Evaluating non-tidal atmospheric products by measuring GRACE K-band range rate residuals

Fan Yang[a,b], Ehsan Forootan[c,d], Maike Schumacher[d,e], C.K. Shum[b,f], Min Zhong[b]

[a] *State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China*
[b] *State Key Laboratory of Geodesy and Earths Geodynamics, Chinese Academy of Sciences, Wuhan, China*
[c] *School of Earth and Ocean Sciences, Cardiff University, Cardiff, UK*
[d] *Institute of Physics and Meteorology (IPM), University of Hohenheim, Stuttgart, Germany*
[e] *School of Geographical Sciences, University of Bristol, Bristol, UK*
[f] *Division of Geodetic Science, School of Earth Sciences, Ohio State University, Columbus, Ohio, USA*

Abstract

In order to reduce high frequency non-tidal mass changes, while inverting for the Earth's time-variable gravity fields from the Gravity Recovery And Climate Experiment (GRACE) measurements, it is usual to apply the Atmospheric and Oceanic De-aliasing (AOD1B) products. However, limitations in these products count as a potential threat to the accuracy of time-variable gravity fields derived from GRACE, as well as its follow-on mission(s). Therefore, in this study, we show to what extent the GRACE-type gravity recovery procedure is sensitive to different non-tidal atmospheric background models. For this, we evaluate the atmospheric parts of the GeoForschungsZentrum (GFZ)'s AOD1B RL05 and RL06, as well as those computed as a part of the European Space Agency Earth System Model ESA-ESM, and the ITG3D model. These data products employ different atmosphere fields (operational and reanalysis data or their combination) from the European Centre for Medium-Range Weather Forecasts (ECMWF) as inputs, and they are also computed by implementing different 2-dimensional or 3-dimensional (2-D or 3-D) integration methods. The accuracy of these products is assessed by comparing the resulting GRACE K-Band Range-Rate (KBRR) residuals computed for time-variable gravity field inversions using each of them separately as a background model. Our investigations during 2006 indicate that: (i) applying ESA-ESM and ITG3D decreases averaged KBRR residuals by 2.8 nm/s and 3.4 nm/s compared to those reduced by the official RL05 products. (ii) Projecting these residuals onto the spatial domain indicates that the improvement covers 78.4% and 78.9% of the globe, respectively. (iii) We find that, compared to ESA-ESM, ITG3D can further reduce the KBRR residuals by 1.8 nm/s at regions of high latitudes, which likely improve the uncertainty of ice mass estimations. Our investigation of the AOD1B RL06 products covers 2006-2010, which indicates the advantage of using the higher temporal sampling, i.e. 3-hourly reanalysis data. Applying the RL06 reduces the averaged KBRR residuals by 44.2 nm/s with respect to the use of the RL05 for gravity field inversion. We, therefore, conclude that the integration method of ITG3D and utilizing reanalysis data with higher (than 6-hourly) temporal sampling rate are beneficial for GRACE-like gravity inversion such as the GRACE Follow-On mission with laser interferometric ranging system.

Keywords: Atmosphere De-aliasing, GRACE, Time-variable Gravity fields, KBRR Residuals

---

1. Introduction

Over the past decade (2002-2017), the Gravity Recovery And Climate Experiment (GRACE) twin-satellite mission (*Tapley et al.*, 2004), has accumulated numerous observations that allow mapping of time-variable gravity of the Earth. From these observations, monthly global gravity field products, which are publicly known as the GRACE level 2 (L2) products (see, e.g., *Dahle et al.*, 2014) are widely used to broaden our knowledge in interdisciplinary science including studying water variability in soil and sub-surface aquifers (*Ramillien et al.*, 2011; *Famiglietti and Rodell*, 2013; *Schumacher et al.*, 2016, 2018; *Forootan et al.*, 2017), and continental ice-sheets (*Sasgen et al.*, 2013) at scales of a few hundred kilometers (see other examples in, e.g., *Kusche et al.*, 2012).

In order to accurately estimate terrestrial water storage changes from GRACE observations, the effects of mass redistributions in the atmosphere and the oceans in response to high-frequency time-variable signals have to be removed or diminished while inverting level 1b (L1b) raw data to solve for gravity fields (for example the commonly used L2 products). To this end, a number of tidal as well as non-tidal background models are forward modeled to reduce these L1b data (bias corrected range rate and position observations). From these, the Atmosphere and Ocean De-Aliasing Product (AOD1B) is released by the official GRACE data processing center GeoForschungsZentrum (GFZ), Potsdam, to account for the non-tidal high frequency mass variations as accurately as possible.

It is well known to the GRACE science team that the anticipated baseline accuracy of GRACE (e.g., *Kim*, 2000) has not been fulfilled, yet, for which the temporal aliasing of high-frequency mass variations is assumed to be a dominant error source (*Elsaka et al.*, 2014; *Sakumura et al.*, 2014). Moreover, *Loomis et al.* (2012) pointed out that even the upcoming GRACE follow-on (GRACE-FO) with laser interferometric ranging system will not help to obtain a better temporal gravity field due to the temporal aliasing problem. In this context, enhancement of the current de-aliasing products (see, *Fagiolini et al.*, 2015) or developing new algorithms to overcome de-aliasing problem (see, *Daras and Pail*, 2017) is necessary for future missions. Some known artifacts in official atmosphere de-aliasing (AD) product, e.g., the data jump pointed by *Duan et al.* (2012); *Forootan et al.* (2014), model drift by *Hardy et al.* (2017), and the imperfect physical assumption by *Forootan et al.* (2013), suggest that the estimation of these products needs to be improved.

Therefore, in addition to the official atmosphere de-aliasing products (abbreviated to 'ATM') provided by GFZ, alternative products have also been released by, for example, *Boy and Chao* (2005), *Zenner et al.* (2010), as well as the [ITG3D model by *Forootan et al.* (2013), and the European Space Agency (ESA) Earth System Model (abbreviated to ESM, see details in *Dobslaw et al.*, 2015, 2016), of which ITG3D and ESM models are publicly accessible. The differences between these data products are mostly caused by the input source data, the integration method, and the sampling of input atmospheric fields. Improvement in any of those factors may lead to a reduction of residual atmospheric signals in the instrument data, which otherwise alias into long wavelength mass signals, and negatively affect the accuracy of monthly mean gravity field solutions. This is especially critical for GRACE-FO and the next generation of gravity missions (*Gruber and Team*, 2014; *Flechtner et al.*, 2014b; *Panet et al.*, 2013) that aim to determine the geoid with an accuracy of 1 mm (*Anselmi et al.*, 2010).

The most recent release 06 of AOD1B (shown here by AOD1B RL06) consists of ocean and atmospheric components, which are both given in sets of 3-hourly sampled series of spherical harmonic coefficients complete up to degree and order 180 (*Dobslaw et al.*, 2017). The AOD1B RL06 has been improved over the previous versions on two aspects: (1) an update ocean bottom pressure from an unconstrained simulation with the global ocean general circulation model MPIOM (Max Planck Institute ocean model) (*Jungclaus et al.*, 2013); and (2) the atmosphere component of AOD1B, which is denoted as ATM RL06, is computed based on an updated analysis and forecast data out of the operational high-resolution global numerical weather prediction (NWP) model from the European Centre for Medium-Range Weather Forecasts (ECMWF) after 2007. ATM RL06 prior to 2007 has been computed using ERA-Interim atmospheric fields.

Previous studies have investigated de-aliasing products on various processing levels (e.g., *Zenner*, 2013): (i) Stokes coefficients (they are usually converted to surface pressure or equivalent water height), (ii) KBRR (K-Band Range Rate) residuals, and (iii) monthly mean gravity field solutions. *Forootan et al.* (2013, 2014) indicated that the ITG3D product results in considerable improvements over the ATM RL04 and RL05 (not RL06) on level (i). But they also addressed that neither the integration technique (2-D versus 3-D or modified 3-D integration) nor the input data (operational ECMWF versus reanalysis) has an impact on level (iii). *Dobslaw et al.* (2015) presented an updated ESM along with some basic validations against the original ESM (*Gruber and Team*, 2014), and ITG3D (*Forootan et al.*, 2013). Their comparisons of the updated ESM with ITG3D are performed on level (i), for which the results indicate very close correspondence at dominant frequencies. Substantially, on levels (i) and (iii), *Dobslaw et al.* (2016) developed a realistically perturbed synthetic de-aliasing models and estimated their impacts on the gravity fields derived by simulated future gravity missions. *Zenner et al.* (2010) proposed a method to take uncertainties of input atmospheric models into account while computing the ATM products, although this procedure has no significant effect on level (ii) or (iii). Recently, *Rudenko et al.* (2016) indicated the significant impact of AOD1B RL04 and RL05 on precise orbits of altimetry satellites. Their study showed the importance of background models for producing more accurate altimetry and gravity l2 and L3 data products.

In this study, to explore the major factors that may affect the quality of atmospheric (non-tidal) de-aliasing products, we carry out an evaluation of available products including: ATM (the atmosphere component of AOD RL05 and RL06, *Flechtner et al.*, 2014a; *Dobslaw et al.*, 2017), ITG3D (*Forootan et al.*, 2013, 2014), and the updated ESM (*Dobslaw et al.*, 2015, 2016). Unlike most of these previous studies, our comparisons here are made mainly on the level of KBRR residuals, since they are directly estimated from GRACE observations and are sensitive to the background models. Generally speaking, the differences between these de-aliasing products, which are hardly distinguished by monthly mean gravity fields due to the downward continuation and filtering process, is prone to be revealed by KBRR residuals analysis. It is our hypothesis that the smaller resulting KBRR residuals represent less misfit of (more accurate) de-aliasing products (see, e.g., *Zenner et al.*, 2012; *Zenner*, 2013). Previous attempts that use KBRR residuals to validate background models or to detect modeling errors can be also found for example, in *Bosch et al.* (2009), *Han et al.* (2009, 2010) and *Dobslaw et al.* (2017).

Our paper is organized as follows: In Sec. 2, a brief data introduction of ATM, ESM, ITG3D products is given. In Sec. 3, the methodology of generating KBRR residuals is first outlined, after that monthly KBRR residuals, as well as the monthly mean gravity fields from 2005 to 2010 are presented to validate our data processing chain. This period is selected since GRACE KBRR measurements contain less noise, and therefore, their quality is reliable. In Sec. 4, we present the resulting daily/monthly/yearly KBRR residuals during 2006-2010 by each atmospheric de-aliasing product in both spatial and temporal domains. In addition, the impact of these three products on the current GRACE gravity fields is analyzed. Finally, Sec. 5 concludes the paper and provides some suggestions for the next version of de-aliasing products.

2. Data

2.1. ATM

ATM RL05 is the atmosphere de-aliasing component of the AOD1B RL05 product released by GFZ (*Flechtner et al.*, 2014a), which is represented by a series of potential coefficients complete up to degree and order (d/o) 100 with temporal resolution of every

6 hours (at 00:00 h, 06:00 h, 12:00 h, and 18:00 h) since year 1976. The procedure of computing this product relies on the input six-hourly atmosphere fields that mainly comprise surface pressure, geopotential, temperature, and specific humidity fields. These input data are all extracted from ECMWF operational analysis (ECMWFop), and are converted into potential coefficients via a three-dimensional (3-D) integration approach including various approximations. ECMWFop is one of the premiere models for medium-range and seasonal-forecasting purposes. Details about the ECMWFop products and ATM products can be found at the Information System and Data Center (ISDC) (http://isdc.gfz-potsdam.de/index.php). The atmospheric part of the latest (RL06) of the AOD1B (ATM RL06) data (*Dobslaw et al.*, 2017) with the temporal sampling of 3 hours is evaluated in our study, which is computed up to d/o 180 (since 2000). The combination between analysis and short-term forecast atmosphere data when producing RL06 (*Dobslaw and Thomas*, 2005) is believed to contribute the most to the 3-hourly samples. Here, we truncate the ATM RL06 data at d/o 100 to be consistent with other products considered in this study.

2.2. ITG3D

ITG3D atmosphere de-aliasing model (*Forootan et al.*, 2013, 2014) is computed up to d/o 100 with the same temporal resolution of 6 h as ATM RL05. The major changes within this new set with respect to the ATM RL05 are twofold: (i) an improved 3-D integration approach with more realistic physical and geometrical Earth's shape, as well as a better numerical integration; and (ii) the input atmospheric data are replaced by the ECMWF's reanalysis data (ERA-Interim; *Dee et al.*, 2011). ERA-Interim includes an improved atmospheric model and assimilation system. Surface and multi-level datasets are available from http://www.ecmwf.int/research/era/do/get/era-interim, and the 6-hourly ITG3D AD products are downloaded from http://www.igg.uni-bonn.de/apmg/index.php.

2.3. ESM

ESM of the European Space Agency (ESA) provides various de-aliasing datasets with the temporal resolution of 3 h and 6 h, as well as the spectral resolution up to d/o 180 and

360. In this study, we only use the atmosphere component (ESM-A) with the temporal sampling of 6 hours without IB-correction (Inverse Barometer) over the oceans. Moreover, to enable the comparisons against ATM and ITG3D models, ESM-A is truncated at d/o 100. In addition to the ESM-A, an alternative ESM-Ac product with the temporal resolution of 3 h is also analyzed, which differs with ESM-A over the Europe, where the atmosphere inputs of the COSMO-EU model (Consortium for Small-Scale Modelling) are used. COSMO-EU model encompasses many local details of the landscape and related flow phenomena that have a pronounced impact on the weather, and therefore, has an improved spatial resolution. Consequently, the main changes within ESM-A(c) (denoted respectively as ESM-A and ESM-Ac) with respect to ATM RL05 are twofold: (i) ESM-A(c) implements the surface pressure integration approach. In another word, a 2-D integration is applied to convert surface pressure values to potential coefficients, rather than using a 3-D integration applied in the ATM and ITG3D; and (ii) ESM-A(c) uses the ERA-Interim archive (or in combination with COSMO-EU) instead of the ECMWFop dataset. All directories that archive the ESM product including the AD component can be downloaded from GFZ website through the web link DOI:10.5880/GFZ.1.3.2014.001. A brief summary of the above three candidate AD products can be found in Table. 1.

Table 1: A summary of the candidate atmospheric de-aliasing products

Product	Period	Temporal resolution	Spectral content	Data source	Integration method
ESM-A	1995-2006	3 h (6 h)	180 (360)	Reanalysis	2-D
ESM-Ac	2006	3 h	360	Reanalysis+COSMO	2-D
ITG3D	2003-2010	6 h	100	Reanalysis	3-D
ATM RL05	1976-2017	6 h	100	Operational(Op)	3-D
ATM RL06	1976-2017	3 h (6 h)	180	Reanalysis+Op	3-D

2.4. GRACE L1b data

A complete set of GRACE L1b data (*Case et al.*, 2002) is the prerequisite to calculate the KBRR residuals. This set mainly includes the KBRR observations, GPS positions, drag-free 3-axis accelerometer measurements along with the star camera measurements, and 6-hourly ocean-de-aliasing product from AOD1B. All data mentioned above are accessible at http://isdc.gfz-potsdam.de/index.php. In what follows, the L1b data along with each of the three AD products mentioned above will be used to calculate the KBRR residuals.

3. Methodology

In this paper, the GRACE level 1B KBRR residuals ($\Delta\dot{\rho}$) are calculated by the Hawk software, Wuhan University, which is designed for gravity recovery (*Yang et al.*, 2017a,b,c) using classical variational-equation approach (*Montenbruck and Gill*, 2000). The light-time and K-Band antenna phase center corrections are applied on the original range-rate observations $\dot{\rho}_{obs}$ (see *Case et al.*, 2002, e.g.,) following

$$\dot{\rho}_{adjust} = \dot{\rho}_{obs} + \Delta_{light-time-correction} + \Delta_{antenna-correction} \quad , \tag{1}$$

where ρ denotes KBRR measurements, and the range rate measurements $\dot{\rho}_{obs}$ are their first-order temporal derivatives. Subsequently, the first version of KBRR residuals $\Delta\dot{\rho}_1$ can be computed by removing the effect of background models from the adjusted observations $\dot{\rho}_{adjust}$ following

$$\Delta\dot{\rho}_1 = \dot{\rho}_{adjust} - \dot{\rho}_{nominal} \quad , \tag{2}$$

where the $\dot{\rho}_{nominal}$ represents the nominal range rate measurements, which are obtained by differentiating ranges $\rho_{nominal}$ as

$$\dot{\rho}_{nominal} = \frac{d\rho_{nominal}}{dt} = (\vec{\nu}_A - \vec{\nu}_B) \cdot \vec{e}_{AB} \quad , \tag{3}$$

where $\vec{\nu}$ are state vectors (velocities) of GRACE satellite A or satellite B; $\vec{e}_{AB}$ denotes the unit vector along the direction of GRACE twin-satellite baseline, which is also the line-of-sight (LOS) unit vector defined by

$$\vec{e}_{AB} = \frac{\vec{X}_{AB}}{d\rho_{nominal}} \quad . \tag{4}$$

To obtain $\rho_{nominal}$, we use

$$\begin{aligned} \vec{X}_{AB} &= \vec{X}_A - \vec{X}_B \\ \rho_{nominal} &= \sqrt{\vec{X}_{AB}^T \cdot \vec{X}_{AB}} \quad , \end{aligned} \tag{5}$$

where $\vec{X}$ contains the positions of GRACE satellite A or satellite B. By substituting Eq. (5) and Eq. (4) into Eq. (3), one can estimate $\dot{\rho}_{nominal}$, which will be further reduced by $\dot{\rho}_{adjust}$ using Eq. (2) to derive the desired KBRR residuals $\Delta\dot{\rho}_1$. However, as shown by Eq. (5), the state vectors $\vec{\nu}, and \vec{X}$ have to be given beforehand, which are usually calculated by implementing an orbit integration (propagation) from the initial state vector $\vec{\nu}_0, \vec{X}_0$, and a-priori force models. Therefore, the nominal range rate ($\dot{\rho}_{nominal}$) does not contain the ranging instrument errors (because $\dot{\rho}_{nominal}$ is not a product of the ranging system but it is computed from orbit positions), and therefore, it could be used as an alternative measure to evaluate AD products. The results of this evaluation are provided in the Appendix, which show that the magnitude of differences between de-aliasing products exceed the noise floor of the ranging system. Therefore, development of AD products should be considered to produce more accurate time-variable gravity fields.

In the following, we introduce the force models used to generate satellite state vectors. The nominal static gravity field is modeled by GIF48 (*Ries et al.*, 2011) complete up to d/o 160. Third-body gravitational perturbations, together with the indirect J2 effect, are computed from the positions and velocities of both Sun and Moon according to JPL DE405 planetary ephemeris (*Standish*, 1995). Subsequently, ocean tides are removed using EOT11a model (*Savcenko and Bosch*, 2012) complete up to d/o 120, associated with 18 major tidal constituents (eight long periodic, four diurnal, five semi-diurnal, one nonlinear constituent) and 238 minor tides. Remaining gravitational force models including solid Earth tides and pole tides, as well as general relativistic perturbations are computed according to the International Earth Rotation Service (IERS) 2010 conventions (*Petit and Luzum*, 2010).

Non-tidal high-frequency oceanic variability is calculated from the ocean component of the GFZ AOD RL05 product. To implement the required comparisons, AD products are chosen from ATM, ITG3D, and ESM. In each scenario, the KBRR residuals are estimated using one of the de-aliasing products. Therefore, all other background models are kept unchanged, and the estimated differences between these scenarios are compared to each other. By doing this, the assessments are independent of the introduced a priori models, and simply represent the effect of changing AD products. We expect that, the choice of other background models does not alter final results and conclusions (assessments are not shown in this paper). More details could be found in Table. 2, where in particular the difference of KBRR-residuals is the key index for assessments that will be used in what follows.

Table 2: An introduction of the related concepts for assessing atmosphere de-aliasing (AD) models
.

	Concept	Remarks	Signal Content
1	True Range Rate	true value (unknown)	gravity fields (static and temporal) + **true AD effect** + other forces
2	KBRR	observations (Obs)	gravity fields (static and temporal) + **true AD effect** + other forces + instrument error
3	Nominal range rate	nominal Obs (Nom)	gravity fields (static) + **AD model** + other force models
4	KBRR-residuals	Obs versus Nom	gravity fields (temporal)+**AD model errors**+ other force model errors + instrument error
5	Difference of KBRR-residuals	KBRR residuals II versus KBRR residuals I	**AD model II errors minus AD model I errors**

We should mention here that the time (t)-dependent KBRR residuals $\Delta\dot{\rho}_1$ estimated above are not the final estimations. We apply another calibration to remove a bias, linear trend, and 1-CPR parameter per orbital revolution (about 94 minutes) from them following *Kim* (2000) and *Zhao et al.* (2011):

$$\Delta\dot{\rho}_2(t) = \Delta\dot{\rho}_1(t) + A + B\ t + C\ sin(\frac{2\pi t}{T_{rev}}) + D\ cos(\frac{2\pi t}{T_{rev}}) + E\ t\ sin(\frac{2\pi t}{T_{rev}}) + F\ t\ cos(\frac{2\pi t}{T_{rev}}) \quad , \tag{6}$$

where T_{rev} is the revolution period, and (A, B, C, D, E, and F) are coefficients that need to be estimated using a least squares method. Subsequently, a 3-σ outlier detection is applied on $\Delta\dot{\rho}_2$ to remove values that are greater than 3 times of the standard deviation value of each arc. By this, the ultimate KBRR residuals ($\Delta\dot{\rho}_2$) are well established. The cleaned up time series of residuals ($\Delta\dot{\rho}_2$) are analyzed in two ways: (1) the RMS of KBRR residuals for each day are computed to form new time-series; and (2) the residuals are cut out over specific areas (continents or globe) by assigning each residual to the mid-point of the orbit positions of the two satellites. Then for each $1^\circ \times 1^\circ$ bin, the RMS of residuals over a given period (daily, monthly, or yearly) is computed and formed global maps that are shown in Sec. 4.

After a complete removal of background models including the AD product in a standard data processing chain, the time-series of daily RMS of KBRR residuals over years 2005-2010 are calculated and presented in Fig. 1 left. Compared to the one presented by *Dahle et al.* (2012), the shape and trend of both KBRR residuals RMS time series agree very well in spite of minor differences in the amplitude. With these KBRR residuals, we produce a monthly gravity model up to d/o 60 using spherical harmonics base-functions (called Hawk-SH60). This comparison indicates a comparable accuracy with the official GRACE level 2 gravity fields in terms of degree variance of the geoid height, see Fig. 1 right panel. Particularly, the correlation coefficient between Hawk-SH60 and CSR RL05 is found to be 0.99. More details and evaluations of Hawk-SH60 products can be found in *Yang et al.* (2017a,b).

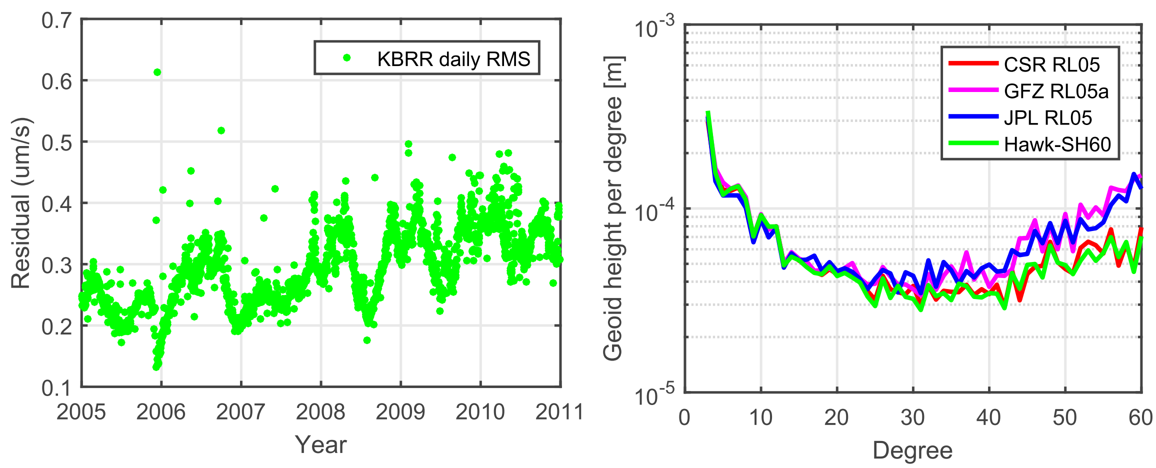

Figure 1: The left panel represents the time-series of daily RMS of KBRR residuals from Jan 2005 to Dec 2010, for which AOD1B RL05 products are applied to remove high-frequency atmospheric and oceanic variability. The right panel represents the degree variance of the geoid height derived from the mean of CSR RL05, GFZ RL05a, JPL RL05 and Hawk-SH60 monthly models averaged during 2005-2010, relative to GIF48. Hawk-SH60 stands for the monthly gravity fields of this study, which are computed up to d/o 60.

Another insight about the performance of estimated gravity fields is shown in Fig. 2, which illustrates the trend and annual amplitude maps in terms of Equivalent Water

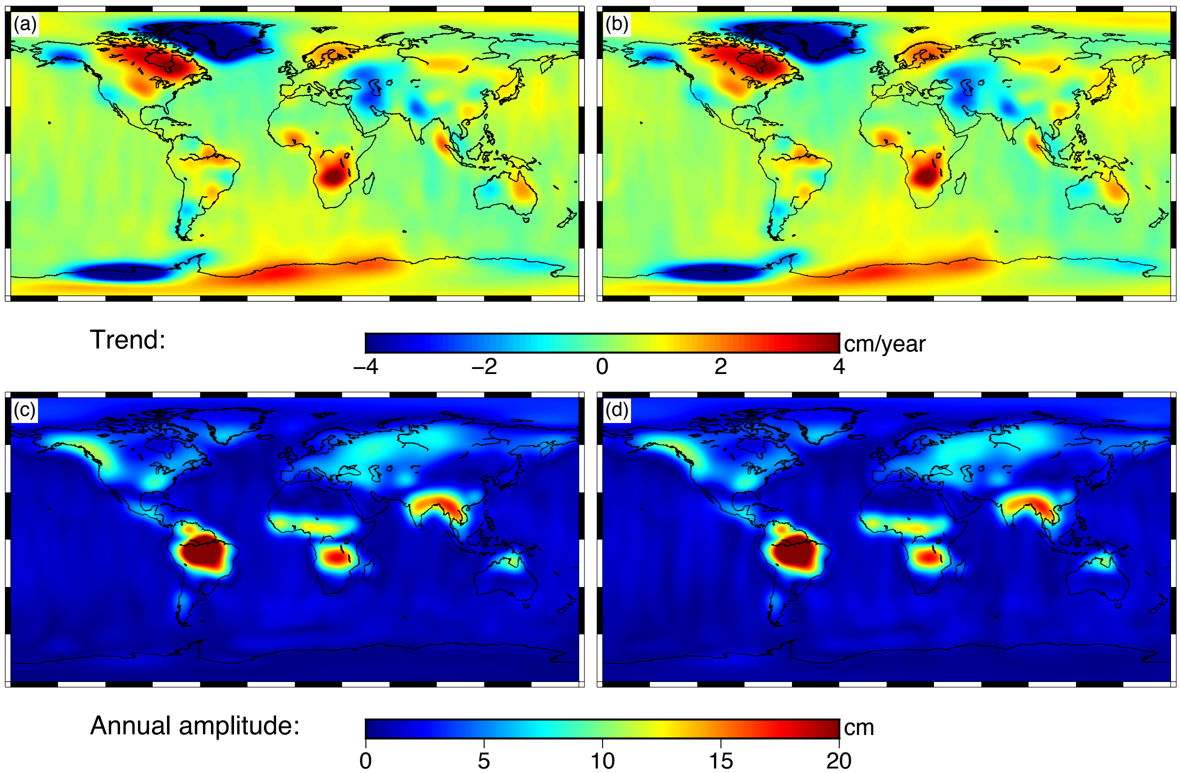

Figure 2: Two maps on top indicate linear trend in terms of Equivalent Water Height (EWH) changes (mm/year) derived from Hawk-SH60 and CSR RL05 monthly gravity fields covering 2005-2010 from (a) CSR RL05, and (b) Hawk-SH60. Bottom maps indicate annual amplitudes in terms of EWH (mm) from (c) CSR RL05, and (d) Hawk-SH60.

Height (EWH) for CSR RL05 and Hawk-SH60, covering Jan 2005-Dec 2010. It is evident that the results of CSR RL05 and Hawk-SH60 are fairly similar (see Fig. 2, and compare the top left to the top right, also compare the bottom left to the bottom right). The similarity of these two models is also supported by comparing the statistics of basin averages as shown in Table. 3, where the differences are mostly less than 5% except for the basins that have weak trend or annual amplitude signals. Therefore, we are confident that there are no potential errors in our data processing chain that introduce an undue adverse effect in the calculation of KBRR residuals, and consequently, they will correctly reflect the impact of the AD products on the gravity inversion.

Table 3: A summary of basin averaged EWH results derived from monthly gravity fields covering Jan 2005 to Dec 2010. Gravity products include the monthly output of our in-house software Hawk using coefficients of up to degree and order (d/o) 60 (Hawk-SH60), and the official products of CSR RL05.

Region	Area [$10^6 km^2$]	Trend [cm/yr]		Annual amplitude [cm]	
		Hawk-SH	CSR RL05	Hawk-SH	CSR RL05
Amazon	6.20	0.53	0.55	17.55	17.40
Nile	5.40	0.25	0.16	3.68	3.41
Congo	3.83	1.26	1.23	4.00	4.16
Mississippi	3.30	0.80	0.83	5.45	5.33
Greenland	2.10	-5.97	-5.97	3.78	3.73
Yangtze	1.81	0.40	0.35	4.03	4.06
Mekong	0.81	-0.20	-0.21	10.45	10.69
Yellow	0.76	-0.10	-0.20	0.93	1.03

4. Results

4.1. Comparisons of KBRR residuals

The following comparisons are carried out for January to December 2006, during which the three 6-hourly de-aliasing data sets of ATM RL05, ITG3D, and ESM are available (see Table. 1). The 3-hourly ATM RL06 will be later evaluated over a longer period of 2006-2010. For each AD product, the KBRR residuals associated with daily RMS values are generated to form the time-series. Subsequently, the time series of KBRR residuals that correspond to one of the AD model is reduced from another as shown in Fig. 3. The results indicate that the daily differences between ATM RL05 and ESM in terms of the RMS of KBRR residuals (the green scattered points), as well as the median value of these differences throughout 2006, i.e. the dashed red line indicating 2.8 nm/s. The amplitude of differences is found to reach up to 20 nm/s, which is below the precision of the current GRACE KBRR system that is only able to distinguish signal stronger than 200 nm/s (see, e.g., *Beutler et al.*, 2010; *Loomis et al.*, 2012; *Chen et al.*, 2015; *Flechtner et al.*, 2016). However, this amplitude is strong enough to be sensed by the next generation of GRACE-type gravity missions or even the planned GRACE-FO mission that will carry a laser interferometer measurement system at an expected precision of 0.6 nm/s (see, *Loomis et al.*, 2012). The averaged improvement of ESM instead of ATM RL05 is found to be 2.8 nm/s, which is above 0.6 nm/s and indicates that the RL05 data will be potentially an error source. In the same way, we can further observe from Fig. 3(b) that ITG3D reduces the median of RMS to 3.4 nm/s (denoted by the red dashed line).

Comparing the RMS results from ESM and ITG3D (see Fig. 3(c)), we find that ITG3D reduces the median by only 0.6 nm/s that can be detected neither by the current K-Band instrument nor by the future laser ranging system of GRACE-FO. A similar result can also be observed by comparing monthly KBRR residuals in Table. 4, and another verification of Table. 4 by a Fisher statistical test (F-test), which indicates that the improvements of 'ATM-ESM' and 'ATM-ITG3D' are reliable ($p = 7.6 \times 10^{-6}$ and $p = 1.0 \times 10^{-5}$, respectively), while the improvement of 'ESM-ITG3D' is found to be less significant ($p = 0.16$).

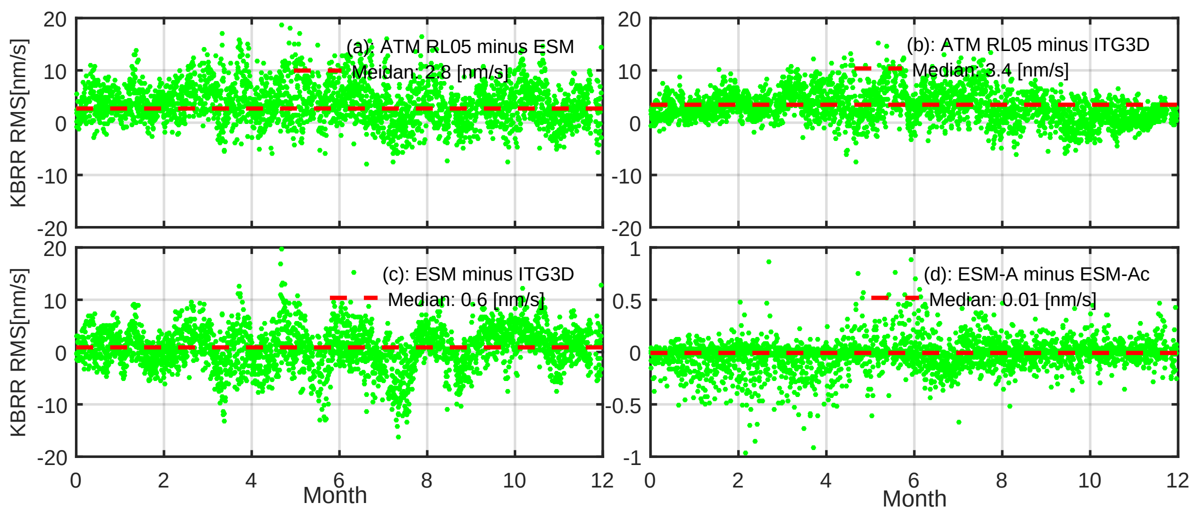

Figure 3: Green dots represent the differences of the RMS of KBRR residuals during 2006 between two pairs of AD products, and the red dashed lines represent the corresponding one-year median values of the green dots. (a) ATM RL05 versus ESM; (b) ATM RL05 versus ITG3D; (c) ITG3D versus ESM; and (d) ESM-A versus ESM-Ac. Note that the differences are computed by reducing the latter model from the former one. Therefore the positive one-year median value indicates that the latter model is found generally better than others.

The estimated RMS of KBRR residuals along GRACE orbit is able to reflect the overall impact of using different AD products on time-variable gravity field recovery. Therefore, in the following, we present the gridded RMS of KBRR residuals to specify the spatial distribution of model misfits. For this purpose, the one year time-series in Fig. 3 are projected onto the spatial domain to generate gridded maps as shown in Fig. 4. These results represent the mean differences of gridded KBRR residuals (for 2006)

Table 4: Comparisons of monthly KBRR residuals RMS during 2006, between ATM RL05 (denoted as ATM below), ESM and ITG3D products. The unit is $[nm/s]$, and the latter product improves over the former one if the computed difference is positive.

Month	ATM	ITG3D	ESM	ATM-ESM	ATM-ITG3D	ESM-ITG3D
Jan	244.4	**240.7**	242.0	2.4	3.7	1.3
Feb	299.0	**295.3**	295.5	3.5	3.7	0.2
Mar	353.9	**349.1**	351.0	2.9	4.8	1.9
Apr	379.1	374.1	**374.0**	5.1	5.0	-0.1
May	358.3	**354.2**	355.0	3.3	4.1	0.8
Jun	372.2	367.5	**367.3**	4.9	4.7	-0.2
Jul	408.1	**403.3**	404.3	3.8	4.8	-1.0
Aug	385.0	383.2	**380.7**	4.3	1.8	-2.5
Sep	401.5	**398.8**	399.3	2.2	2.7	0.5
Oct	364.2	**361.2**	363.3	0.9	3.0	2.1
Nov	305.1	**300.7**	304.2	0.9	4.4	3.5
Dec	262.2	**260.1**	260.9	1.3	2.1	0.8

estimated by replacing the ATM RL05, ITG3D and ESM as done before. In Fig. 4, the $1^\circ \times 1^\circ$ quadratic grid boxes marked in red (positive) suggest that the latter model has a smaller RMS, and therefore, it is a better model for reducing high-frequency atmospheric mass changes from gravity products, and the blue boxes represent vice versa.

In support of our assumption, a supplementary experiment between ESM-A and ESM-Ac is setup to examine the impact of using a regional atmospheric model on the RMS of KBRR residuals. The results are shown in Fig. 3(d) and the Fig. 4(d), where Fig. 3(d) indicates that the median value of RMS after reducing ESM-A and ESM-Ac is too small (0.01 nm/s) to be detected by the K-Band or the laser ranging systems. This likely indicates that the regional improvement of input atmosphere variability from the COSMO-EU model can hardly lead to a global improvement. In Fig. 4(d) (with the ocean being masked), the effects on KBRR residuals are successfully confined within the region where the COSMO-EU is supposed to take effect. The border of COSMO-EU model is marked by a thick red line, which includes the whole Europe associated with a part of the northern Africa, as shown in Fig. 4(d). Considering the fact that COSMO-EU has a better spatial resolution, ESM-Ac mostly performs better than ESM-A as expected (the red grids are dominant, see Fig. 4(d)). Out of the COSMO-EU's domain, the differences vanish fast as expected. Since no significant differences can be detected over the regions out of Europe, we can be sure that the projection of KBRR residuals to the spatial domain remains within the location of interest and hardly leaks to other regions. In summary, we conclude that using KBRR residuals is an efficient and straight-forward approach to assess the quality of AD products.

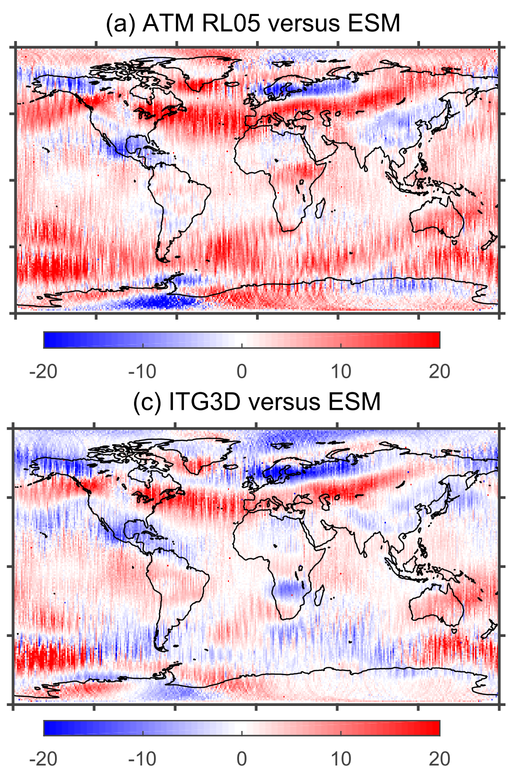

Figure 4: Differences of one-year KBRR residuals RMS over 2006 in $1^\circ \times 1^\circ$ bin, the unit is $[nm/s]$: (a) ATM RL05 versus ESM; (b) ATM RL05 versus ITG3D; (c) ITG3D versus ESM; and (d) ESM-A versus ESM-Ac. Note that the differences are made by reduction of the latter model from the former one, therefore, the regions in red indicate the latter model is better.

Based on the assumption mentioned above, Fig. 4(a) demonstrates the differences between ATM RL05 and ESM model (the former minus the latter), where it can be observed that ESM has a considerable improvement over ATM RL05 at the majority of the globe. The ratio of the red and the blue grid points is 78.9% : 21.1%, which could be regarded as a global improvement in spite of some local deterioration. Particularly, 65% of the red points has the strength beyond 1 nm/s that is sufficient to be monitored by the laser ranging system (0.6 nm/s). Furthermore, Fig. 4(b) compares the differences between the RMS of KBRR residuals derived from ATM RL05 minus that of ITG3D, where the ratio of the red and the blue grids is found to be 78.4% : 21.6%. We note that

Table 5: The percentage of the red parts in monthly spatial maps over 2006, and the maps denote the differences of monthly KBRR residuals RMS between two aliasing products. The red part in maps is where the latter de-aliasing product performs better than the former one, see Fig. 4.

Month	ATM-ESM	ATM-ITG3D	ITG3D-ESM
Jan	67.5%	65.8%	50.1%
Feb	64.0%	67.1%	47.4%
Mar	61.8%	62.3%	53.1%
Apr	60.2%	67.1%	49.2%
May	61.2%	64.8%	49.0%
Jun	60.1%	64.3%	49.0%
Jul	62.5%	64.0%	53.8%
Aug	57.4%	61.8%	48.9%
Sep	60.3%	60.5%	50.9%
Oct	63.8%	60.7%	55.6%
Nov	66.1%	60.4%	51.5%
Dec	65.4%	67.2%	41.8%

the percentage is calculated without weighting grid cells, because the spatial KBRR RMS values have already been weighted using latitude-dependent scales. Another investigation of the monthly performance has been concluded in Table. 5, where we also find that the red parts in scenarios (ATM RL05-ESM) and (ATM RL05-ITG3D) are over 50% of the globe in each month. In terms of the statistics we obtained so far, the ITG3D and ESM models apparently perform better than ATM RL05 in the performed spatial analysis, and this finding is consistent with the one contained in Fig. 3. According to the official description of ITG3D, ESM and ATM RL05 products, the major differences between ATM RL05 and the other two products are caused by the input atmosphere fields (source data, *Forootan et al.*, 2013; *Dobslaw et al.*, 2015), which consequently cause the derived differences in the KBRR residuals. Therefore, the ERA-interim data applied in ITG3D and ESM seems to be better suited for reducing the high frequency atmospheric mass changes than the ECMWFop used in ATM RL05.

However, given the same input ERA-interim fields, the generated AD products may not be identical due to the various assumptions within the 2-D and 3-D integration approaches. Figure 4(c) illustrates the comparison between ITG3D products (*Forootan et al.*, 2013) derived from a 3-D approach, and ESM products (*Dobslaw et al.*, 2015) derived from a 2-D integration approach. Although the 3-D method is theoretically more comprehensive than the 2-D (e.g., it considers all vertical structure of atmosphere), only small improvements are found in Fig. 4(c) as the ratio of the red with respect to the blue is found to be 52% : 48%, see Table. 5 for more details. This result confirms previous findings that the choice of 3-D integration approach has much smaller impact than the input atmospheric fields.

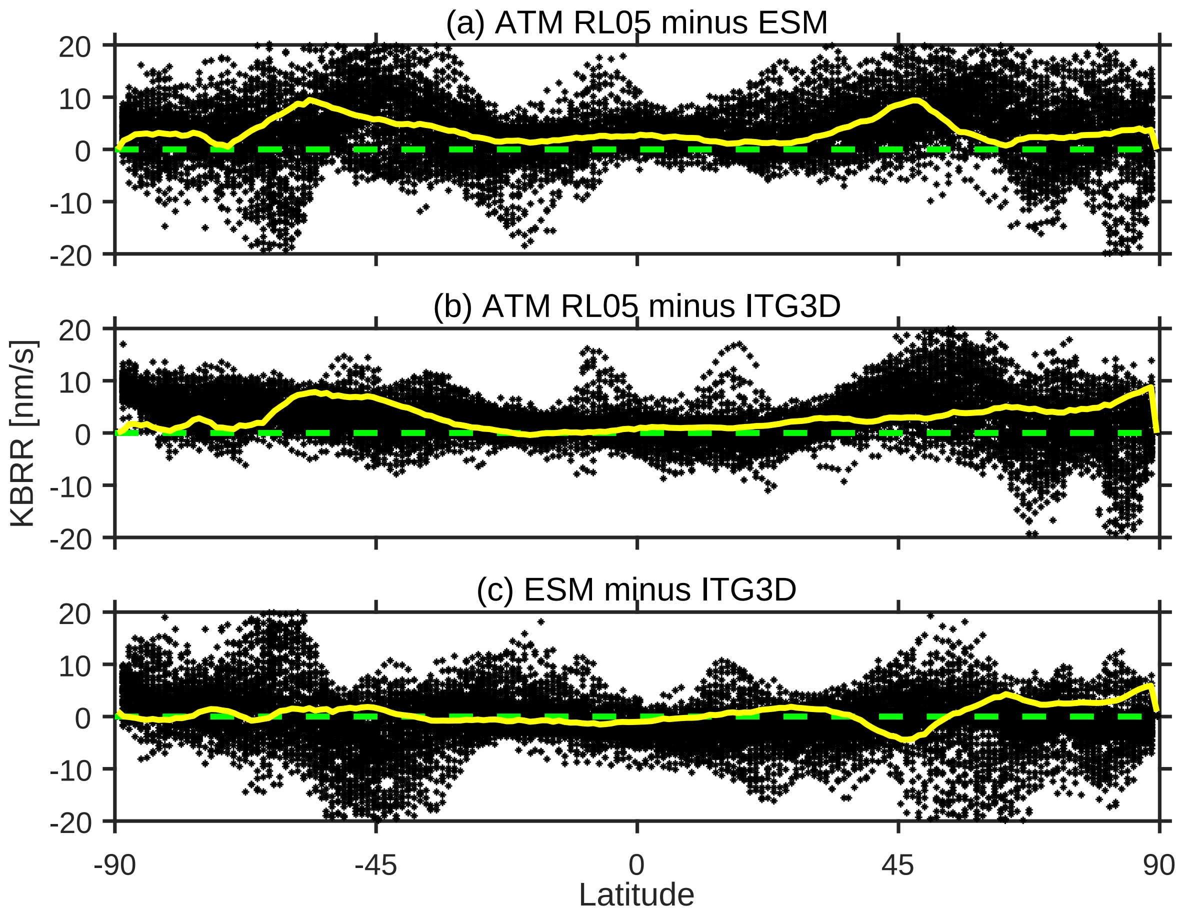

Figure 5: The black dots denote the spatial distribution of the differences of the KBRR residuals during 2006, plotted against the latitude; the yellow solid line are the median value of the black dots, and the green dashed line stands for zero. (a) ATM RL05 versus ESM-A products; (b) ATM RL05 versus ITG3D; and (c) ESM-A versus ITG3D.

To further understand the nature of Fig. 4, the differences of the estimated RMS of KBRR residuals during 2006 are plotted against latitude in black dots as shown in Fig. 5. The green dashed line denotes the zero, and the yellow represents the median value of the KBRR residuals that are distributed at given latitude. From Fig. 5(a)(b), the superiority of ESM or ITG3D model with respect to ATM RL05 model can be concluded, as the yellow solid lines are entirely above the green dashed lines: 99.7% of the yellow solid line stays above 0 nm/s in Fig. 5(a), and it is 99.5% in Fig. 5(b). Our conclusion is that both ESM-A and ITG models reduce the KBRR residuals globally.

Considering Fig. 5(c), which indicates the differences of RMS between ESM and ITG3D models, we find only 48% of the yellow solid line stays above 0 nm/s. In this sense, ITG3D has no evident global improvement over ESM, however, we also notice that the ITG3D model has a stable and better performance at particular regions with latitude $[60°, 90°]$. The average value of the yellow line in Fig. 5(c) at latitude $[60°, 90°]$ is 1.9 nm/s, which denotes that ITG3D can reduce the KBRR residuals over high-latitude regions. This conclusion is consistent with those of previous studies (*Berrisford et al.*, 2011; *Forootan et al.*, 2013), which indicate major differences between atmospheric fields are distributed over high-latitude regions. Therefore, we suggest that ITG3D is more appropriate for gravity recovery in particular regions such as Greenland.

4.2. Effects on monthly mean gravity field

In previous sections, we show the approach and results of using in-orbit KBRR residuals to assess AD products. In this section, we will assess how differences in these products might be transferred to an ultimate monthly gravity solution. We should mention here that because of the complexity of numerical procedure within time-variable gravity inversion, the differences between the de-aliasing products that are captured by the KBRR residuals analysis might not be one-by-one reflected in the estimated gravity fields. This is another motivation for carrying out investigations in this section.

February 2006 is selected as an example to compare gravity fields that are reduced by different AD products. In Fig. 6(a), our numerical results are shown in terms of degree variance of geoid height. In addition to the curves of gravity fields recovered from ATM RL05 and ESM (ITG3D result is not shown because it almost overlaps the one reduced by ESM), the GRACE pre-launch baseline, as well as the current CSR RL05 and its calibrated errors are plotted as our references. It could be observed that the gravity field derived from ATM RL05 (in solid red line) and ESM (in solid blue line) are both comparable to that from CSR RL05 (in solid green line). However, the discrepancies between the solutions still exist and should not be neglected.

From the results, it can be seen that the current accuracy of the GRACE recovered signal (in black dashed line) is one order of magnitude away from the baseline accuracy (in cyan dashed line), and minor modification of the background model might not be able to improve the accuracy, considerably. From Fig. 6(a), the red dashed line (ATM RL05 versus ESM) is between the current GRACE accuracy (calibrated error) and the baseline accuracy, which indicates that ESM (or ITG3D) is able to contribute to the next generation of GRACE rather than the current one. Figure 6(b) illustrates the geoid height transformed from the red dashed line in Fig. 6(a), which shows the amplitude of differences between gravity fields obtained from ATM RL05 and ESM can reach up to more than 1 mm after applying a Gaussian filter with 500km half-width radius. For discussion about the impact of filtering we refer to *Forootan et al.* (2014). The errors estimated here should be considered especially for the next generation of gravity missions that aim to reach the accuracy of 1 mm in terms of geoid height.

The blue dashed line in Fig. 6(a) indicates the differences between monthly gravity fields recovered from ITG3D and ESM, and lies much lower than the curves that correspond to other differences. Moreover, the blue dashed line is even lower than the pre-launch accuracy of GRACE (the cyan dashed line) after degree 15. This result suggests that the use of 3D instead of 2D integration approach can slightly improve the recovered gravity signal in terms of baseline accuracy, and its impact can be detected for the low degree terms (i.e., <15).

4.3. Revisit the jump in the ATM RL05 products

A jump in 2006 has been previously reported in the ATM RL05 product (e.g., *Duan et al.*, 2012; *Forootan et al.*, 2014; *Rudenko et al.*, 2016). Their analysis shows that the changes in the vertical layers of the input atmosphere fields from ECMWFop are responsible for the jump that occurs between January and February. Their finding of this jump is based on the level-2 GAC/GAA monthly mean atmosphere non-tidal products (*Flechtner et al.*, 2014a).

In Fig. 7(a), we show how differences of KBRR residuals can be related to EWH

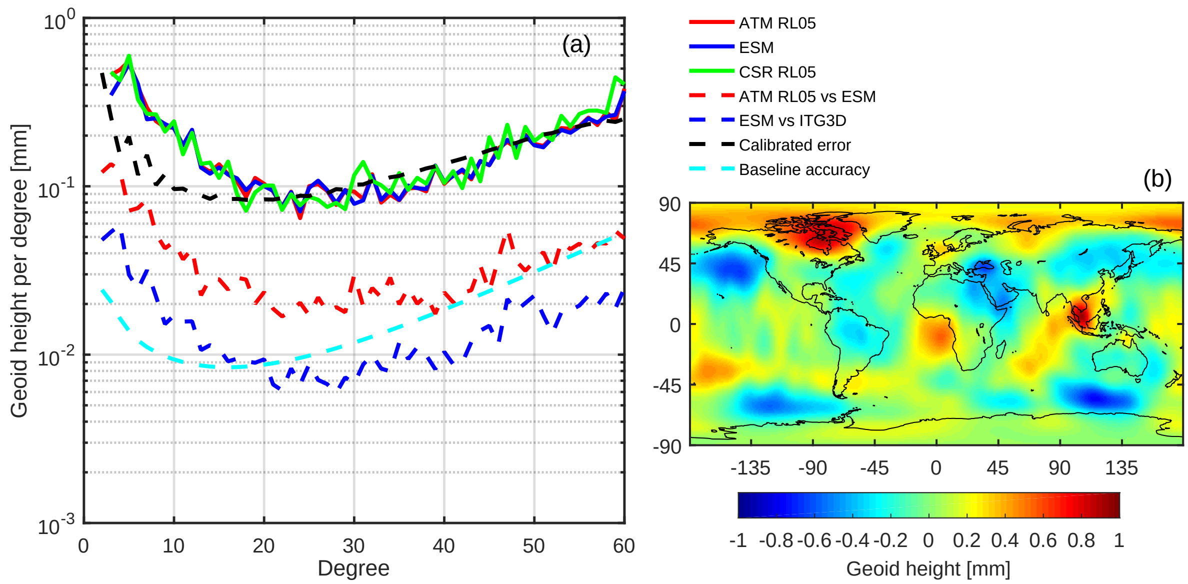

Figure 6: (a) Impact of using different atmospheric de-aliasing products on a GRACE monthly solution, on February 2006, computed up to degree and order (d/o) 60, in terms of geoid height. (b) Differences between the gravity fields inverted by reducing ITG3D minus the one reduced by ATM RL05 in terms of geoid heights after applying a Gaussian filter with 500km half-width radius.

changes caused by these jumps. ATM RL05 and ITG3D are averaged daily and transformed to EWH. Subsequently, we calculate the differences between these daily EWH maps. The results demonstrate that the daily differences before January 29th are similar to that on January 29th (indicated by Fig. 7(a)), min/max/spatially weighted root mean squares (wrms) of the differences are found to be roughly 3.9/5.9/0.6. However, after January 29th, min/max/wrms of the differences change sharply to -9.0/9.0/1.0, as shown in Fig. 7(b). We find that these changes mainly happen within the continents (see Fig. 7(b)), therefore, the daily mean of EWH within continents are plotted in Fig. 7(c). The results clearly distinguish the jump occurring on January 29th, 2006. This conclusion could be further supported by analyzing the KBRR residuals, which are also shown in Fig. 7(c). Our estimations indicate that a jump of $\sim 3\ mm$ in terms EWH causes a change of $\sim 2\ nm/s$ in terms of the RMS of KBRR residuals. This relationship (i.e. 3 mm EWH $\sim$=2 nm/s RMS of KBRR) might be considered as a measure to evaluate the impact of possible jumps in the AD products on the gravity recovery procedure.

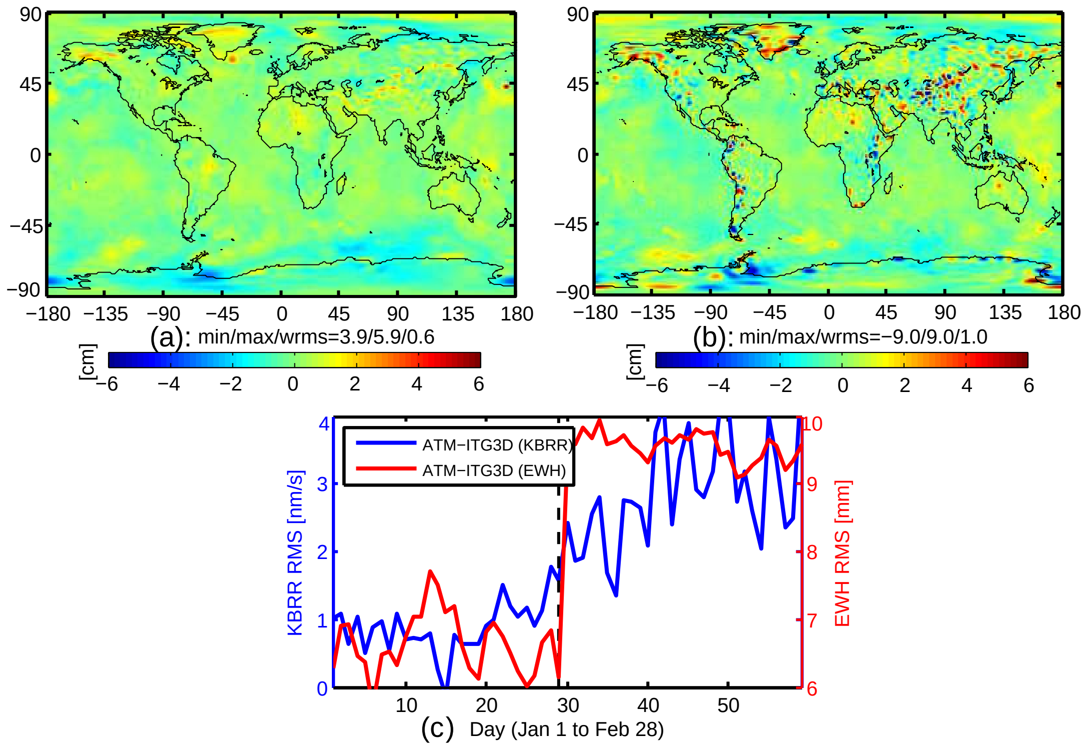

Figure 7: (a) Differences of EWH maps between ITG3D and ATM RL05 on January 29th, 2006. (b) Differences of EWH maps between ITG3D and ATMRL05 on Jan 30th, 2006. (c) The blue solid line shows differences between daily RMS of the continental KBRR residuals derived from ITG3D and ATM RL05, and the red solid line denotes differences between the RMS of continental EWH from ITG3D and ATM RL05.

4.4. A validation of ATM RL06

Here, we carry out an extended evaluation (2006-2010) of the atmospheric part of the latest release of AOD1B products (shown by ATM RL06), which are compared with ATM RL05 and ITG3D products. ESM products are excluded since they are only available up to 2006 (see Table. 1). ATM RL06 is truncated at d/o 100 to be spectrally consistent with other products. We also do not add the ocean part of RL06 to have a consistent comparison.

The time-series of calibrated daily RMS of KBRR residuals derived from ATM RL05 is shown in Fig. 8a, from which a range of changes that is about 200-600 nm/s can be detected. These values can be considered as our reference, to continue our comparisons. An average improvement derived from using ITG3D instead of ATM RL05 is found to be 2.4 nm/s (shown in Fig. 8(b)), whose magnitude is insignificant. Figure 8(c) and (d) indicate that using ATM RL06 instead of RL05 (or ITG3D) largely reduces the RMS of KBRR residuals by 44.2 nm/s on average (see the red dashed lines in Figs. 8(c) and (d)). Considering the high accuracy of laser ranging system, we believe the averaged value of 44.2 nm/s is significant enough to affect gravity fields that will be estimated by the next generation of satellite gravity missions. In fact, the improvement of ATM RL06 over RL05 in terms of KBRR-residuals might be under-estimated because of the ranging instrument error (see Eq. (2)). In order to study the potential maximal improvement that might be offered by changing AD models, we conduct another comparison in terms of nominal ranging rate (see Eq. (3)) over year 2006. The numerical results indicate that the median value of daily RMS of the differences between ATM RL05 and ATM RL06 has reached 295.6 nm/s. Similarly, it is 29.8 nm/s between ATM RL05 and ITG3D. We also computed the RMS between ATM RL05 and RL06, and between that of RL05 and ITG3D, which are found to be 301.2 nm/s and 29.5 nm/s, respectively. These results indicate that it is necessary to update AD products for the GRACE-FO and future missions.

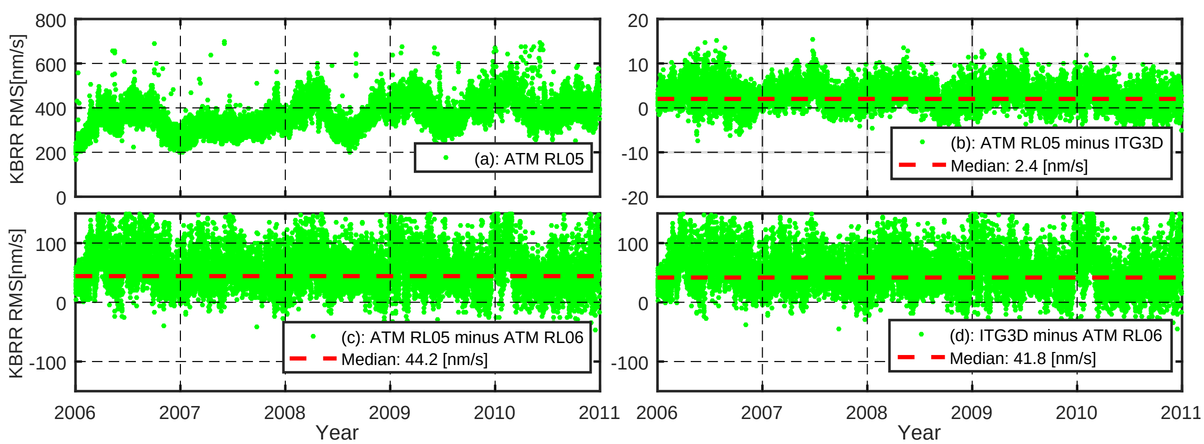

Figure 8: Green dots represent the RMS of KBRR residuals, and the red dashed lines represent the median values of the RMS during 2005-2010. Time series are derived by reducing (a) ATM RL05; (b) ATM RL05 versus ITG3D; (c) ATM RL05 versus ATM RL06; and (d) ITG3D versus ATM RL06. The differences are made by reducing the latter model from the former one, therefore the positive median value indicates that the latter model is better.

These results can be further supported by Fig. 9, where the time-series of daily RMS of KBRR residuals are yearly averaged and projected onto the spatial domain. It is evident from Figs. 9(a)(c)(e)(g) and (i) that ATM RL06 products are more efficient than ATM RL05 to reduce the RMS, and the amplitudes of yearly average differences can reach even up to 100 nm/s. The portions of positive values (red color) in Figs. 9(a)(c)(e)(g) and (i) that correspond to years 2006-2010 are found to be 83%, 79%, 83%, 80%, and 74%, respectively. In parallel, when comparing ITG3D to ATM RL06 in Figs. 9(b)(d)(f)(h)(j), we find the portions of the red are 78%, 75%, 80%, 79%, and 69%, respectively. Better results derived from RL06 is likely due to its higher temporal and spatial resolution, while a detailed assessment will be performed in future.

5. Summary and Conclusion

In this study, an evaluation of existing non-tidal atmospheric de-aliasing (AD) products is carried out. Differences between these products, including ATM RL05 and RL06 from AOD1B, ITG3D, and ESM, are analyzed and their impact on GRACE and future satellite gravity missions is evaluated using the root mean squares (RMS) of KBRR residuals as a measure. Our assessments during 2006 indicate a reduction of 3.4 nm/s in the RMS while using ITG3D instead of ATM RL05, and reduction of 2.8 nm/s while using ESM instead of ATM RL05. The differences are found below the accuracy of the current KBRR system (i.e. ~ 200 nm/s) but above that of the laser ranging system (i.e. ~ 0.6 nm/s) designed for the GRACE-FO mission.

We also assess the spatial distribution of the estimated RMS of KBRR residuals during 2006. Our results indicate that ITG3D and ESM respectively perform better than ATM RL05 over 78.4% and 78.9% of the globe. By averaging the KBRR residuals against latitudes, we could show that ITG3D performs better than ESM in the high latitude regions $[60°, 90°]$. Time-variable gravity fields from GRACE, while considering various atmospheric de-aliasing products within the inversion, demonstrate that future GRACE-like missions are likely sensitive to the improvements of ESM and ITG3D over ATM RL05. Initial validations of the atmospheric part of the latest version of AOD1B (ATM RL06) are carried out during 2006-2010 as well. Our results indicate that the 3-hourly RL06 data result in a significant decrease in the KBRR residuals. The averaged reduction of RMS of KBRR residuals (computed against ATM RL05) during 2006-2010 is found to be 44.2 nm/s, indicating that RL06 performs much better than all the current

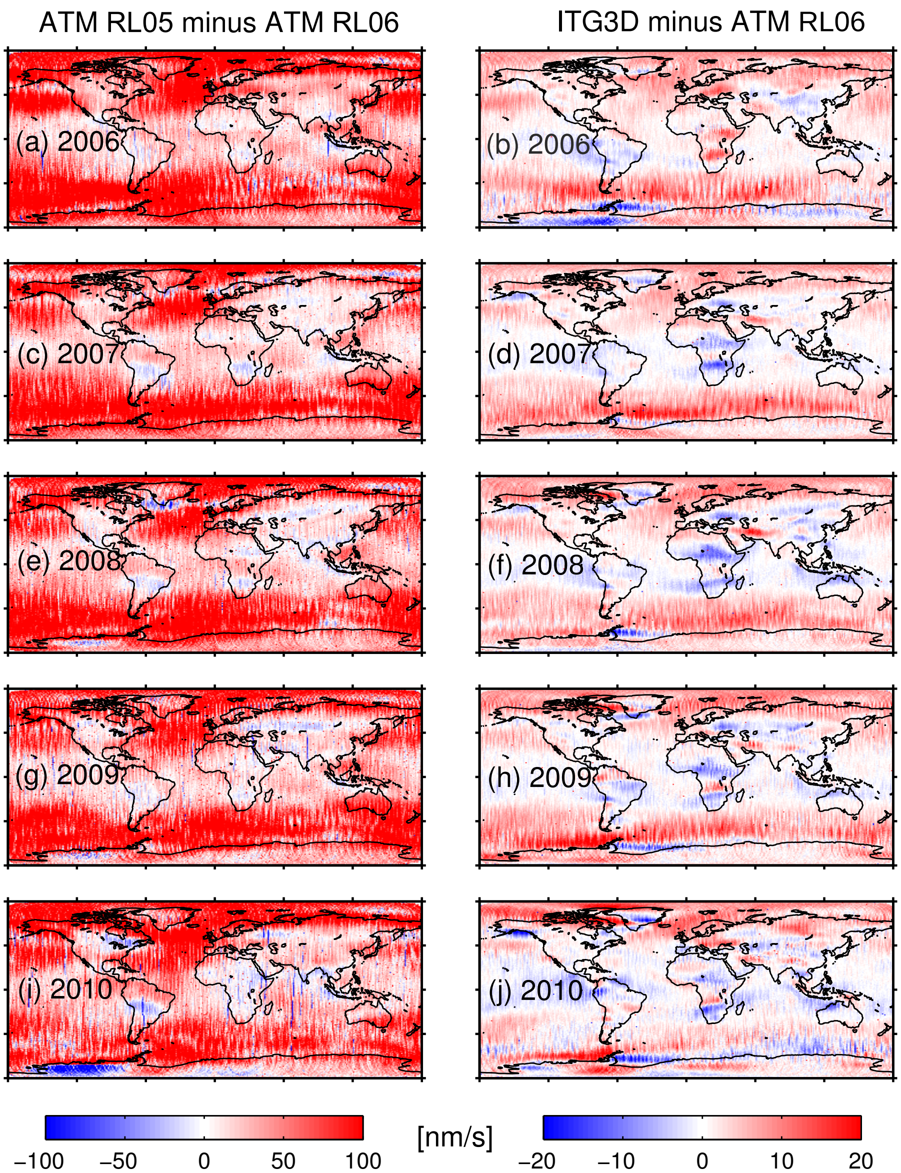

Figure 9: Year-to-year spatial gridded RMS of KBRR residuals. Maps are generated at $1^\circ \times 1^\circ$ bin, and their unit is $[nm/s]$: (a)(c)(e)(g) and (i) represent the results of ATM RL05 versus ATM RL06 over years 2006-2010, respectively; (b)(d)(f)(h) and (j) represent the results of ITG3D versus ATM RL06 over years 2006-2010, respectively. The regions in red color are where the latter product performs better than the former one.

6-hourly AD products.

In our view, the above results suggest that changing input fields data from ECMWFop to ERA-Interim is beneficial and improves the final quality of AD products, and this is also found as the major cause of discrepancies between ITG3D, ESM, ATM RL05 and RL06. In particular for RL06, the combination of 3-hourly reanalysis data and hourly forecast data has well improved the quality of AD products. Compared to the option of input atmosphere fields, as well as their temporal resolution, the option of integration approach (3-D or 2-D) is found globally less significant considering the current accuracy of GRACE-like gravity inversion. The impact of integration at high-latitude regions is found considerable. Thus, the improved 3-D integration in ITG3D is suggested to be used for generating next versions of AD products.

Our work may contribute to the GRACE community in the following aspects. (1) While acknowledging the outstanding performance of ATM RL06 produced by GFZ, there is still room to work on the different options, which can be set during estimating AD products, e.g., changing sampling rate, input data, and the details of 3-D integration method. These settings must be ensured to be consistent with the processing strategy of gravity inversion. Therefore, as a further validation, the improved 3-D integration method as well as 3-hourly reanalysis data will be used to update the ITG3D model in our future work. (2) Apart from the AD products considered in this contribution, a possible investigation will be carried out to evaluate other background models, for example, tidal and non-tidal ocean models, while considering GRACE or GRACE-FO, as well as LISA-type or Bender-type future satellite gravity missions.

Acknowledgments

Authors are grateful to Dr. Kosuke Heki and two anonymous reviewers for their valuable comments, which we used them to improve this study. We thank JPL and GFZ for providing the GRACE level 1b data and atmospheric de-aliasing data products. This work is supported by open foundation of SKLGED2018-1-4-E and China Postdoctoral Science Foundation (Grant No. 2017M622517), as well as the Strategic Priority Research Program of the Chinese Academy of Sciences, Grant No. XDB23030100. We also acknowledge supports by grants from the National Natural Science Foundation of China (NSFC, Grant No. 41374020), and by the Belmont Forum/IGFA G8 via US National Science Foundation (Grant No. ICER-1342644).

References

Anselmi, A., S. Cesare, and R. Cavaglia (2010), Assessment of a Next Generation Mission for Monitoring the Variations of Earths gravity, *Final Report, Document SD-RP-AI-0668, Thales Alenia Space, 22 December 2010*.

Berrisford, P., P. Kllberg, S. Kobayashi, D. Dee, S. Uppala, A. J. Simmons, P. Poli, and H. Sato (2011), Atmospheric conservation properties in ERA-Interim, *Quarterly Journal of the Royal Meteorological Society*, *137*(659), 1381–1399, doi:10.1002/qj.864.

Beutler, G., A. Jäggi, L. Mervart, and U. Meyer (2010), The celestial mechanics approach: application to data of the grace mission, *Journal of Geodesy*, *84*(11), 661–681, doi:10.1007/s00190-010-0402-6.

Bosch, W., R. Savcenko, F. Flechtner, C. Dahle, T. Mayer-Gürr, D. Stammer, E. Taguchi, and K.-H. Ilk (2009), Residual ocean tide signals from satellite altimetry, GRACE gravity fields, and hydrodynamic modelling, *Geophysical Journal International*, *178*(3), 1185–1192, doi:10.1111/j.1365-246X.2009.04281.x.

Boy, J.-P., and B. F. Chao (2005), Precise evaluation of atmospheric loading effects on Earth's time-variable gravity field, *Journal of Geophysical Research: Solid Earth*, *110*, B08,412, doi:10.1029/2002JB002333.

Case, K., G. Kruizinga, and S. Wu (2002), GRACE level 1B data product user handbook, *JPL Publication D-22027*.

Chen, Q., Y. Shen, X. Zhang, and et al (2015), Monthly gravity field models derived from GRACE level 1b data using amodified short arc approach, *Journal of Geophysical Research: Solid Earth*, *120*(3), 1804–1819, doi:10.1002/2014JB011470.

Dahle, C., F. Flechtner, C. Gruber, D. König, R. König, G. Michalak, and K.-H. Neumayer (2012), The new GFZ RL05 GRACE gravity field model time series, in *Geophys. Res. Abstracts*, vol. 14.

Dahle, C., F. Flechtner, C. Gruber, D. König, R. König, G. Michalak, and K.-H. Neumayer (2014), GFZ RL05: an improved time-series of monthly GRACE gravity field solutions, in *Observation of the System Earth from Space-CHAMP, GRACE, GOCE and future missions*, pp. 29–39, Springer.

Daras, I., and R. Pail (2017), Treatment of temporal aliasing effects in the context of next generation satellite gravimetry missions, *Journal of Geophysical Research: Solid Earth*, *122*(9), 7343–7362, doi:10.1002/2017JB014250.

Dee, D., S. Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, P. Bauer, et al. (2011), The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Quarterly Journal of the royal meteorological society*, *137*(656), 553–597, doi:10.1002/qj.828.

Dobslaw, H., and M. Thomas (2005), Atmospheric induced oceanic tides from ECMWF forecasts, *Geophysical Research Letters*, *32*(10), 189–191, doi:10.1029/2005GL022990.

Dobslaw, H., I. Bergmann-Wolf, R. Dill, E. Forootan, V. Klemann, J. Kusche, and I. Sasgen (2015), The updated ESA Earth System Model for future gravity mission simulation studies, *Journal of Geodesy*, *89*(5), 505–513, doi:10.1007/s00190-014-0787-8.

Dobslaw, H., I. Bergmann-Wolf, E. Forootan, C. Dahle, T. Mayer-Gürr, J. Kusche, and F. Flechtner (2016), Modeling of present-day atmosphere and ocean non-tidal de-aliasing errors for future gravity mission simulations, *Journal of Geodesy*, *90*(5), 423–436, doi:10.1007/s00190-015-0884-3.

Dobslaw, H., I. Bergmann-Wolf, R. Dill, L. Poropat, M. Thomas, C. Dahle, S. Esselborn, R. König, and F. Flechtner (2017), A new high-resolution model of non-tidal atmosphere and ocean mass variability for de-aliasing of satellite gravity observations: AOD1B RL06, *Geophysical Journal International*, *211*(1), 263–269, doi: 10.1093/gji/ggx302.

Duan, J., C. Shum, J. Guo, and Z. Huang (2012), Uncovered spurious jumps in the GRACE atmospheric de-aliasing data: potential contamination of GRACE observed mass change, *Geophysical Journal International*, *191*(1), 83–87, doi:10.1111/j.1365-246X.2012.05640.x.

Elsaka, B., E. Forootan, and A. Alothman (2014), Improving the recovery of monthly regional water storage using one year simulated observations of two pairs of GRACE-type satellite gravimetry constellations, *Journal of Applied Geophysics*, *109*, 195–209, doi:10.1016/j.jappgeo.2014.07.026.

Fagiolini, E., F. Flechtner, M. Horwath, and H. Dobslaw (2015), Correction of inconsistencies in ECMWF's operational analysis data during de-aliasing of GRACE gravity models, *Geophysical Journal International*, *202*(3), 2150–2158, doi:10.1093/gji/ggv276.

Famiglietti, J. S., and M. Rodell (2013), Water in the balance, *Science*, *340*(6138), 1300–1301, doi:10.1126/science.1236460.

Flechtner, F., H. Dobslaw, and E. Fagiolini (2014a), AOD1B product description document for product release 05 (Rev. 4.2, May 20, 2014), *Technical Note, GFZ German Research Centre for Geosciences Department*, *1*.

Flechtner, F., P. Morton, M. Watkins, and F. Webb (2014b), Status of the GRACE follow-on mission, in *Gravity, geoid and height systems*, pp. 117–121, Springer, doi:10.1007/978-3-319-10837-7_15.

Flechtner, F., K. H. Neumayer, C. Dahle, H. Dobslaw, E. Fagiolini, J. C. Raimondo, and A. Güntner (2016), What can be expected from the GRACE-FO laser ranging interferometer for Earth science applications?, *Surveys in Geophysics*, *37*(2), 453–470, doi:10.1007/s10712-015-9338-y.

Forootan, E., O. Didova, J. Kusche, and A. Löcher (2013), Comparisons of atmospheric data and reduction methods for the analysis of satellite gravimetry observations, *Journal of Geophysical Research: Solid Earth*, *118*(5), 2382–2396, doi:10.1002/jgrb.50160.

Forootan, E., O. Didova, M. Schumacher, J. Kusche, and B. Elsaka (2014), Comparisons of atmospheric mass variations derived from ECMWF reanalysis and operational fields, over 2003–2011, *Journal of Geodesy*, *88*(5), 503–514, doi:10.1007/s00190-014-0696-x.

Forootan, E., A. Safari, A. Mastafaie, M. Schumacher, M. Delavar, and J. Awange (2017), Large-scale total water storage and water flux changes over the arid and semiarid parts of the middle east from grace and reanalysis products, *Surveys in Geophysics*, *38*(3), 591–615, doi:10.1007/s10712-016-9403-1.

Gruber, T., and N.-D. Team (2014), e2. motion: Earth system mass transport mission (square)concept for a next generation gravity field missionfinal report of project satellite gravimetry of the next generation (NGGM-D). Deutsche Geodätische Kommission der Bayerischen Akademie der Wissenschaften, Reihe B Angewandte Geodäsie Heft Nr. 318, München, *Tech. rep.*, ISBN 978-3-7696-8597-8.

Han, S.-C., H. Kim, I.-Y. Yeo, P. Yeh, T. Oki, K.-W. Seo, D. Alsdorf, and S. B. Luthcke (2009), Dynamics of surface water storage in the Amazon inferred from measurements of inter-satellite distance change, *Geophysical Research Letters*, *36*(9), 8379–8387, doi:10.1029/2009GL037910.

Han, S.-C., J. Sauber, and S. Luthcke (2010), Regional gravity decrease after the 2010 Maule (Chile) earthquake indicates large-scale mass redistribution, *Geophysical Research Letters*, *37*(23), 817–824, doi:10.1029/2010GL045449.

Hardy, R., R. Nerem, and D. Wiese (2017), The impact of atmospheric modeling errors on GRACE estimates of mass loss in greenland and antarctica, *Journal of Geophysical Research*, *122*(30), 10–28, doi:10.1002/2017JB014556.

Jungclaus, J. H., N. Fischer, H. Haak, K. Lohmann, J. Marotzke, D. Matei, U. Mikolajewicz, D. Notz, and J. S. Storch (2013), Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPIEarth system model, *Journal of Advances in Modeling Earth Systems*, *5*(2), 422–446, doi:10.1002/jame.20023.

Kim, J. (2000), Simulation study of a low-low satellite-to-satellite tracking mission, Ph.D. thesis, The University of Texas at Austin.

Kusche, J., V. Klemann, and W. Bosch (2012), Mass distribution and mass transport in the Earth system, *Journal of Geodynamics*, *59*(60), 1–8, doi:10.1016/j.jog.2012.03.003.

Loomis, B. D., R. S. Nerem, and S. B. Luthcke (2012), Simulation study of a follow-on gravity mission to GRACE, *Journal of Geodesy*, *86*(5), 319–335, doi:10.1007/s00190-011-0521-8.

Montenbruck, O., and E. Gill (2000), *Satellite orbits*, vol. 2, 384pp pp., Springer.

Panet, I., J. Flury, R. Biancale, T. Gruber, J. Johannessen, M. van den Broeke, T. van Dam, P. Gegout, C. Hughes, G. Ramillien, et al. (2013), Earth system mass transport mission (e. motion): a concept for future earth gravity field measurements from space, *Surveys in Geophysics*, *34*(2), 141–163, doi:10.1007/s10712-012-9209-8.

Petit, G., and B. Luzum (2010), IERS conventions 2010. International earth rotation and reference systems service, *Tech. rep.*, IERS Technical Note.

Ramillien, G., R. Biancale, S. Gratton, X. Vasseur, and B. S (2011), GRACE-derived surface water mass anomalies by energy integral approach: application to continental hydrology, *Journal of Geodesy*, *85*(6), 313–328, doi:10.1007/s00190-010-0438-7.

Ries, J., S. Bettadpur, S. Poole, and T. Richter (2011), Mean background gravity fields for GRACE processing, in *GRACE science team meeting, Austin, TX*, pp. 8–10.

Rudenko, S., D. Dettmering, S. Esselborn, E. Fagiolini, and T. Schöne (2016), Impact of atmospheric and oceanic de-aliasing level-1b (AOD1B) products on precise orbits of altimetry satellites and altimetry results, *Geophysical Journal International*, *204*(3), 1695–1702, doi:10.1093/gji/ggv545.

Sakumura, C., S. Bettadpur, and S. Bruinsma (2014), Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models, *Geophysical Research Letters*, *41*(5), 1389–1397, doi:10.1002/2013GL058632.

Sasgen, I., H. Konrad, E. Ivins, M. Van den Broeke, J. Bamber, Z. Martinec, and V. Klemann (2013), Antarctic ice-mass balance 2003 to 2012: regional reanalysis of GRACE satellite gravimetry measurements with improved estimate of glacial-isostatic adjustment based on GPS uplift rates, *The Cryosphere*, *7*(5), 1499–1512, doi:10.5194/tc-7-1499-2013.

Savcenko, R., and W. Bosch (2012), EOT11a-empirical ocean tide model from multi-mission satellite altimetry, *Report No.89, Deutsches Geodätisches Forschungsinstitut (DGFI)*.

Schumacher, M., J. Kusche, and P. Döll (2016), A systematic impact assessment of GRACE error correlation on data assimilation in hydrological models, *Journal of Geodesy*, *90*(6), 537–559, doi:10.1007/s00190-016-0892-y.

Schumacher, M., E. Forootan, van Dijk AIJM, H. Müller Schmied, R. Crosbie, J. Kusche, and P. Döll (2018), Improving drought simulations within the Murray-Darling Basin by combined calibration/assimilation of GRACE data into the WaterGAP Global Hydrology Model, *Remote Sensing of Environment*, *204*, 212–228, doi:doi:10.1016/j.rse.2017.10.029.

Standish, E. (1995), The JPL planetary and lunar ephemerides DE402/LE402, in *Bulletin of the American Astronomical Society*, vol. 27, p. 1203.

Tapley, B. D., S. Bettadpur, M. Watkins, and C. Reigber (2004), The gravity recovery and climate experiment: Mission overview and early results, *Geophysical Research Letters*, *31*(9), L09,607, doi:10.1029/2004GL019920.

Yang, F., Hsu, M. Zhong, C. Wang, and Z. Z (2017a), GRACE global temporal gravity recovery through the radial basis function approach, *Chinese Journal of Geophysics*, *60*(4), 1332–1346.

Yang, F., J. Kusche, E. Forootan, and R. Rietbroek (2017b), Passive-ocean radial basis function approach to improve temporal gravity recovery from GRACE observations, *Journal of Geophysical Research: Solid Earth*, *122*, 6875–6892, doi:10.1002/2016JB013633.

Yang, F., C. Wang, and Hsu (2017c), Towards a more accurate temporal gravity model from GRACE observations through the kinematic orbits, *Chinese Journal of Geophysics*, *60*(1), 37–49, doi:10.6038/cjg20170104.

Zenner, L. (2013), Atmospheric and oceanic mass variations and their role for gravity field determination, Ph.D. thesis, München, Technische Universität München, Diss., 2013.

Zenner, L., T. Gruber, A. Jäggi, and G. Beutler (2010), Propagation of atmospheric model errors to gravity potential harmonics: impact on GRACE de-aliasing, *Geophysical Journal International*, *182*(2), 797–807, doi:10.1111/j.1365-246X.2010.04669.x.

Zenner, L., E. Fagiolini, I. Daras, F. Flechtner, T. Gruber, T. Schmidt, and G. Schwarz (2012), Non-tidal atmospheric and oceanic mass variations and their impact on GRACE data analysis, *Journal of Geodynamics*, *59*, 9–15, doi:10.1016/j.jog.2012.01.010.

Zhao, Q., J. Guo, Z. Hu, C. Shi, J. Liu, H. Cai, and X. Liu (2011), GRACE gravity field modeling with an investigation on correlation between nuisance parameters and gravity field coefficients, *Advances in Space Research*, *47*(10), 1833–1850, doi:10.1016/j.asr.2010.11.041.

Appendix A

The Hawk software classifies the KBRR processing into three phases: (1) preprocessing the KBRR observations, (2) calculating the nominal ($d\rho_{nominal}$) range rate (or known as the predicted range rate $d\rho_{predicted}$), and (3) computing the KBRR-residuals by subtracting the nominal range rate from the KBRRs. Each phase and the errors involve in them can be found in Fig. 10.

Table 6: Corresponding statistics (the median value of the time series) of Fig. 11 in 2006, where the results are reported in [nm/s].

Scenarios	Comparison at the level of nominal rates	Comparison at the level of KBRR-residuals (reported in Sec. 4)
ATM RL05 v.s. RL06	295.6	44.6
ITG3D v.s. ATM RL05	29.8	3.4
ITG3D v.s. ESM	3.9	0.6

The main results of this manuscript (presented in the Sec. 4) are based on the KBRR residuals (Phase 3 in Fig. 10), which contain errors from both the ranging system, as well as those of the background models. Although, selecting this measure is useful to directly justify whether changes in the de-aliasing products affect the final gravity solutions, one cannot isolate the maximum potential of replacing AD products. Therefore, as an alternative measure, we use the nominal range rates of Eq. (3), which are computed from orbital positions, to measure the impact of de-aliasing products. The RMS of differences (between nominal range rates) does not contain errors of the ranging system, and therefore, it reflects the maximum impact one might expect after changing the AD products. In Figs. 11 and 12, the daily RMS of differences in nominal rates are shown that correspond to the year 2006 and 2007-2010, respectively. Figure 11 compares ATM RL05, ATM Rl06, ITG3D, and ESM, while Fig. 12 contains ATM RL05, RL06, and ITG3D. The corresponding statistics are reported in Tables. 6 and 7, which indicate that, as expected, the magnitude of differences in terms of KBRR-residuals is much smaller than those obtained here from the nominal rates. Furthermore, both measures indicate that the differences are bigger than the noise level of laser-ranging system. Additionally, we note that, the nominal (predicted) KBRR differences between ATM RL05 and RL06 even exceed the noise level of the K-band ranging system, which indicates the importance of using ATM RL06 and other possible future atmospheric de-aliasing products for the GRACE-FO mission.

Table 7: Corresponding statistics (the median value of the time series) of Fig. 12 covering 2007-2010, where the results are reported in [nm/s].

Scenarios	Comparison at the level of nominal rates	Comparison at the level of KBRR-residuals (reported in Sec. 4)
ATM RL05 v.s. RL06	285	43.6
ITG3D v.s. ATM RL05	18.5	2.2

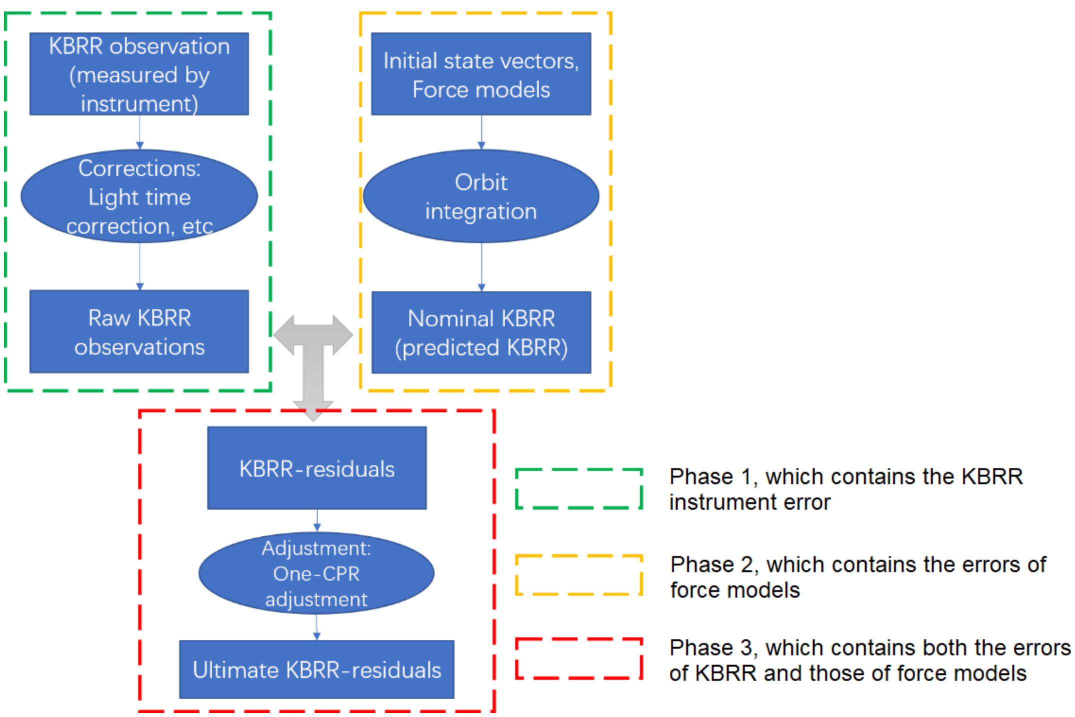

Figure 10: Flowchart of processing KBRR data in the Hawk software.

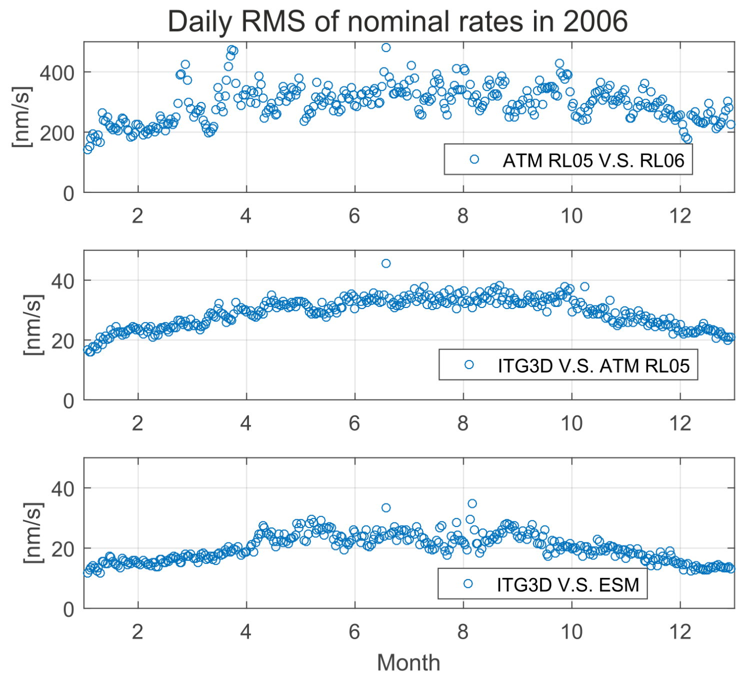

Figure 11: Time series of daily RMS of nominal rates ($d\rho_{nominal}$) in 2006.

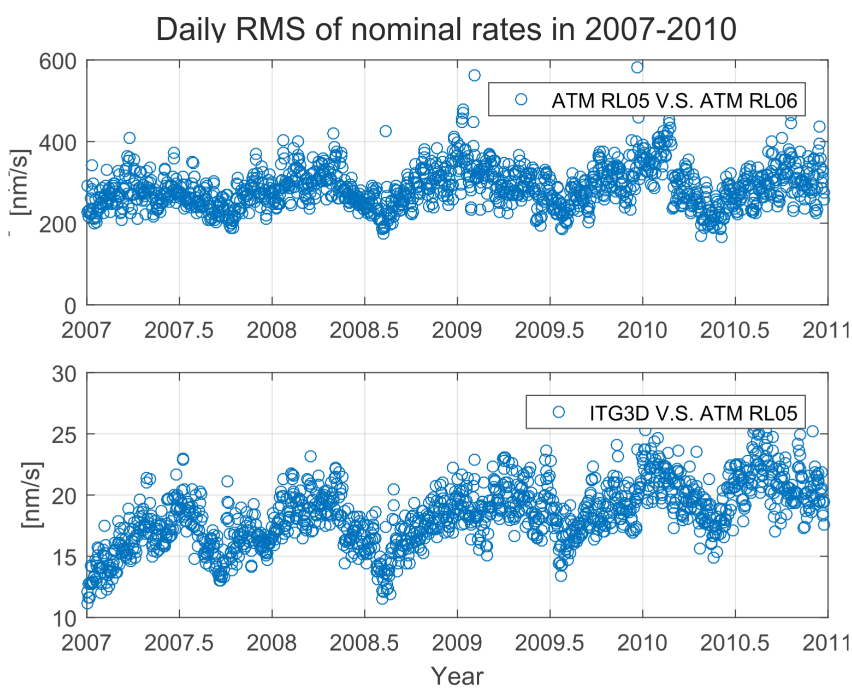

Figure 12: Time series of daily RMS of nominal rates ($d\rho_{nominal}$) covering 2007-2010.