1.9)	3.7 (2.7, 2.0)	4.5 (3.2, 2.4)	3.8 (3.1, 2.4)
	MAGIC	2.5 (1.7, 0.9)	2.5 (2.1, 1.6)	3.5 (2.5, 2.1)	3.5 (2.9, 2.4)
	CAI3	2.6 (1.7, 1.0)	2.6 (2.2, 1.7)	3.4 (2.4, 2.4)	3.9 (3.3, 3.2)
	CAI6	2.6 (1.7, 1.0)	2.5 (2.0, 1.5)	3.2 (2.2, 1.3)	4.2 (3.5, 2.9)

contrast to *in-situ* and satellite soil moisture (SSM) or land surface temperature (LST) assimilation, and for example, to identify problems in soil processes representation (Swenson & Lawrence 2014).

In this study we investigate the effect of assimilating simulated data from GRACE-FO, MAGIC and quantum accelerometer multipair scenarios CAI3 and CAI6 into the Community Land Model CLM (Oleson *et al.* 2004). We follow Naz *et al.* (2019) and Springer *et al.* (2019) in the use of the TerrSysMP-PDAF assimilation framework (Shrestha *et al.* 2014; Kurtz *et al.* 2016), which in principle allows us to integrate observations of SSM, LST and others jointly with TWSA. CLM is set up here as in Springer *et al.* (2019) for the Euro-CORDEX region at 12 km grid scale, and forced by COSMO-REA6 meteorological data (Bollmeyer *et al.* 2015). In our experiments, we use the most recent version CLM version 5.0 (CLM5, Lawrence *et al.* 2019) as the 'truth' model that is perturbed using the errors from the mission scenarios and assimilate into CLM3.5; in other words we seek to correct for both random and non-random deficiencies of the model implemented in the DA framework. Results will be compared as basin-averages for the major catchments in Europe as in Springer *et al.* (2017) with area down to about 25 000 km^2.

Within our regional observing system simulation experiment, the water storage changes simulated by CLM5 are accumulated to gridded monthly TWSA over Europe, used as a reference truth, and averaged to the observation grid space at 2° resolution. We first observe that the CLM5 simulations provide distinctly different TWSA over Eastern Europe, with more pronounced seasonal variations and extremes compared to CLM3.5 (Fig. 8). Previous investigations have demonstrated that CLM3.5 underestimates these signals compared to GRACE observations over the region (Springer 2019), which we attribute to deficiencies in the representation of snow processes or inaccuracies in the atmospheric forcings. Long-term trends in CLM3.5 are too high in TWSA (wetting) in Central Europe and too low (drying) in parts of Eastern Europe—in fact CLM5 runs suggest that most of Europe appears as drying in the GRACE/-FO period. In the context of our assessment here, it means that we pose a somewhat more realistic and potentially challenging experiment as compared to our global assimilation experiment, as the satellite-derived data now have to 'compensate' not only for random model

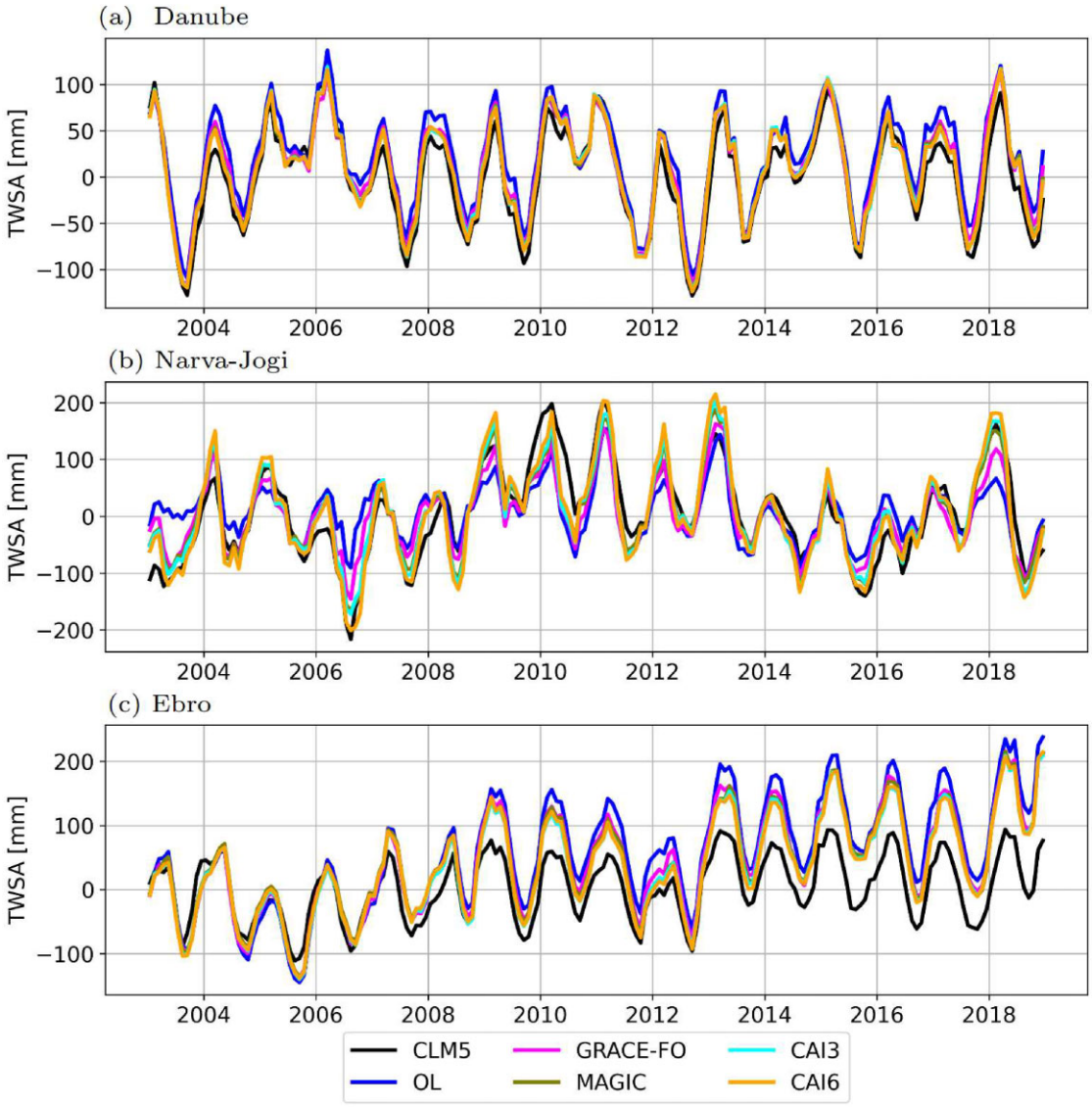

Figure 8. Basin-averaged TWSA for Danube, Narva and Rio Ebro basin. CLM5 truth, CLM3.5 open-loop simulation and assimilation results for the four mission scenarios are represented.

noise, but also systematic errors due to mismodelled processes. In other words, we expect the assimilation to pull the model simulation towards the truth, against the incorrect model dynamics. While this is indeed the reality in assimilation systems, we are not aware of similar impact simulation studies in the literature.

As in the WaterGAP DA, the synthetic truth observations are perturbed with correlated noise which is generated from the scaled full formal error covariance matrices for the four different mission scenarios GRACE-FO, MAGIC, CAI3 and CAI6. The previous WaterGAP-DA results and the following results were derived with the EnKF filter, where within the DA scheme only the main diagonal of the error covariance matrix is taken into account (i.e. assumed as known to the developer of the DA scheme).

In Fig. 8 below, we show three exemplary time-series plots of basin-averaged TWSA, that is, for the Danube basin (area 817 000 km^2, the Rio Ebro basin (area 80 093 km^2) and the Narva basin (area 56 225 km^2). In these plots, the CLM5 truth, the CLM3.5 open-loop simulation and the assimilation results for the mission scenarios are provided.

We find the representation of the seasonal cycle as well as of interannual signals, that is, dry and wet years, in the open-loop simulation increasingly improved when ingesting simulated satellite-gravimetric data. For the Rio Ebro basin, a systematic and growing offset between truth CLM5 and our CLM3.5 simulations can only be slightly reduced.

Maps in Fig. 9 demonstrate correlations for all basins considered in this study. Indeed, the largest improvements are generally obtained for CAI6, while MAGIC, CAI3 and CAI6 perform all clearly better than GRACE-FO (Table 3). Since the MAGIC constellation shows similar results as the CAI3 constellation, the figures of MAGIC are left out. Data assimilation generally improves the representation of month-to-month variability, the annual cycle as well as the representation of the more extreme years. In this study, however, in some of the basins data assimilation was not able to completely correct for time-variable trend errors, that is, overcome an incorrectly represented model dynamic even for multipair mission scenarios. We suspect the cause may be an unrealistic ensemble spread, related to our scheme for generating perturbations of soil parameters.

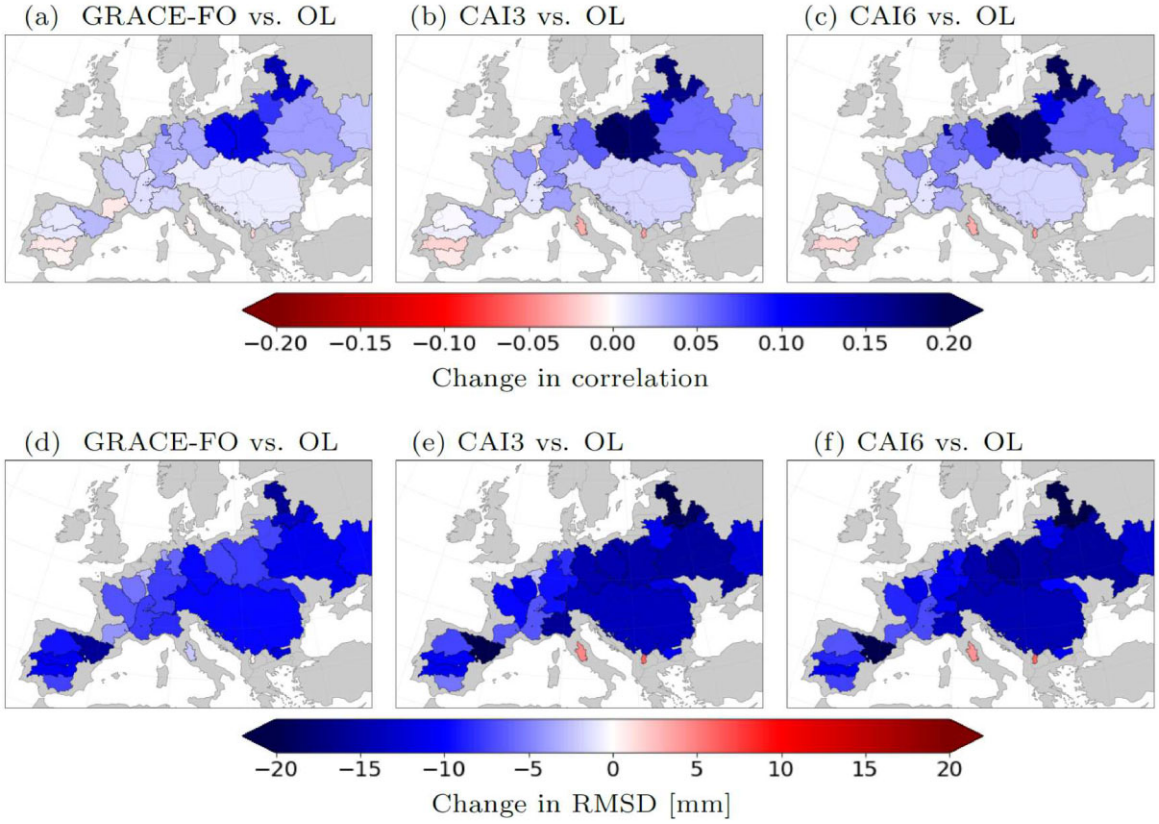

Figure 9. Correlation and RMSE for basin-averaged TWSA in the CLM3.5 model simulation (OL) and for the results of assimilation with the four scenarios, as compared to the truth (CLM5 simulation).

Table 3. RMSE (mm), correlation and trends for three selected river catchments.

	Scen.	RMSE (mm)	Corr	Trend (mm yr^{-1}]
Danube	CLM5			−0.3
	OL	25	0.97	0.2
	GRACE-FO	16	0.98	0.2
	MAGIC	13	0.98	0.2
	CAI3	12	0.99	0.1
	CAI6	11	0.99	0.2
Rio Ebro	CLM5			3.3
	OL	71	0.88	12.4
	GRACE-FO	56	0.91	10.5
	MAGIC	53	0.91	9.7
	CAI3	50	0.91	10.0
	CAI6	49	0.91	9.6
Narva-Jogi	CLM5			2.4
	OL	55	0.74	−1.8
	GRACE-FO	40	0.88	0.0
	MAGIC	31	0.92	1.1
	CAI3	33	0.92	1.3
	CAI6	32	0.93	1.8

Potential science targets for improved land surface modelling are ubiquitous in climate sciences, hydrology and carbon budgeting, but could be identified as well in geohazard monitoring. For example, we suggest that advanced future missions have the potential to enhance the monitoring and prediction of rainfall-triggered landslides, particularly for large-scale events involving mass movements on the order of cubic kilometres. Yet, the direct detection of such events may still be limited to the largest events—typically those larger than 5–10 km^3, corresponding to approximately 10 per cent of all recorded landslides (Kirschbaum *et al.* 2010). However, the improved spatial and temporal resolution from assimilation in high-resolution land surface models would also allow for near real-time monitoring of root-zone soil moisture (Felsberg *et al.* 2021), in particular when combined with satellite soil moisture products which are operationally provided at resolutions of one kilometre to few 10 km but with coarse orbital sampling. Since soil moisture is a key

Table 4. Percentage of error level improvement (per cent) for individual sea level budget components based on the individual quantum simulation scenarios. All results are relative to the MAGIC scenario. Contributions of each mass component to the overall ocean mass error are provided in the third column.

Basin	CAI3 versus MAGIC	CAI6 versus MAGIC	Contribution to mass error
Antarctica	38.4	46.1	1.5
Greenland	38.0	53.3	1.5
Glaciers	42.7	59.9	5.0
Hydrology	46.4	57.6	5.0
Internal mass variability	6.5	9.8	87.0
Mass sum	6.5	9.8	100.0
Steric	0.5	0.7	–
Total sum	2.5	3.8	–

factor influencing shear stress along topographic inclines and thus landslide susceptibility, these advancements can be essential for refining landslide risk assessments and early warning systems, such as NASA's Landslide Hazard Assessment for Situational Awareness (LHASA) Model (Stanley *et al.* 2021).

3.4 Global ocean mass and sea level budget

The imbalance of water mass fluxes into and out of the global ocean leads to a change in total ocean mass, while at regional scale also barotropic mass rearrangements play a significant role. Ocean mass change (OMC) responds to mass loss of the Greenland and Antarctica ice sheets, coastal and land glaciers, net river discharge driven by terrestrial hydrology and changes in the land-ocean precipitation–evapotranspiration balance. In combination with measured or modelled sea level variations, estimates of OMC can be utilized to construct sea level budgets (Rietbroek *et al.* 2016; Dieng *et al.* 2017; WCRP Global Sea Level Budget Group 2018; Uebbing *et al.* 2019; Hakuba *et al.* 2021; Horwath *et al.* 2022), that is, attributing the altimetric sea level to individual mass and steric contributors and better understanding contemporary sea level drivers. This is in particular relevant for steric sea level changes which are related to ocean warming and salinity change, and which are notoriously difficult to obtain from *in-situ* measurements for the deep ocean.

Approaches exist for retrieving contributions either individually and/or as a residual (Chambers *et al.* 2010; WCRP Global Sea Level Budget Group 2018; Hakuba *et al.* 2021; Horwath *et al.* 2022) or jointly in an inverse approach (Rietbroek *et al.* 2016). Here, we first look at the inverse approach as it involves an explicit weighting between gravimetric spherical harmonic coefficients and (along-track) altimetric sea surface heights based on *a-priori* error information, and thus, facilitates error propagation for mission scenarios. Our approach works via least-squares fitting of several hundreds of dominating (*a-priori*) spatial fingerprints of sea level change to monthly gravimetric and altimetric data (Rietbroek *et al.* 2016; Uebbing *et al.* 2019); these fingerprints include static 'passive' sea level patterns related to ice sheet, glaciers and land hydrological changes derived through the sea level equation, steric patterns that we derive from an EOF decomposition of modelled thermo- and halosteric expansion, and patterns that we interprete as 'internal mass variability' and that are mostly related to barotropic motions in the major ocean basins. For this study, we replace the GRACE/-FO error model in our inversion processing (Uebbing *et al.* 2019) by the normal equations from each of the scenarios introduced above, while keeping the radar altimeter error model (assuming Jason-1 and Jason-2, with errors 5-10 cm derived from binning 20 Hz data into 1 Hz blocks standard deviation) unmodified.

Our sensitivity studies with respect to sea level budget partitioning should be understood as the result of straightforward error propagation with simplified mission scenario errors, which are integrated with assumptions on altimetry errors. We follow the forward-modelling-fitting inversion framework as outlined in Rietbroek *et al.* (2016) and shown to match nearly exactly the 'direct' approach for deriving global ocean mass change from real GRACE data in Uebbing *et al.* (2019). We find in our error propagation study that the ocean mass retrieval error will be dominated for all scenarios by the internal ocean mass variability, followed by the uncertainty of the land hydrology contribution, and only minor contributions from the glacier and ice sheet components (Table 4). This is an expected result since the mass balance for these larger regions can be determined quite well already now, and the spatial resolution is not very relevant for the global mean sea level contribution. In other words, the formal errors for global or large-basin ocean mass change are already very low with the GRACE and GRACE-FO mission; and we thus compare potential relative improvements with respect to the MAGIC scenario here, that is, we derive the error improvement (per cent) by comparing formal 1σ-errors of three- and six-pair CAI scenarios to the corresponding MAGIC-derived 1σ.

Fig. 10 summarizes the reduction in error level with respect to the simulated MAGIC scenario. In addition, Table 4 provides global average error level improvements for each of the individual subcomponents of the sea level budget. For the 3-pair and 6-pair CAI scenarios we find a general improvement in ocean mass error levels (Fig. 10), as well as for the individual contributions solved within the inversion; the reduction in error standard deviation reaches up to 60 per cent for the six-pair scenario (Table 4). On the contrary, the improvements for the steric level contributions and the 'explained total sea level' are close to zero; this is expected as they are dominated by the altimetry errors, which had been assumed as present-day errors for all experiments.

Along a tropical band, in Fig. 10 we find regions with slightly higher error level for the quantum scenarios. These errors result from the internal ocean mass variations contribution, for example, barotropic redistributions of ocean mass, which represents the most dominant error source in the mass budget (Table 4). This is also the most spatially variable error source as it is closely related to the dealiasing signal, which is removed during gravity L2 processing. However, the dealiasing error is not comprehensively modelled in the individual simulations, and so we may consider this as an artefact of our overall assessment approach. On the other hand, disregarding this effect by excluding a fingerprint for such internal variability in the inversion scheme would likely lead to overoptimistic errors.

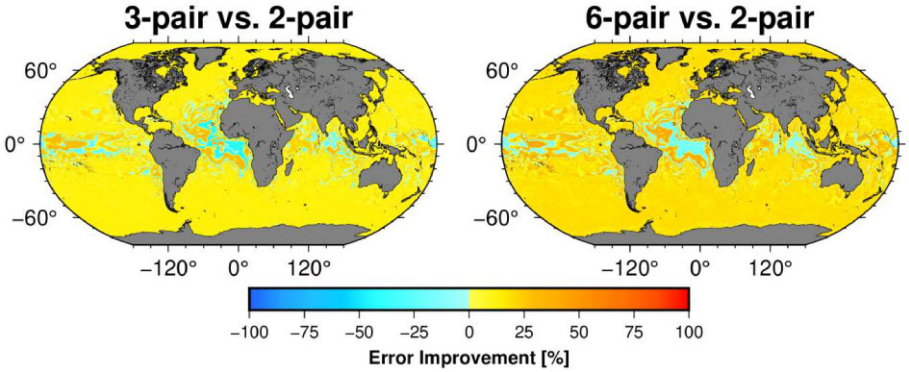

Figure 10. Map of error level improvement of the ocean mass change component based on individual quantum simulation scenarios (3- and 6-pair CAI); results are relative to the MAGIC-like scenario.

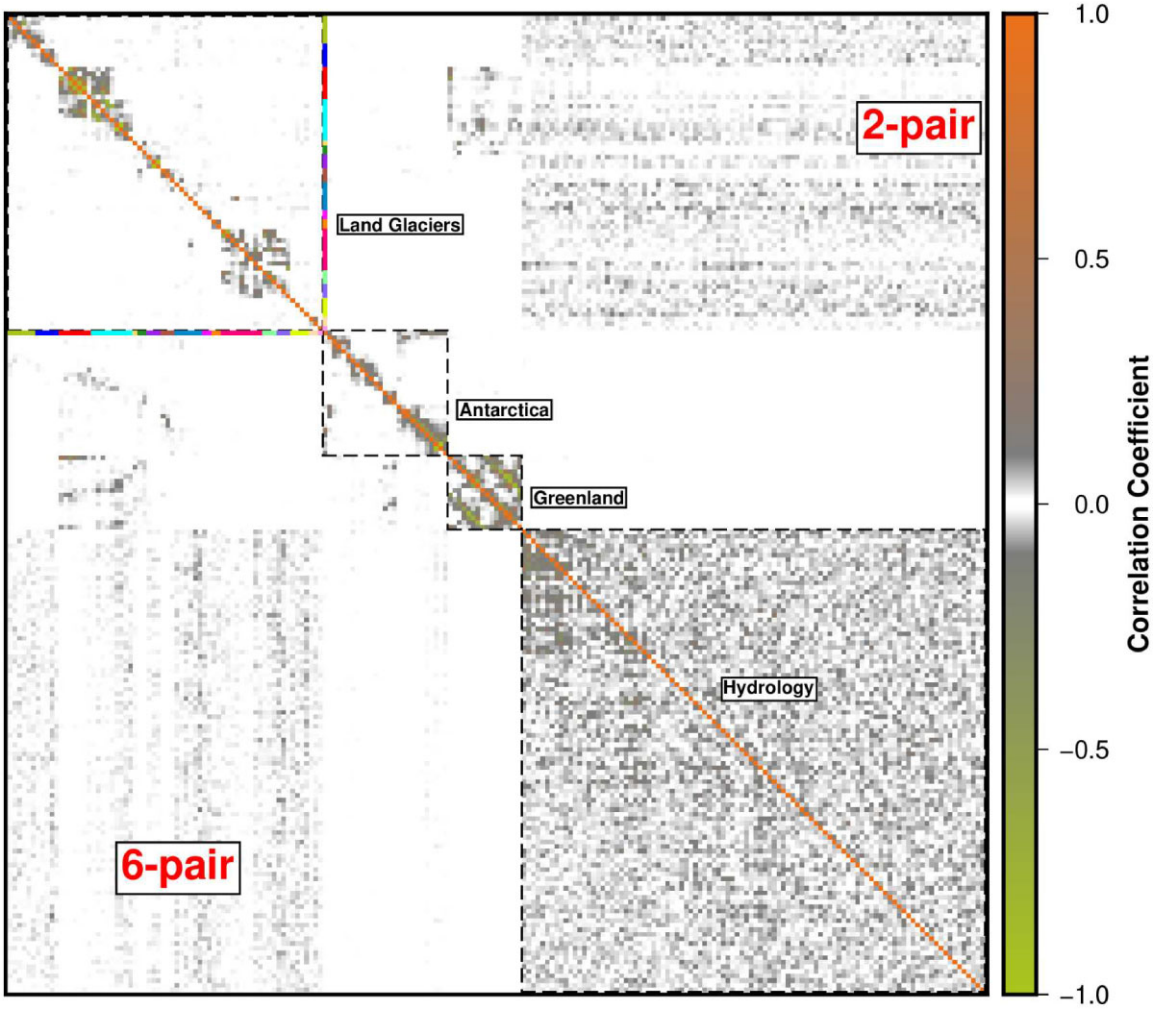

Figure 11. Improvement in signal separability from reduced correlations between individual sea level components exemplarily shown for June 2006. We compare the correlations of MAGIC (upper triangle) to those of a six-pair CAI solution (lower triangle). The figure only shows the major mass components for land glaciers and ice sheets Greenland and Antarctica as well as variations in terrestrial hydrology.

Moreover, it turns out that in our sea level inversion scheme, the major contribution of advanced gravity missions is identified in terms of source separability, rather than in reduction of (already small) errors. Our results show that the correlation between individual solved-for fingerprints is reduced (i.e. the separability of individual contributors to sea level is improved); this is demonstrated in Fig. 11. Despite our somewhat heuristic error covariance modelling for the scenarios, we find a general improvement in signal separability for individual basins and groups of mass sea level contributors. We identify a reduction in interbasin correlations, especially for the Arctic Canada North and South glaciers (red and cyan). Similarly, interbasin correlations for Greenland and Antarctic ice sheets are

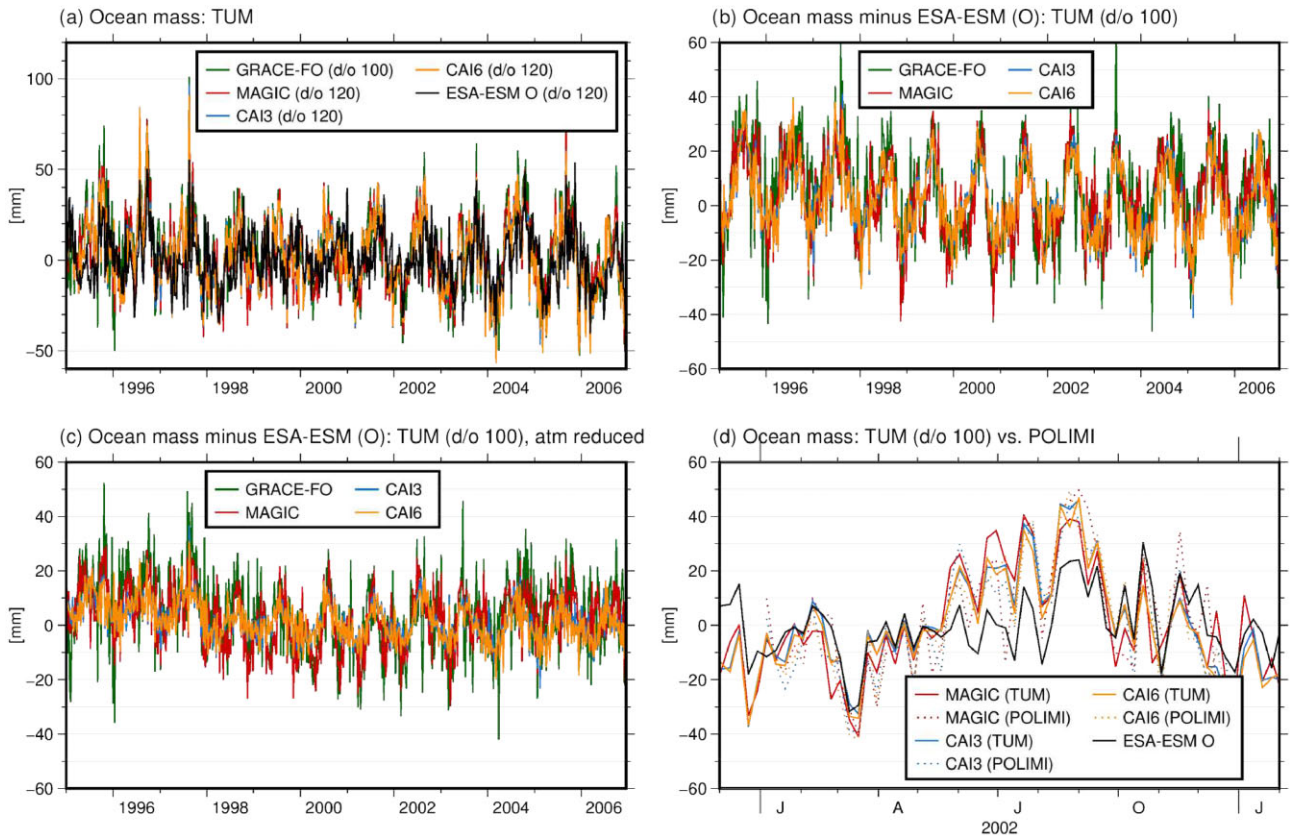

Figure 12. Time-series of ocean mass change in the East China sea for different simulation scenarios/retrievals and for the ESAESM (O) reference. All time-series (except POLIMI) are DDK3-filtered. (a) Ocean mass change in the ECS from TUM simulations and ESAESM; (b) Ocean mass change in the ECS from TUM w.r.t. ESAESM; (c) same as (b), but the weekly averaged 'atm' or A coefficients are additionally reduced; (d) Ocean mass change in the ECS from TUM, POLIMI and ESAESM.

also reduced. Furthermore, some decrease in correlation between individual sources of the terrestrial hydrology contribution is observed (Fig. 11).

In summary, these results suggest that signal separability will improve significantly with future gravity mission concepts, but we point out that exploiting these data will also require a thorough and realistic error covariance modelling to maximize the resolution of such inverse solutions. We suggest that a potential science target could be the mass change of Greenland and Antarctic peripheral glaciers, for which the separation from the surrounding ice-sheet's mass balance is not possible today without relying on auxiliary data such as ice-altimetry.

3.5 Regional ocean mass retrieval

While the global sea level budget is often considered as closed, it is unclear whether this holds also true for regional sea level budgets. These budgets are generally more difficult to derive, due to a variety of local physical effects, as for example, wind-driven sea level changes or sedimentation, but also retrieval challenges related to satellite gravimetry and altimetry. Earlier studies (e.g. Frederikse *et al.* 2016; Rietbroek *et al.* 2016) pointed out problems and demonstrated limited agreement in many areas. The improved spatial resolution associated with future gravity missions is indeed expected to aid in better closing budgets at regional scale.

Therefore, here we also consider the common 'direct' method where ocean mass and volumetric sea level are combined at global or regional basin-average scale (Chambers *et al.* 2010), as this enables us to study the impact of spatial resolution for challenging regions in a more straightforward way. In order to derive regional OMC from the simulated (i.e. ESAESM) gravity fields, we add the GIA correction (A *et al.* 2013) to the spherical harmonic coefficients, apply DDK3 filtering (Kusche 2007) and restore the operational AO dealiasing product (Dobslaw *et al.* 2016) (see Section 3.5 for a discussion of AO and O). After converting to water heights and computing the basin average, we apply a leakage correction derived from DDK3-filtering the LSDM hydrology component of the ESAESM.

Basin averages are shown below for the East China Sea (ECS), a Western Pacific marginal sea with an area of 770 000 km^2. What makes this region unique and challenging for budget studies is its complex current system, the mostly shallow bathymetry and a large amount of sediments that is transported by rivers or results from coastal erosion. *In-situ* observations, as, for example, tide gauge data, for external validation are relatively sparse. Simulation results are thus compared to the 'truth' from the ocean component of the ESAESM.

In what follows, we compare regional ocean mass change for the ECS from simulated retrievals for all mission scenarios with the 'direct' method (Fig. 12a), to the 'truth' mass change represented in the ESAESM data set. In the figure, the ocean component of the ESAESM, that is, the true signal to be recovered, is shown as reference. It is evident that retrievals from all scenarios retrace the large seasonal oscillation of ocean mass well, with the GRACE-FO

Table 5. Errors of ocean mass change in the ECS (mm)

Scenario	GRACE-FO	MAGIC	CAI3	CAI6
RMS TUM	18	15	13	13
RMS TUM (atm reduced)	15	11	7	7
error propagation TUM	23	09	6	7

solution exhibiting the largest noise and the quantum CAI simulations with a lower noise floor, as expected. In order to better assess the performance of the individual scenarios, we next remove the ESAESM truth signal from the outcomes of the individual scenarios, as shown in Fig. 12(b). In other words, the smaller the differences are, the closer the simulations are to the reference truth, with RMSD 1.8 cm for GRACE-FO, 1.5 cm for MAGIC and 1.3 cm for both three- and six-pair CAI missions (Table 5). There is an obvious annual signal of up to 4 cm, which we hypothesize is in part attributable to remaining aliasing errors.

In order to derive ocean mass change maps from the simulated gravity fields, the O dealiasing data product (Dobslaw *et al.* 2016) should be restored. However, in the ESAESM setting this is only available together with the atmospheric dealiasing component as AO. Even though the atmospheric part should be small, it could lead to larger discrepancies here that we would like to avoid. Thus, in Fig. 12(c) the weekly average of the 'atm' coefficients from Flechtner *et al.* (2015) is subtracted, which reduces the discrepancies further and results in RMSD of 1.5 cm (GRACE-FO), 1.1 cm (MAGIC) and 0.7 cm (CAI3 and CAI6) (Table 5). These numbers are of the same order compared to what we derive from error propagation with the scaled formal covariance matrices (Section 2.3), that is, 2.3, 0.9, 0.6 and 0.7 cm, respectively.

For this experiment, we had the opportunity to compare the outcome of TUM gravity retrieval simulations with the regional POLIMI approach (Reguzzoni *et al.* 2014), where the signal covariance function applied in collocation had been tailored to the specific area, that is, the ECS, and which therefore does not require further DDK or other filtering. A comparison for the year 2002 is shown in Fig. 12(d); we find the difference between TUM and POLIMI retrievals smaller as compared to the difference with respect to the ESAESM reference. This confirms our assumption that our results are robust with respect to the specific recovery approach.

Various studies have suggested that at regional scales, sediment inflow from large river systems like the Yangtze River can significantly impact ocean mass change and the sea level budget. Sediment load is originally controlled by erosion caused on various timescales by tectonics, climate and weather, but in many deltas modified by deforestation, agriculture, river damming and sand mining (plastic debris is thought to add to few 0.001 Gt yr^{-1}; Schmidt *et al.* 2017). Today, it is possible to derive the sediment deposition mass rate from combining satellite altimetry and GRACE-FO for regions down to about 80 000 km^2, corresponding to about 30 per cent of the world's largest rivers (Mouyen *et al.* 2018). From our results above, we suggest that OBP data from future CAI multipair constellations would enable one to monitor sediment deposition rates for at least about 60 per cent of the largest river systems, with the same accuracy as with the GRACE-FO data today.

3.6 Earthquake and submarine volcano signal sensitivity

The fault rupture recovered by seismologic observations generates also long period deformations and gravity signals detected by GRACE/-FO (e.g. Chao & Liau 2019; Tanaka & Heki 2014) which are part of the entire process, but are elusive to seismological instrumentation because they do not necessarily generate seismic waves, but generate long-lasting deformation rates and changes in the gravitational field. The slow changes are not only observed postseismically (Han *et al.* 2015, 2016) but also anticipating the main shock rupture, as observed for the Maule, Chile earthquake (Bouih *et al.* 2022). Some of these slow signals have been observed to be much bigger than the coseismic signal, demonstrating long-lasting alteration of the crust due to the earthquake rupture. The above-mentioned various studies have used data from the GRACE/-FO missions to detect the imprint of major earthquakes on gravity change, revealing the post-seismic signal generated by a combination of afterslip on the fault and relaxation of the lithosphere (Tanaka & Heki 2014), adding an independent signal to test seismologically derived source models. The seismogravity signals detected by GRACE/-FO are limited to megathrust events, but we expect that the improved performance of the MAGIC mission and in particular future scenarios with quantum sensors on multipair constellations will push the noise limit, resulting in an improved time latency, and detection of smaller magnitude compared to GRACE and GRACE-FO.

The Singkil Earthquake (Indonesia, 2005 March 28, magnitude 8.6, thrust mechanism) was chosen as a representative example to evaluate the expected improvements in resolution (here derived with the weekly solutions). The coseismic pattern is derived as the difference of the gravity signal in the week following the event and the week before the earthquake. Retrieval errors are defined as the difference of the observed signal and the simulated earthquake signal included in the time-variable gravity field to be recovered, and overlaid with hydrological and other signals from the ESAESM model. The simulated Earthquake rupture has been retrieved from the USGS—National Earthquake Information Center, which routinely publishes fault rupture mechanisms based on the worldwide seismographs networks, through inversion of the seismic wave field. The event is modeled with a point source model given the Earthquake moment tensor, which defines the source for calculating the coseismic gravity field for a self-gravitating, spherically symmetric, isotropic and viscoelastic layered standard Earth model (Wang *et al.* 2017). If the focus would have been the exact modelling of the specific source, a finite source model composed of a series of patches, each defined by its moment tensor, would be the choice, as well as the analysis of the effects of a refined layered earth model. In the present study with the focus on demonstrating the relative improvement of the four scenarios, we chose to use a simple point source of given magnitude and fault plane mechanism.

The global census of seamounts is based on the gravity field derived from satellite altimetry (Wessel *et al.* 2010); while more recently the SWOT mission has been proposed for detecting seamounts through an automatic analysis of the vertical gravity gradient (Yu *et al.* 2024). The problem of detecting unrest of submarine volcanoes is of particular relevance for remote locations, and is complicated by the fact that volcano growth up to several hundreds of metres can occur without violent eruptions and degassing of magma, when the pressure of the overlying water column is sufficiently large. An example is the Fany Maoré submarine volcano,

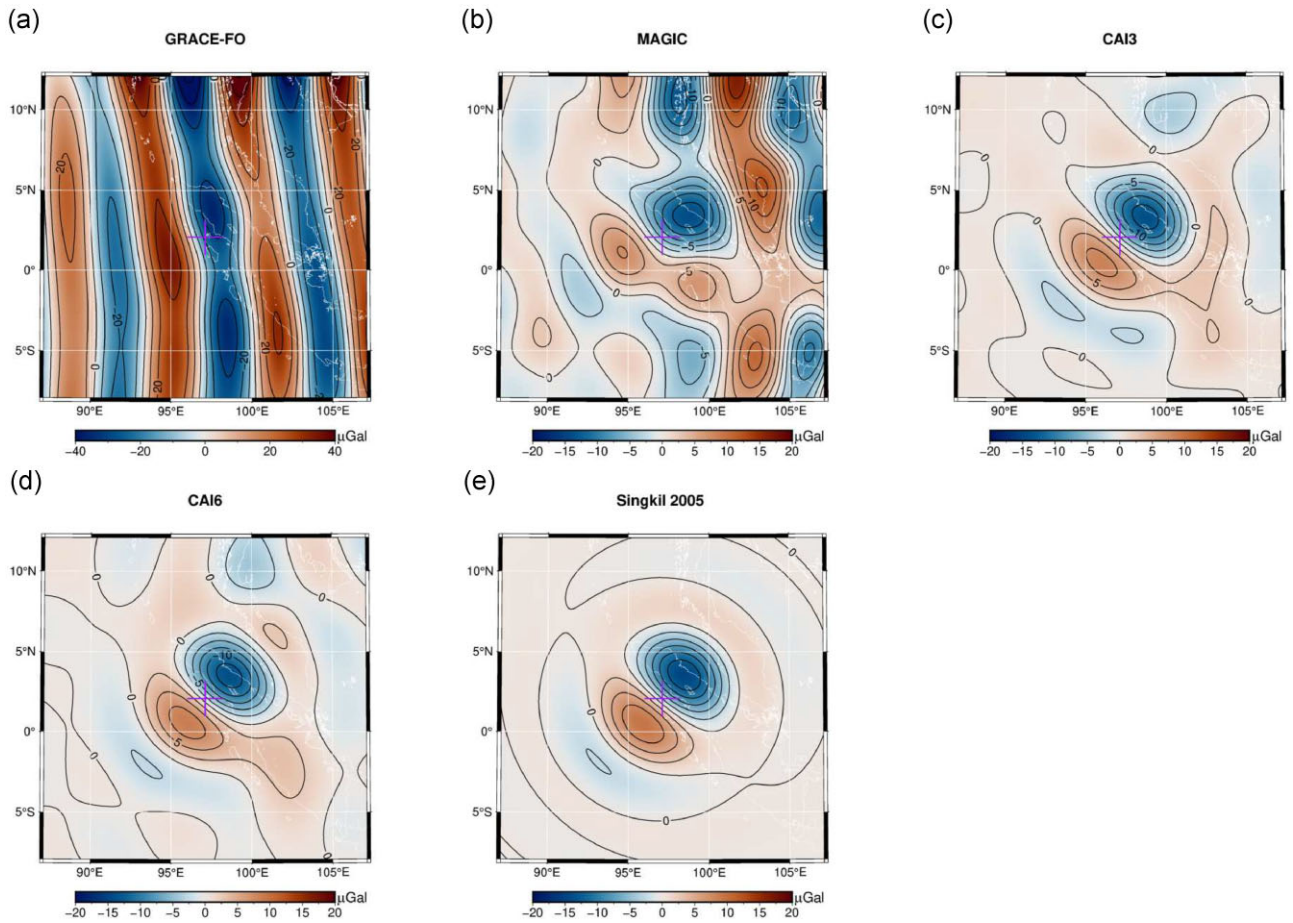

Figure 13. Sensitivity to a M 8.6 earthquake; (a) to (d): Observed coseismic signal for the four mission scenarios; (e) modelled coseismic gravity change for the Singkil 2005 earthquake.

offshore Mayotte in the Indian Ocean (Berthod *et al.* 2022). Here, the space-gravimetric method would offer a global means to monitor growth, through mass-change estimation, supposed the signal is sufficiently large to emerge above the noise level of the gravity observations (Braitenberg & Pastorutti 2024). Evidently, knowledge of the oceanic tidal and non-tidal variations is essential in separating the volcano signal. The size of a submarine volcano itself may be in the order of kilometres to a few tens of kilometres, thus beyond the smallest wavelength resolved in these satellite gravity data products. However, the gravity field that a seamount unrest generates contains sufficient energy at low harmonic degrees, allowing a detectability assessment (comparing the spectral content of signal and retrieval errors) in the measurement bandwidth of the satellites.

In this part, we compare the simulated truth earthquake signal, that is, as implemented in the ESAESM S-upd, with the simulated retrievals for the four mission scenarios. We illustrate maps of the true coseismic signal to be recovered, and of the retrieved signals. We then compare the earthquake and seamount signals with the retrieval error, defined by subtracting the true HIS signal from the retrieved signal in the spectral domain, with the help of the spectral degree amplitude of the retrieval error of the respective signals. Since the seamount signal is not included in the ESAESM, we examine its detectability by requiring that at least for a portion of the entire bandwidth, the signal spectrum amplitude is above the retrieval error spectral curve.

The 'true' coseismic gravity change for the Singkil 2005 earthquake is shown in Fig. 13(e), whereas the retrieved coseismic signal is seen in 13 (a to d). It is clearly seen that the retrieved signal (Figs 13a to d) is dominated by the typical striping pattern of the single-pair (GRACE-FO), greatly reduced in the MAGIC scenario and no more visible in the CAI3 and CAI6 scenarios. The striping pattern has amplitude of ± 35 microGals for the single-pair, rendering it a challenge to discern an earthquake signal of smaller size. It is possible, but either filtering must be applied, or an *a-priori* gravity pattern must be fitted to the data to allow one to extract the signal, which otherwise is difficult to discern. The striping is reduced to about ± 10 microGal for two pairs, drastically improving the detectability of an earthquake. For the detection of our earthquake case example, it appears that the coseismic signal would be well visible already with two satellite pairs, with a further improvement for the three-pair scenario, and another slight improvement for the six-pair scenario.

Concerning the spectral representation of signal and retrieval errors, we compare the signal spectrum in terms of spherical harmonic degree amplitudes with the one-week spectrum of the retrieval errors (Fig. 14j). We applied a localization procedure (Wieczorek & Simons 2005; 9° localization radius) to the spherical harmonic spectrum, in order to facilitate the comparison of the global noise degree variances with the spectrum of the local signal. The reason for this is that gravity signals due to an earthquake or a volcanic eruption have a very local extent, with the consequence that the

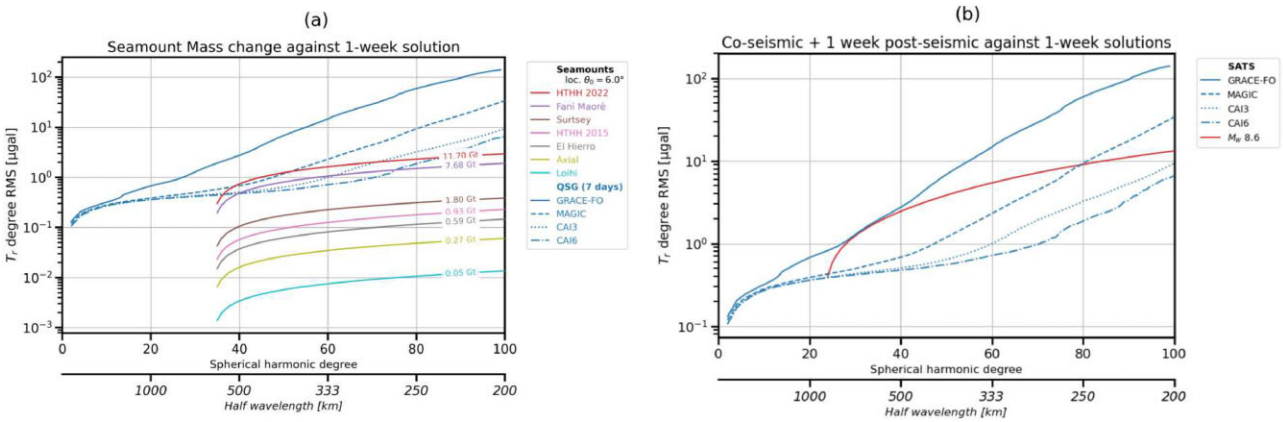

Figure 14. Signal and noise spectra in terms of degree amplitudes: (a) signal and noise spectra for a selection of seamount eruptions; (b) for the M 8.6 Singkil 2005 earthquake.

global spherical harmonic degree amplitude is not representative of the local signal amplitude. The localization algorithm is discussed in Wieczorek & Simons (2005) and its effect on the SH degree amplitudes for seamounts was illustrated in (Braitenberg & Pastorutti 2024). Results from our spectral analysis are consistent with the maps of the retrieved earthquake signal: considering one week of data acquisition, the noise of the GRACE-FO mission is larger than the signal and the signal is above the noise up to degrees 80 for the MAGIC mission, with the signal spectrum being above the noise curves up to degrees 100 for the CAI3 and CAI6 constellations. The earthquake gravity amplitude increases linearly with the seismic moment (of which the moment magnitude M_w is a logarithmic scale), but depends nonlinearly on the source depth and fault plane mechanism. While the latter have a smaller contribution to the signal with respect to magnitude (Daras *et al.* 2024a), assigning an absolute magnitude of detectability for any earthquake in a given mission scenario is challenging. In any case, our example shows that there is a marked increase in detectability from one to two pairs due to the successive reduction of the disturbing striping pattern. Advancing from three to six pairs gives only a minor reduction in the earthquake retrieval error, however the decrease in error at the shorter wavelengths (apparent in the spectra of Fig. 14) allows for a significant increase in spatial resolution. However, we note here that the rupture process can continue as a slow afterslip generating deformation, and inducing a viscoelastic relaxation response in the crust without emission of seismic waves. Geodetic measurements with GNSS, SAR capture this movement through observations of the surface displacements, but the movement cannot be seen if it occurs on the ocean bottom where the satellite gravity observations can detect a signal which otherwise remains elusive.

Switching to the second solid earth topic, in what follows we describe the scenario sensitivity with respect to submarine volcano unrests, choosing six seamounts distributed worldwide as example. Estimates for seamount volume and mass change are described in detail in Braitenberg & Pastorutti (2024), where they were based on documented morphological changes following an eruption. The mass change was then estimated assuming plausible density values, or relying on published values. These data served as input for the gravity field modelling, with which we obtained the target signals for the seamount detection exercise. The mass change is mostly positive, due to a magma eruption typically placing mass at the top of the seamount. Only for the Hunga Tonga 2022 explosion the mass change is negative, since the upper top of the caldera in the volcano summit remained at a lower level after the explosion. The estimated mass changes range from a minimum of 0.04 Gt to a maximum of 11.7 Gt. This mass change is modelled with a truncated cone, of which the spherical harmonic expansion is determined, and the localized spectrum is calculated, as done for the earthquake (6° localization radius). We refer to Braitenberg & Pastorutti (2024) for details of the mass modelling and the spectral localization operation. The signal spectral amplitude is then compared with the noise spectrum for the four satellite constellations, at weekly sampling. As seen in Fig. 14(a), only seamounts of the size of Hunga Tonga 2022 (11.7 Gt) or Fany Maoré (7.68 Gt) can be detected; with the smaller eruptions being too small and requiring a longer acquisition time to lower the noise curves. With MAGIC, a small part of the signal spectrum of Hunga Tonga 2022 could be detected. The detectability is greatly improved for the three- and six-pair constellations, allowing us to capture the signal for degrees up to 60 or above, for the two greatest eruptions Hunga Tonga and Fani Maoré. The useful spectral band is the one for which the signal spectrum is above the noise spectrum.

4 CONCLUSIONS

We find significant improvements for the advanced mission scenarios considered here over the GRACE-FO scenario, across a range of scientific applications. In other words, our results suggest that the MAGIC mission will provide data of an improved usability when compared to GRACE and GRACE-FO, and solutions from future quantum multipair solutions would again improve over MAGIC data products and likely generate new applications. This is of course not a surprise.

We find that these relative improvements will differ across applications, and they may be generally not as strong as a naive assessment on the basis of formal errors would let one expect. For example, errors in weekly basin-averaged total water storage will drop by a factor of about ten from the GRACE-FO scenario to the two-pair MAGIC scenario (with around 2 cm for 70 per cent of the 400 largest basins worldwide) when the same post-processing is applied, and then again by a factor of (only) two and slightly better to the six-pair CAI scenario. This is partly due to the presence of background model errors, which again is not a new result, and it depends inevitably on the chosen geopotential retrieval (level-1 to level-2) processing strategy and our assumptions on background model errors. Our results are broadly consistent with recent papers that assumed similar error models, for example, Daras *et al.*

(2024b). However we do not know of any study hitherto that had considered more complex applications such as involving data fusion or assimilation.

We showcase a number of potential new targets in ocean and sea level science, water resource management and geohazard monitoring and potentially forecasting that appear achievable only with a future multipair CAI mission. We strongly believe the societal and scientific benefits would warrant the realization of such missions, but we caution the reader here again that all our simulations are likely on the conservative side, and more in-depth investigations will be required.

We find stronger improvements for more 'straightforward' applications such as deriving time-series of basin-averaged TWSA or ocean mass, and this is also confirmed in sensitivity studies for geophysical signals. Extracting the simulated coseismic M 8.6 earthquake signal from the satellite observations is greatly improved for MAGIC when compared to GRACE-FO, with further significant improvement for CAI3 and minor improvement for CAI6. The spectral representation demonstrates that the signal is below the noise level of the weekly solutions for GRACE-FO, while being well above the noise level for MAGIC. CAI3 and CAI6 constellations would extent the smallest resolved wavelength significantly further as compared to MAGIC. For seamount unrest, all studied cases appear below the noise level of the weekly GRACE-FO solution, the largest unrest involving 11.7 Gt mass change being just above the threshold of MAGIC and well above noise level for the simulated CAI constellations.

We find improvements are less clear-cut in applications that involve the fusion of gravimetric data with other EO data, here along-track radar altimetry, or with hydrological and land surface models. For example, after assimilation of MAGIC (or CAI multi-pair) data into a 0.5° global hydrological model, we find that 20-yr trend errors at the native grid scale 50 km drop to about 1 cm yr^{-1}, about what we expect from GRACE-FO assimilation at 300 km scale. In other words, multipair TWSA maps, after assimilation, would resolve much smaller signals, of about a factor of 30–40 in area smaller, at the same noise level. For annual and semi-annual signals in total water storage, MAGIC assimilation solution errors at 100 km would be still lower than GRACE-FO assimilation errors at the 300 km spatial scale. However, we found no significant benefit from the simulated six-pair CAI data as compared to three-pair CAI, and only small from CAI3 compared to MAGIC—this is in contrast to the TWSA retrieval with our assimilation, and we attribute it here to the fact that we did not tune our assimilation specifically to each post GRACE-FO scenario. In combination with radar altimetry, we find a similar mixed picture with some sea level contributions already now very well determined, where big improvements with advanced missions in terms of ice sheet basin separability would not add much in terms of ocean mass budget knowledge, whereas the most uncertain component, that is, the barostatic ocean-internal mass redistribution, would improve only slightly with 10–20 per cent. We suggest this is due to the fact that these applications need to weigh mass change data against other satellite observations and/or model data, and the synthesized results (e.g. from data assimilation) will depend on both mass change data errors and model uncertainties as well as the strategies to deal with these errors, potentially in simplified ways, in scientific and potentially operational data analysis systems. This is in particular true when one is interested in improvements in variables that a mass change mission cannot observe at all without integrating with other data sets, such as ocean heat or groundwater change. We would like to add that formal data fusion and assimilation approaches force the user to explicitly specify uncertainties for all model and observational quantities. This is not the case for approaches that express unknown quantities (e.g. ocean heat change or groundwater storage change) directly as a residual; however since these approaches work essentially with the same data as more formal approaches they tend to underestimate error contamination.

Some obvious limitations of this study are related to the realism of the simulated scenarios, the used sensor error and noise models, the background model error assumptions, other assumptions that we had to make about the processing, the number of simulations that we could carry out within computational resources and finally the resulting limitations in representing errors and error correlations on the basis of a limited number of samples. In the assessment of the potential of mass change data in new applications, further limitations related to these analysis frameworks must be admitted; for example, when integrating MAGIC or quantum mission data with radar altimetry it is challenging to represent the entire complexity in mission simulation assessments. The same is also true in data assimilation frameworks. Our study here was still limited in terms of integrating spatial observation correlations in data assimilation (a limitation that many current real assimilation systems face). It should also be noted that we did not account for potential improvements in radar altimetry coverage, improvements in corrections and noise levels, and we did not attempt to consider future lower uncertainties of hydrological models resulting, for example, from improved forcing data.

However, we would like to clarify again that we intentionally did not tune our assessments towards the specific mission scenario simulations, for example, by choosing individually optimized (DDK) filters. Our focus here was on assuming as a baseline the post-processing that was developed for GRACE/-FO, for consistency in the relative comparison. The simulated GRACE-FO errors in our study are, even after applying DDK filters, somewhat on the conservative side when compared to what many real-data analysis suggest and that may lead one to speculate that all MAGIC and quantum multipair results simulated here may also be on the conservative side.

We believe this paper covers a few very relevant applications that are often cited for showcasing the benefit of continued and enhanced mass change observation, with additional work required to improve the representativeness across the full spectrum. It seems, however, evident that with the NGGM mission in addition to the upcoming GRACE-C continuation of GRACE-FO, that is, with the MAGIC constellation, all science applications will benefit and new applications will likely become possible. Dramatically improved latency times and higher temporal resolution will likely enable operational services, something which follow-on studies could explore in more detail. What is additionally required to unlock this potential are background model improvements, gravity retrieval method improvement, tailoring of data fusion and assimilation frameworks and ways for representing and implementing error models better in post-processing schemes. We note that the questions of 'how accurate are the GRACE/-FO data' and 'how correlated are TWSA maps' had hampered many applications and we urge that attention is payed to these questions for MAGIC. As long as assessment studies are limited to a quite finite number of simulations, we suggest that also improved approaches for estimating error covariance matrices need to be studied.

ACKNOWLEDGMENTS

Most work presented here was performed in the framework of the project 'Quantum Space Gravimetry for monitoring Earth's Mass Transport Processes—QSG4EMT', Contract No. 4000138395/22/NL/SD funded by the European Space Agency. In addition, JK, AS and YE acknowledge funding from Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for SFB 1502/1-2022 (project no 450058266). BU acknowledges funding from DFG within FOR AlgoforGe, project no. 459420781. We acknowledge Alejandro Blasquez (LEGOS, Toulouse, France) and Srinivas Bettadpur (U Texas, Austin, USA) for discussing an early version of the manuscript. We are further grateful to Joao de Teixeira da Encarnacao (TU Delft, The Netherlands) for helpful discussions on sensor models.

AUTHOR CONTRIBUTION

JK, RP and ID led the study, JK, AE, CB and MR conceived the analysis, HG, AS, BU, YE, AE, CB, AP, PZ, MS, MR, LR, FM conducted the experiments and analyses, RP, PZ and MS created and provided the mission simulations. All authors contributed to the text and reviewed the manuscript.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY

Data generated in this study can be shared upon reasonable request to the corresponding author.

REFERENCES

A, G., Wahr, J. & Zhong, S., 2013. Computations of the viscoelastic response of a 3-D compressible Earth to surface loading: an application to Glacial Isostatic Adjustment in Antarctica and Canada, *J. Geophys. Int.,* **192**(2), 557–572.

Abrykosov, P., Murböck, M., Hauk, M., Pail, R. & Flechtner, F., 2022. Data-driven multi-step self-de-aliasing approach for GRACE and GRACE-FO data processing, *J. geophys. Int.,* **232**(2), 1006–1030.

Alonso, I.E., 2022. Cold atoms in space: community workshop summary and proposed road-map, *EPJ Quantum Technol.,* **9,** 30. doi:10.1140/epjqt/s40507-022-00147-w.

Bender, P.L., 2022. An improved next generation gravity mission, *Remote Sensing,* **14**(4), doi:10.3390/rs14040948.

Berthod, C. *et al.*, 2022. Temporal magmatic evolution of the Fani Maoré submarine eruption 50 km east of Mayotte revealed by in situ sampling and petrological monitoring, *C. R. Géosci.,* **354**(S2), 1–29.

Bollmeyer, C. *et al.*, 2015. Towards a high-resolution regional reanalysis for the European CORDEX domain, *Quart. J. R. Meteorol. Soc.,* **141**(686), 1–15.

Bouih, M., Panet, I., Remy, D., Longuevergne, L. & Bonvalot, S., 2022. Deep mass redistribution prior to the 2010 mw 8.8 Maule (Chile) earthquake revealed by GRACE satellite gravity, **584,** 117465. doi:10.1016/j.epsl.2022.117465.

Braitenberg, C. & Pastorutti, A., 2024. Detectability of seamount eruptions through a quantum technology gravity mission MOCAST+: Hunga Tonga, Fani Maoré and other smaller eruptions, *Surv. Geophys.,* **45**(4), 1331–1361.

Cambiotti, G., Douch, K., Cesare, S., Haagmans, R., Sneeuw, N., Anselmi, A., Marotta, A.M. & Sabadini, R., 2020. On earthquake detectability by the next-generation gravity mission, *Surv. Geophys.,* **41**(5), 1049–1074.

Carraz, O., Siemes, C., Massotti, L., Haagmans, R. & Silvestrin, P., 2014. A spaceborne gravity gradiometer concept based on cold atom interferometers for measuring earth's gravity field, *Microgravity Sci. Technol.,* **26,** 139–145.

Chambers, D.P., Wahr, J., Tamisiea, M.E. & Nerem, R.S., 2010. Ocean mass from GRACE and glacial isostatic adjustment, *J. geophys. Res.: Solid Earth,* **115**(B11), B11415.

Chao, B.F. & Liau, J.R., 2019. Gravity changes due to large earthquakes detected in GRACE satellite data via empirical orthogonal function analysis, **124**(3), 3024–3035.

Colombo, O.L., 1981. *Numerical methods for harmonic analysis on the sphere, Vol. 310, Department of Geodetic Science,* The Ohio State University.

Daras, I., 2023. Next Generation Gravity Mission (NGGM) Mission Requirements Document, Issue 1.0, Tech. rep., *Earth and Mission Science Division, European Space Agency*.

Daras, I. *et al.*, 2024a. Mass-change and geosciences international constellation (magic) expected impact on science and applications, *J. geophys. Int.,* **236**(3), 1288–1308.

Daras, I. *et al.*, 2024b. Mass-change And Geosciences International Constellation (MAGIC) expected impact on science and applications, *J. geophys. Int.,* **236**(3), 1288–1308.

Dieng, H.B., Cazenave, A., Meyssignac, B. & Ablain, M., 2017. New estimate of the current rate of sea level rise from a sea level budget approach, *Geophys. Res. Lett.,* **44**(8), 3744–3751.

Dobslaw, H., Bergmann-Wolf, I., Dill, R., Forootan, R., Klemann, V., Kusche, J. & Sasgen, I., 2015. The updated ESA earth system model for future gravity mission simulation studies, *J. Geod.,* **89,** 505–513.

Dobslaw, H., Bergmann-Wolf, I., Forootan, E., Dahle, C., Mayer-Gürr, T., Kusche, J. & Flechtner, F., 2016. Modeling of present-day atmosphere and ocean non-tidal de-aliasing errors for future gravity mission simulations, *J. Geod.,* **90**(5), 423–436.

Douch, K., Wu, H., Schubert, C., Müller, J. & Pereira dos Santos, F., 2018. Simulation-based evaluation of a cold atom interferometry gradiometer concept for gravity field recovery, *Adv. Space Res.,* **61**(5), 1307–1323.

Dueben, P.D. & Bauer, P., 2018. Challenges and design choices for global weather and climate models based on machine learning, *Geosci. Model Develop.,* **11**(10), 3999–4009.

Eicker, A., Schumacher, M., Kusche, J., Döll, P. & Müller Schmied, H., 2014. Calibration/data assimilation approach for integrating GRACE data into the WaterGAP Global Hydrology Model (WGHM) using an ensemble Kalman filter: First results, *Surv. Geophys.,* **35**(6), 1285–1309.

Eicker, A., Strohmenger, C., Braitenberg, C., Kusche, J., Pail, R. & Daras, I., 2024. *Community assessment on user requirements for future satellite gravity missions* EGU24-14722, Copernicus Meetings, Conference Name: EGU24.

Elsaka, B. *et al.*, 2014. Comparing seven candidate mission configurations for temporal gravity field retrieval through full-scale numerical simulation, *J. Geod.,* **88**(1), 31–43.

Encarnação, J., Siemes, C., Daras, I., Carraz, O., Strangfeld, A., Zingerle, P. & Pail, R., 2024. *Towards a realistic noise modelling of quantum sensors for future satellite gravity missions*.

European Commission, 2003. *Regulation (EC) No 1059/2003 of the European Parliament and of the Council of 26 May 2003 on the establishment of a common classification of territorial units for statistics (NUTS)*.

Felsberg, A., Lannoy, G. J. M.D., Girotto, M., Poesen, J., Reichle, R.H. & Stanley, T., 2021. Global soil water estimates as landslide predictor: the effectiveness of SMOS, SMAP, and GRACE observations, land surface simulations, and data assimilation, *J. Hydrometeorol.,* **22**(5), 1065–1084.

Fisher, R.A. & Koven, C.D., 2020. Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems, *J. Adv. Model. Earth Syst.,* **12**(4), e2018MS001453. doi:10.1029/2018MS001453.

Flechtner, F., Dobslaw, H. & Fagiolini, E., 2015. *GRACE 327-750 - AOD1B product description document for product release 05,* Tech. Rep. Rev. 4.3, GFZ German Research Centre for Geosciences.

Flechtner, F., Neumayer, K.-H., Dahle, C., Dobslaw, H., Fagiolini, E., Raimondo, J.-C. & Güntner, A., 2016. What can be expected from the GRACE-FO laser ranging interferometer for earth science applications?, *Surv. Geophys.,* **37**(2), 453–470.

Flechtner, F., Landerer, F., Save, H., Mccullough, C., Dahle, C., Bettadpur, S., Gaston, R. & Snopek, K., 2024a. *GRACE-FO: science results, project status and further plans,* EGU24-7426, Copernicus Meetings, Conference Name: EGU24.

Flechtner, F., Wiese, D., Webb, F., Landerer, F., Gross, M., Snopek, K., Fischer, S. & Dahle, C., 2024b. Global gravity and mass change observations beyond GRACE-FO: updates on the upcoming GRACE-Continuity mission, in *GRACE/GRACE-FO Science Team Meeting,* Potsdam, Germany, pp.8–10. Oct 2024, GSTM2024-21.

Forman, B.A. & Reichle, R.H., 2013. The spatial scale of model errors and assimilated retrievals in a terrestrial water storage assimilation system, *Water Resour. Res.,* **49**(11), 7457–7468.

Frederikse, T., Riva, R., Kleinherenbrink, M., Wada, Y., van den Broeke, M. & Marzeion, B., 2016. Closing the sea level budget on a regional scale: trends and variability on the Northwestern European continental shelf, *Geophys. Res. Lett.,* **43**(20), 10 864–10 872.

Gerdener, H., Kusche, J., Schulze, K., Ghazaryan, G. & Dubovyk, O., 2022. Revising precipitation–water storages–vegetation signatures with GRACE-based data assimilation, *J. Hydrol.,* **612:A,** 128096. doi:10.1016/j.jhydrol.2022.128096.

Gerdener, H., Kusche, J., Schulze, K., Döll, P. & Klos, A., 2023. The global land water storage data set release 2 (GLWS2.0) derived via assimilating GRACE and GRACE-FO data into a global hydrological model, *J. Geod.,* **97,** 73. doi:10.1007/s00190-023-01763-9.

Haagmans, R., Siemes, C., Massotti, L., Carraz, O. & Silvestrin, P., 2020. ESA's next-generation gravity mission concepts, *Rendiconti Lincei. Scienze Fisiche e Naturali,* **31**(Suppl. 1), S15–S25.

Hakuba, M.Z., Frederikse, T. & Landerer, F.W., 2021. Earth's Energy Imbalance From the Ocean Perspective (2005–2019), *Geophys. Res. Lett.,* **48**(16), e2021GL093624. doi:10.1029/2021GL093624.

Han, S.-C., Sauber, J. & Pollitz, F., 2015. Coseismic compression/dilatation and viscoelastic uplift/subsidence following the 2012 indian ocean earthquakes quantified from satellite gravity observations, *Geophys. Res. Lett.,* **42**(10), 3764–3772.

Han, S.-C., Sauber, J. & Pollitz, F., 2016. Postseismic gravity change after the 2006–2007 great earthquake doublet and constraints on the asthenosphere structure in the central Kuril Islands, *Geophys. Res. Lett.,* **43**(7), 3169–3177.

Han, S.-C., Sauber, J., Broerse, T., Pollitz, F., Okal, E., Jeon, T., Seo, K.-W. & Stanaway, R., 2024. GRACE and GRACE follow-on gravity observations of intermediate-depth earthquakes contrasted with those of shallow events, *J. geophys. Res.: Solid Earth,* **129**(2), e2023JB028362. doi:10.1029/2023JB028362.

Horwath, M. *et al.*, 2022. Global sea-level budget and ocean-mass budget, with a focus on advanced data products and uncertainty characterisation, *Earth Syst. Sci. Data,* **14**(2), 411–447.

Jensen, L., Eicker, A., Dobslaw, H., Stacke, T. & Humphrey, V., 2019. Long-Term Wetting and Drying Trends in Land Water Storage Derived From GRACE and CMIP5 Models, *J. geophys. Res.: Atmospheres,* **124**(17–18), 9808–9823.

Jensen, L., Gerdener, H., Eicker, A., Kusche, J. & Fiedler, S., 2024. Observations indicate regionally misleading wetting and drying trends in CMIP6, *npj Clim. Atmos. Sci.,* **7**(1), 1–12.

Kalnay, E., 2002. *Atmospheric Modeling, Data Assimilation and Predictability,* Cambridge Univ. Press.

Kirschbaum, D.B., Adler, R., Hong, Y., Hill, S. & Lerner-Lam, A., 2010. A global landslide catalog for hazard applications: method, results, and limitations, *Natural Hazards,* **52**(3), 561–575.

Kurtz, W., He, G., Kollet, S.J., Maxwell, R.M., Vereecken, H. & Hendricks Franssen, H.-J., 2016. Terrsysmp–pdaf (version 1.0): a modular high-performance data assimilation framework for an integrated land surface–subsurface model, *Geosci. Model Develop.,* **9**(4), 1341–1360.

Kusche, J., 2007. Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models, *J. Geod.,* **81**(11), 733–749.

Landerer, F.W. *et al.*, 2020. Extending the global mass change data record: GRACE Follow-On instrument and science data performance, *Geophys. Res. Lett.,* **47**(12), e2020GL088306. doi:10.1029/2020GL088306.

Lawrence, D.M. *et al.*, 2019. The community land model version 5: description of new features, benchmarking, and impact of forcing uncertainty, *J. Adv. Model. Earth Syst.,* **11**(12), 4245–4287.

Lévèque, T. *et al.*, 2019. Correlated atom accelerometers for mapping the Earth gravity field from Space, in *International Conference on Space Optics—ICSO 2018,* Vol. **11180,** p. 111800W, eds Sodnik, Z., Karafolas, N. & Cugny, B., International Society for Optics and Photonics, SPIE.

Lévèque, T. *et al.*, 2023. CARIOQA: definition of a Quantum Pathfinder Mission, in *International Conference on Space Optics—ICSO 2022,* Vol. **12777,** p. 127773L, eds Minoglou, K., Karafolas, N. & Cugny, B., International Society for Optics and Photonics, SPIE.

Li, B. *et al.*, 2019. Global GRACE data assimilation for groundwater and drought monitoring: advances and challenges, *Water Resour. Res.,* **55**(9), 7564–7586.

Liu, Y., Wang, P., Ruan, H., Wang, T., Yu, J., Cheng, Y. & Kulmatov, R., 2020. Sustainable use of groundwater resources in the transboundary aquifers of the Five Central Asian Countries: challenges and perspectives, *Water,* **12**(8), 2101. doi:10.3390/w12082101.

Marti, F. *et al.*, 2022. Monitoring the ocean heat content change and the Earth energy imbalance from space altimetry and space gravimetry, *Earth Syst. Sci. Data,* **14**(1), 229–249.

Massotti, L., Siemes, C., March, G., Haagmans, R. & Silvestrin, P., 2021. Next generation gravity mission elements of the mass change and geoscience international constellation: from orbit selection to instrument and mission design, *Remote Sensing,* **13**(19). doi:10.3390/rs13193935.

Meza, I. *et al.*, 2020. Global-scale drought risk assessment for agricultural systems, **20**(2), 695–712.

Migliaccio, F., Reguzzoni, M., Sansò, F. & Tselfes, N., 2006. On the use of gridded data to estimate potential coefficients, in *Proceedings 3rd GOCE user workshop, Frascati, ESRIN,* pp. 311–318.

Migliaccio, F., Reguzzoni, M., Sansö, F. & Tselfes, N., 2009. An error model for the goce space-wise solution by Monte Carlo methods, in *Observing our changing earth,* pp. 337–344, ed. Sideris, M. G., Springer.

Migliaccio, F. *et al.*, 2019. MOCASS: A satellite mission concept using cold atom interferometry for measuring the earth gravity field, *Surv. Geophys.,* **40,** 1029–1053.

Miyoshi, T., Kalnay, E. & Li, H., 2013. Estimating and including observation-error correlations in data assimilation, **21**(3), 387–398 . doi:10.1080/17415977.2012.712527.

Müller Schmied, H. *et al.*, 2021. The global water resources and use model WaterGAP v2.2d: model description and evaluation, *Geosci. Model Develop,* **14**(2), 1037–1079.

Monahan, T., Tang, T., Roberts, S. & Adcock, T.A., 2024. *Tidal and mean sea surface corrections from and for SWOT using a spatially coherent variational Bayesian harmonic analysis*.

Mouyen, M., Longuevergne, L., Steer, P., Crave, A., Lemoine, J.-M., Save, H. & Robin, C., 2018. Assessing modern river sediment discharge to the ocean using satellite gravimetry, *Nat. Commun.,* **9**(1), 1–9.

Murböck, M., Pail, R., Daras, I. & Gruber, T., 2014. Optimal orbits for temporal gravity recovery regarding temporal aliasing, *J. Geod.,* **88**(2), 113–126.

Naz, B.S. *et al.*, 2019. Improving soil moisture and runoff simulations at 3 km over Europe using land surface data assimilation, *Hydrol. Earth Syst. Sci.,* **23**(1), 277–301.

Nerger, L., Tang, Q. & Mu, L., 2020. Efficient ensemble data assimilation for coupled models with the Parallel Data Assimilation Framework: example of AWI-CM (AWI-CM-PDAF 1.0), *Geosci. Model Develop.,* **13**(9), 4305–4321.

Ojha, C., Werth, S. & Shirzaei, M., 2019. Groundwater Loss and Aquifer System Compaction in San Joaquin Valley During 2012–2015 Drought, *J. geophys. Res.: Solid Earth,* **124**(3), 3127–3143.

Oleson, K. *et al.*, 2004. *Technical Description of the Community Land Model (CLM)*.

Pail, R. *et al.*, 2015. Science and user needs for observing global mass transport to understand global change and to benefit society, *Surv. Geophys.*, **36**(6), 743–772.

Pail, R. *et al.*, 2024. Report of MAGIC Science and Applications Workshop 2023, *European Space Agency and National Aeronautics and Space Administration,* eds Daras, I. & Tsaoussi, L.,

Purkhauser, A.F. & Pail, R., 2020. Triple-pair constellation configurations for temporal gravity field retrieval, *Remote Sensing,* **12**(5). doi:10.3390/rs12050831.

Ray, R.D., 2008. A preliminary tidal analysis of icesat laser altimetry: Southern ross ice shelf, *Geophys. Res. Lett.,* **35**(2). doi:10.1029/2007GL032125.

Reager, J.T., Thomas, B.F. & Famiglietti, J.S., 2014. River basin flood potential inferred using GRACE gravity observations at several months lead time, *Nature Geoscience,* **7**(8), 588–592.

Reguzzoni, M. & Tselfes, N., 2009. Optimal multi-step collocation: application to the space-wise approach for goce data analysis, *J. Geod.,* **83,** 13–29.

Reguzzoni, M., Gatti, A., De Gaetani, C., Migliaccio, F. & Sansò, F., 2014. Locally adapted space-wise grids from GOCE data, in *EGU General Assembly Conference Abstracts,* pp. 14010.

Reguzzoni, M., Migliaccio, F. & Batsukh, K., 2021. Gravity field recovery and error analysis for the MOCASS mission proposal based on cold atom interferometry, *Pure appl. Geophys.,* **178,** 2201–2222.

Rietbroek, R., Brunnabend, S.-E., Kusche, J., Schröter, J. & Dahle, C., 2016. Revisiting the contemporary sea-level budget on global and regional scales, *Proc. Natl Acad. Sci.,* **113**(6), 1504–1509.

Rodell, M. & Reager, J., 2023. Water cycle science enabled by the GRACE and GRACE-FO satellite missions, *Nat. Water,* **1,** 47–59.

Rossi, L., Reguzzoni, M., Koç, Ö., Rosi, G. & Migliaccio, F., 2023. Assessment of gravity field recovery from a quantum satellite mission with atomic clocks and cold atom gradiometers, *Quantum Sci. Technol.,* **8,** 014009. doi:10.1088/2058-9565/aca8cc.

Savcenko, R., Bosch, W., Dettmering, D. & Seitz, F., 2012. EOT11a - global empirical ocean tide model from multi-mission satellite altimetry, with links to model results, in supplement to: Savcenko, Roman; Bosch, Wolfgang (2012): Eot11a - empirical ocean tide model from multi-mission satellite altimetry. Deutsches Geodätisches Forschungsinstitut (DGFI), München, p. 49.

Schlaak, M., Pail, R., Jensen, L. & Eicker, A., 2023. Closed loop simulations on recoverability of climate trends in next generation gravity missions, *J. geophys. Int.,* **232**(2), 1083–1098.

Schmidt, C., Krauth, T. & Wagner, S., 2017. *Export of Plastic Debris by Rivers into the Sea,* American Chemical Society. doi:10.1021/acs.est.7b02368.

Shihora, L., Liu, Z., Balidakis, K., Wilms, J., Dahle, C., Flechtner, F., Dill, R. & Dobslaw, H., 2024. Accounting for residual errors in atmosphere–ocean background models applied in satellite gravimetry, *J. Geod.,* **98**(27). doi:10.1007/s00190-024-01832-7.

Shrestha, P., Sulis, M., Masbou, M., Kollet, S. & Simmer, C., 2014. A scale-consistent terrestrial systems modeling platform based on cosmo, clm, and parflow, *Monthly Weath. Rev.,* **142**(9), 3466–3483.

Springer, A., 2019. *A water storage reanalysis over the European continent: assimilation of GRACE data into a high-resolution hydrological model and validation, Thesis,* Universitäts- und Landesbibliothek Bonn.

Springer, A., Kusche, J., Hartung, K., Ohlwein, C. & Longuevergne, L., 2014. New estimates of variations in water flux and storage over Europe based on regional (re)analyses and multisensor observations, *J. Hydrometeorol.,* **15**(6), 2397–2417.

Springer, A., Eicker, A., Bettge, A., Kusche, J. & Hense, A., 2017. Evaluation of the water cycle in the European COSMO-REA6 reanalysis using GRACE, *Water,* **9**(4). doi:10.3390/w9040289.

Springer, A., Karegar, M.A., Kusche, J., Keune, J., Kurtz, W. & Kollet, S., 2019. Evidence of daily hydrological loading in GPS time series over Europe, *J. Geod.,* **93**(10), 2145–2153.

Springer, A., Mielke, C., Liu, Z., Dixit, S., Friederichs, P. & Kusche, J., 2024. A regionally refined and mass-consistent atmospheric and hydrological de-aliasing product for GRACE, GRACE-FO and future gravity mission, *J. geophys. Res. (Solid Earth),* **129,** e2023JB027883. doi:10.1029/2023JB027883.

Stanley, T.A., Kirschbaum, D.B., Benz, G., Emberson, R.A., Amatya, P.M., Medwedeff, W. & Clark, M.K., 2021. Data-driven landslide nowcasting at the global scale, *Front. Earth Sci.,* **9,** 640043. doi:10.3389/feart.2021.640043.

Swenson, S.C. & Lawrence, D.M., 2014. Assessing a dry surface layer-based soil resistance parameterization for the Community Land Model using GRACE and FLUXNET-MTE data, *J. geophys. Res.: Atmospheres,* **119**(17), 10299–10312.

Tanaka, Y. & Heki, K., 2014. Long- and short-term postseismic gravity changes of megathrust earthquakes from satellite gravimetry, **41**(15), 5451–5456.

Tapley, B.D. *et al.*, 2019. Contributions of GRACE to understanding climate change, *Nat. Clim. Change,* **9**(5), 358–369.

Trimeche, A. *et al.*, 2019. Concept study and preliminary design of a cold atom interferometer for space gravity gradiometry, *Class. Quantum Gravity,* **36**(21), 215004. doi:10.1088/1361-6382/ab4548.

Uebbing, B., Kusche, J., Rietbroek, R. & Landerer, F., 2019. Processing choices affect ocean mass estimates from GRACE, *J. geophys. Res.: Oceans,* **124**(Supplementary). doi:10.1029/2018JC014341.

van der Gun, J., 2022. *Large Aquifer Systems Around the World | The Groundwater Project,* https://gw-project.org/ (accessed 1 March 2025).

Wahr, J., Molenaar, M. & Bryan, F., 1998. Time variability of the Earth's gravity field: hydrological and oceanic effects and their possible detection using GRACE, *J. geophys. Res.: Solid Earth,* **103**(B12), 30 205–30 229.

Wang, J., Forman, B.A., Girotto, M. & Reichle, R.H., 2021. Estimating terrestrial snow mass via multi-sensor assimilation of synthetic AMSR-E brightness temperature spectral differences and synthetic GRACE terrestrial water storage retrievals, *Water Resour. Res.,* **57**(9), e2021WR029880. doi:10.1029/2021WR029880.

Wang, R., Heimann, S., Zhang, Y., Wang, H. & Dahm, T., 2017. Complete synthetic seismograms based on a spherical self-gravitating earth model with an atmosphere–ocean–mantle–core structure, *J. geophys. Int.,* **210**(3), 1739–1764.

WCRP Global Sea Level Budget Group, 2018. Global sea-level budget 1993–present, *Earth Syst. Sci. Data,* **10**(3), 1551–1590.

Wessel, P., Sandwell, D.T. & Kim, S.-S., 2010. The global seamount census, *Oceanography,* **23**(1), 24–33.

Wieczorek, M.A. & Simons, F.J., 2005. Localized spectral analysis on the sphere, *J. geophys. Int.,* **162**(3), 655–675.

Wiese, D., Visser, P. & Nerem, S., 2011. Estimating low resolution gravity fields at short time intervals to reduce temporal aliasing errors, *Adv. Space Res.,* **48**(6), 1094–1107.

Wiese, D.N. *et al.*, 2022. The mass change designated observable study: overview and results, *Earth Space Sci.,* **9**(8), e2022EA002311. doi:10.1029/2022EA002311.

Yan, Z., Ran, J., Xiao, Y., Xu, Z., Wu, H., Deng, X.-L., Du, L. & Zhong, M., 2023. The temporal improvement of earth's mass transport estimated by coupling GRACE-FO with a Chinese polar gravity satellite mission, *J. geophys. Res.: Solid Earth,* **128**(9), e2023JB027157. doi:10.1029/2023JB027157.

Yu, D., Weng, Z., Hwang, C., Zhu, H., Luo, J., Yuan, J. & Ge, S., 2024. Seamount detection using SWOT-derived vertical gravity gradient: advancements and challenges, *J. geophys. Int.,* **237**(3), 1780–1793.

Zaitchik, B.F., Rodell, M. & Reichle, R.H., 2008. Assimilation of GRACE terrestrial water storage data into a land surface model: results for the Mississippi River Basin, *J. Hydrometeorol.,* **9**(3), 535–548.

Zingerle, P., Pail, R., Gruber, T. & Daras, I., 2024. Constellation design and performance of future quantum satellite gravity missions, *Earth Planets Space,* **76**(101). doi:10.1186/s40623-024-02034-3.