



## Article

# Monsoon-Based Linear Regression Analysis for Filling Data Gaps in Gravity Recovery and Climate Experiment Satellite Observations

Hussein A. Mohasseb <sup>1,2</sup> , Wenbin Shen <sup>1,2,\*</sup> and Jiashuang Jiao <sup>2,3</sup>

<sup>1</sup> State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan 430079, China; hussein985@whu.edu.cn

<sup>2</sup> School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China; jiashuang.jiao@whu.edu.cn

<sup>3</sup> Wuhan Gravitation and Solid Earth Tides National Observation and Research Station, Wuhan 430071, China

\* Correspondence: wbshen@sgg.whu.edu.cn; Tel.: +86-27-6877-1539

**Abstract:** Over the past two decades, the Gravity Recovery and Climate Experiment (GRACE) satellite mission and its successor, GRACE-follow on (GRACE-FO), have played a vital role in climate research. However, the absence of certain observations during and between these missions has presented a persistent challenge. Despite numerous studies attempting to address this issue with mathematical and statistical methods, no definitive optimal approach has been established. This study introduces a practical solution using Linear Regression Analysis (LRA) to overcome data gaps in both GRACE data types—mascon and spherical harmonic coefficients (SHCs). The proposed methodology is tailored to monsoon patterns and demonstrates efficacy in filling data gaps. To validate the approach, a global analysis was conducted across eight basins, monitoring changes in total water storage (TWS) using the technique. The results were compared with various geodetic products, including data from the Swarm mission, Institute of Geodesy and Geoinformation (IGG), Quantum Frontiers (QF), and Singular Spectrum Analysis (SSA) coefficients. Artificial data gaps were introduced within GRACE observations for further validation. This research highlights the effectiveness of the monsoon method in comparison to other gap-filling approaches, showing a strong similarity between gap-filling results and GRACE's SHCs, with an absolute relative error approaching zero. In the mascon approach, the coefficient of determination ( $R^2$ ) exceeded 91% for all months. This study offers a readily usable gap-filling product—SHCs and smoothed gridded observations—with accurate error estimates. These resources are now accessible for a wide range of applications, providing a valuable tool for the scientific community.

**Keywords:** GRACE; GRACE-FO; gap; linear regression; TWS; hydrology



**Citation:** Mohasseb, H.A.; Shen, W.; Jiao, J. Monsoon-Based Linear Regression Analysis for Filling Data Gaps in Gravity Recovery and Climate Experiment Satellite Observations. *Remote Sens.* **2024**, *16*, 1424. <https://doi.org/10.3390/rs16081424>

Academic Editors: Liming He and José Fernández

Received: 20 January 2024

Revised: 8 April 2024

Accepted: 13 April 2024

Published: 17 April 2024



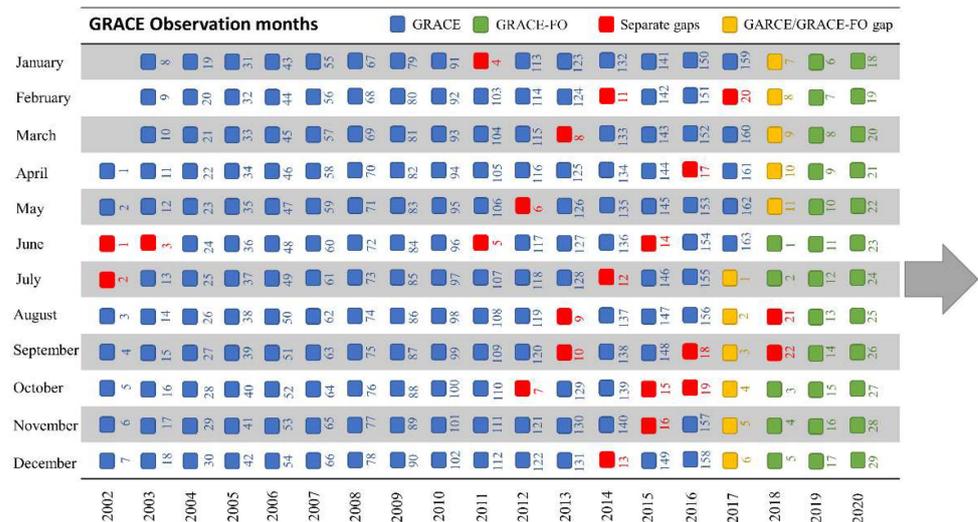
**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The GRACE and GRACE-FO missions have a dedicated purpose: to provide continuous and highly detailed monitoring of the Earth's gravity field [1,2]. This invaluable data resource offers us the chance to delve into various aspects of our planet, including comprehending its internal composition, tracking global environmental shifts, and monitoring alterations in the climate [3,4]. Moreover, it facilitates the study of changes in glaciers, mass changes caused by ice melting, groundwater levels, and numerous other hydrological and geophysical applications [5–7]. The GRACE gravity satellites provide critical scientific data in the form of SHCs representing Earth's monthly mean gravity field. These coefficients are resolved up to degrees 60 or 96 [8,9]. The key organizations involved in processing GRACE data include the Center for Space Research (CSR) at the University of Texas at Austin, the German Research Centre for Geosciences (GFZ), and the NASA Jet Propulsion Laboratory (JPL). Additionally, GRACE and GRACE-FO offer mass concentration data in the form of

mascons, processed by three main centers: CSR, JPL, and NASA Goddard Space Flight Center (GSFC) [10].

GRACE successfully completed its scientific mission over a span of 15 years, concluding in June 2017. Subsequently, GRACE's successor satellite, GRACE-FO, was launched in May 2018 [1,11]. GRACE-FO continues to fulfill the GRACE science mission, but the 11-month gap between these two missions inevitably imposes limitations on our capacity to systematically analyze and fully exploit the satellite observations acquired by both GRACE and GRACE-FO over the past two decades [12–15]. Nonetheless, the presence of data gaps, particularly the nearly one-year hiatus between missions, has hindered the uninterrupted analysis of the extensive 20-year dataset encompassing both GRACE and GRACE-FO missions. In total, there were approximately 20 individual months with missing data in GRACE and 2 consecutive months with gaps in GRACE-FO, as well as 11 months without observations during the transition from GRACE to GRACE-FO (as illustrated in Figure 1) [12,16,17]. Isolated instances of missing data for one or two months in a row are not a significant concern, considering the standard temporal resolution of the data. While filling-in the 11-month data gap presents a considerable challenge, previous studies have explored three main categories of solutions to address the data gaps for filling-in GRACE and GRACE-FO observations: using other satellite data, artificial intelligence, and mathematical and statistical methods.



**Figure 1.** GRACE observation months. The blue cells indicate the months that correspond to the existing GRACE observations, while the green cells indicate the months that correspond to the existing GRACE-FO observations. The red cells represent the months for which both GRACE and GRACE-FO observations are missing. Finally, the yellow cells show the gap that exists between the GRACE and GRACE-FO observations.

### 1. Utilizing other satellite data.

Some studies have attempted to fill the gaps by incorporating data from other satellite missions (e.g., Swarm, global position system (GPS), SLR) [18,19]. For instance, European satellite Swarm mission data have been employed for this purpose [19,20]. However, the limitation here is that the data quality of Swarm is primarily suitable for lower-degree gravity fields (below degree 12) [19,21,22]. GPS data have also been considered as a means to fill in data gaps in gravity observations [23]. However, this approach faces two significant challenges: firstly, the uneven distribution of GPS stations: The distribution of GPS stations around the world is not uniform. Some regions have a dense network of GPS stations, while others have fewer or none at all. This non-uniform distribution can lead to data gaps in areas with limited GPS coverage. Secondly, the unavailability of GPS data in some regions: In certain parts of the world, particularly in most of Africa, GPS data may be unavailable or

scarce. This lack of data in these regions makes it challenging to use GPS to fill in the data gaps in a comprehensive manner. As a result of these challenges, most GPS-based solutions for filling data gaps tend to be regional in nature, focusing on areas where GPS data are more readily available and reliable. Furthermore, an additional approach to filling the data gaps involves the use of Satellite Laser Ranging (SLR) data [24,25]. However, this method has its limitations. It can only produce gravity field models up to a 10-degree resolution, which means it captures primarily the long-term variations in large-scale gravity fields [26]. This approach results in lower-resolution gravity field models compared to what can be achieved with GRACE data, which offer higher spatial detail and resolution.

## 2. Artificial Intelligence, Machine Learning, and Deep Learning.

Another approach involves leveraging artificial intelligence techniques, such as machine learning and deep learning [14,27–31]. Machine learning is a popular branch of artificial intelligence widely used for solving scientific and engineering problems. However, it is typically designed for regional gridded observations of hydrological signals, making it less suitable for global and generic SHCs.

## 3. Mathematical and Statistical Methods.

The third category of solutions involves mathematical and statistical methods [15,16,21,32,33]. An example is the method proposed by [16], which uses Singular Spectrum Analysis (SSA) to fill data gaps in GRACE mission data. Initially, SSA was applied to regional terrestrial water storage anomalies in China, limiting its applicability to other research areas like oceanography and solid Earth physics. However, subsequent work by Yi and Sneeuw [16], as well as Wang et al. [21], improved upon this by devising new schemes to fill data gaps with SHCs, making it more broadly applicable.

These various approaches reflect ongoing efforts to address the challenges posed by data gaps in the GRACE and GRACE-FO missions, each with its own strengths and limitations. However, no recommendation has yet been made on the most effective method. On the contrary, despite the widespread use of mascon data by both geodesists and non-geodesist scientists, there is a notable deficiency in research efforts aimed at addressing the filling-in of gaps in mascon data.

In this study, we've taken an innovative approach to fill-in the data gaps in both the GRACE and GRACE-FO missions, including the 11-month gap in mascon and SHC data. The LRA method is employed for estimation, and the approach is novel in how the estimations are conducted. Specifically, the value for each missing month is calculated individually by utilizing the available observations for the same month from each year. For instance, to estimate the missing data for January in a particular year, observations from all the available GRACE January months across different years are drawn upon. This approach is based on the assumption that observations for the same month annually exhibit similar climate conditions, including factors like rainfall, evapotranspiration, temperature, and more.

There are some studies attempting to solve the gap issue with linear regression, and all of them use artificial intelligence to estimate the new coefficients. Zhang et al. [14] evaluated multiple linear regression (MLR), a back propagation neural network (BPNN), and a deep belief network (DBN) to fill the data gap in Greenland and its six sub-regions. They employed these methods to establish the relationships between precipitation, runoff, evapotranspiration, and ice discharge and GRACE-estimated ice mass changes. Sun et al. [30] used three learning-based models, including a deep neural network, MLR, and seasonal autoregressive integrated moving average with exogenous variables, and six GRACE solutions, three SHCs (CSR, JPL, and GFZ), and three mascons (CSR, JPL, and GSFC) to reconstruct the missing monthly data at a grid cell scale.

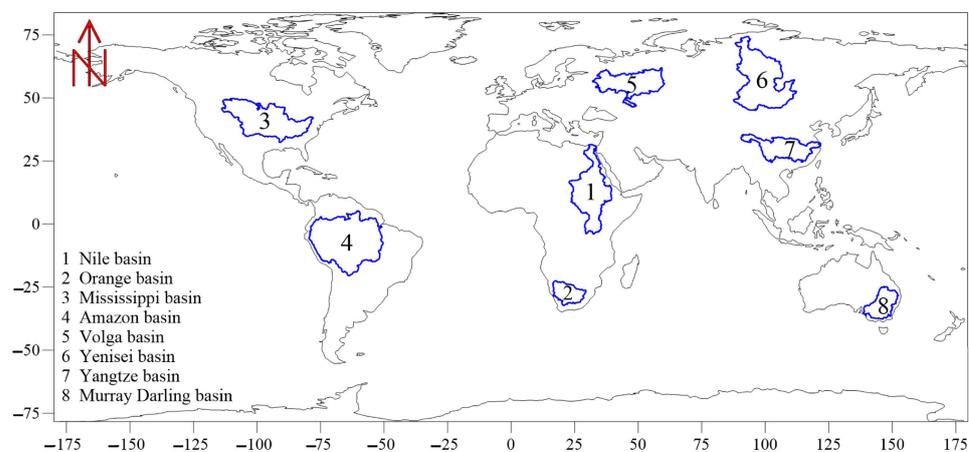
In the present studies bridging the 11-month data gap between GRACE missions, an increasing trend is observed in the adoption of artificial intelligence techniques. However, this study presents improvements in four crucial aspects: Firstly, this analysis is based on the monsoon condition, which gives accurate results in a simple manner. Secondly,

this approach stands out for its simplicity and elegance. Unlike the intricate and resource-intensive experimental design often associated with artificial intelligence methods, this data-filling method is straightforward and easily understandable. This simplicity makes it accessible for implementation with minimal computational burden, making it a practical choice for a wide range of researchers and scientists. Thirdly, this approach aims to provide data products in the form of SHCs, which are versatile and applicable for various GRACE data purposes. Additionally, data products in the form of mascons are offered, catering to hydrology applications and serving the needs of both geodesists and non-geodesist scientists. Fourthly, a notable distinction from previous studies is that these previous studies were primarily designed for hydrological applications, limiting their utility in other domains, including oceanography and solid Earth studies. In contrast, this method is more versatile and adaptable across different research areas, making it valuable for a wider range of applications.

## 2. Materials and Methods

### 2.1. Study Area

In this study, the global scale is generally used for estimation. However, for particular parts, the focus is on eight major basins distributed across various continents, as illustrated in Figure 2. These selected basins include the Nile basin, Orange basin, Mississippi basin, Amazon basin, Volga basin, Yenisei basin, Yangtze basin, and Murray Darling basin. Importantly, all of these basins have a surface area that exceeds the spatial resolution of GRACE data, which is approximately 160,000 square kilometers. This ensures that this analysis covers basins with sufficiently large surface areas for meaningful observations and assessments [34] (Table 1).



**Figure 2.** Eight major basins, including the Nile, Orange, Mississippi, Amazon, Volga, Yenisei, Yangtze, and Murray Darling.

**Table 1.** The river basins include their area and their locations.

Basin	Approximate Area (km <sup>2</sup> )	Location
Nile basin	3,038,100	East and North Africa
Orange basin	850,000	Southern Africa
Mississippi basin	2,980,000	North America
Amazon basin	5,500,000	South America
Volga basin	1,360,000	Europe
Yenisei basin	2,580,000	Asia
Yangtze basin	1,800,000	Asia
Murray Darling basin	1,061,469	Australia

## 2.2. Gravity Data

### 2.2.1. GRACE Data

While the gap-filling method described below can be applied to any mascon or GSM (Gravity Recovery and Climate Experiment Science Team Mass Concentration) GRACE product (CSR, JPL, and GFZ) and is relevant to both d/o 60 and 96, the Release-06 product provided by the CSR processing center [35] for d/o 60 was used for this study. The selection was inspired by its widespread usage in hydrological research.

Data were downloaded spanning from April 2002 to June 2017 for GRACE and from June 2018 to December 2021 for GRACE-FO. The mascon data in this context represent scaled GRACE solutions, allowing for the direct estimation of the time-series of TWS. On the other hand, the SHCs represent Earth's gravity field. CSR SHCs data with a maximum degree of 60 were specifically chosen for this analysis. To estimate the change in water mass using SHCs, the Equivalent Water Height (EWH) can be expressed as [36]

$$\Delta\sigma(\theta, \lambda) = \frac{a\rho_{av}}{3\rho_w} \sum_{n=0}^{\infty} \sum_{m=0}^n \frac{2n+1}{1+k_n} W_n(\bar{P}_{nm}(\cos\theta)) (\Delta\bar{C}_{nm} \cos(m\lambda) + \Delta\bar{S}_{nm} \sin(m\lambda)) \quad (1)$$

where  $a$  is the equatorial radius of the Earth (its value corresponding to the GRACE processing center),  $\rho_{av}$  is the average density of the Earth ( $5517 \text{ Kg/m}^3$ ),  $\rho_w$  is the density of the fresh water ( $1000 \text{ Kg/m}^3$ ),  $P_{nm}$  is the normalized associated Legendre functions in degree  $n$  and order  $m$ ,  $K_n$  is the elastic love number of degree  $n$ ,  $W_n$  corresponds to the Gaussian smoothing operator,  $\theta$  is the latitude,  $\lambda$  is the eastward longitude, and  $\Delta\bar{C}_{nm}$  and  $\Delta\bar{S}_{nm}$  are the monthly spherical harmonics coefficient anomalies. The residual spherical harmonics coefficients,  $\Delta\bar{C}_{nm}$ ,  $\Delta\bar{S}_{nm}$ , of the  $i$ -th monthly are defined by

$$\begin{pmatrix} \Delta C_{nm} \\ \Delta S_{nm} \end{pmatrix}_i = \begin{pmatrix} C_{nm} \\ S_{nm} \end{pmatrix}_i - \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} C_{nm} \\ S_{nm} \end{pmatrix}_i \quad (2)$$

In the analysis, the long-term value was removed using the baseline from 2004 to 2009 of the  $\Delta C_{nm}$  and  $\Delta S_{nm}$  SHCs, where “ $n$ ” and “ $m$ ” are specific degrees and orders. This step is necessary to eliminate the mean field, primarily influenced by the static density distribution within the solid Earth. By removing the static field, all contributions from stored water are also eliminated, allowing us to focus solely on the time-variable component of water storage changes. To enhance the accuracy of the analysis, several corrections to the GRACE and GRACE-FO data were implemented. Firstly, the  $C_{20}$  and  $C_{30}$  coefficients were replaced with their values obtained from SLR as described in [37]. Additionally, to account for geocentric motion, the degree-1 coefficients were corrected using results from Swenson et al. [38]. Furthermore, a smoothing process was applied to mitigate stripes noise in the data. This smoothing was accomplished using a 300 km Gaussian filter, enhancing the quality and clarity of the data.

### 2.2.2. Swarm Data

Swarm satellite data are processed by multiple centers, but only two of these centers have maintained data continuity from December 2013 to the present day. These two centers are the Astronomical Institute of the Czech Academy of Sciences (ASU) and the International Combination Service for Time-Varying Gravity (COST-G). Both of these centers provide monthly SHC processed data, and for this study, the data were truncated at the maximum degree of 40 [21]. Therefore, in this research, processed data from COST-G and ASU up to a degree 40 were chosen.

### 2.2.3. IGG

This study employed an SLR-based recovery method that incorporates the GRACE empirical orthogonal function decomposition model developed by the Institute of Geodesy and Geoinformation at the University of Bonn, referred to as the IGG [39]. The IGG dataset

was derived from SLR (Satellite Laser Ranging) observations and processed using empirical orthogonal functions (EOFs) obtained from the GRACE solutions. This dataset covers the time period from January 2003 to June 2020 and has been truncated to degree 60. The IGG dataset was created using the time-varying gravity field data from ITSG-Grace2018, which spans from April 2002 to August 2016, as the source for EOFs. The  $C_{20}$  coefficient was replaced by the SLR result. To ensure the stability of signals beyond the GRACE mission time frame, the estimation of high-degree and low-degree terms of the SHCs was separated. The high-degree terms were estimated based on the leading modes in the EOF decomposition of the GRACE solutions, while the low-degree terms were determined from the SLR observations. Löcher and Kusche [39] presented six different combinations of time-varying gravity field models based on various maximum degrees of SLR solutions. However, they concluded that the weighted average of these models, known as EnsMean, offers a favorable compromise in terms of global coverage and spatially homogeneous signals. Therefore, for subsequent analyses, the researchers adopted the EnsMean model because it strikes a suitable balance among the available models, as recommended by Löcher and Kusche [39].

#### 2.2.4. Quantum Frontiers (QFs)

The QF dataset comprises monthly gravity field models that encompass SHCs up to degree-and-order (d/o) 60. These models are constructed using tracking data obtained from two sources: HLSST (high-low satellite-to-satellite tracking) and SLR (Satellite Laser Ranging). The computation process involves the utilization of nine SLR geodetic satellites and 27 low-Earth orbiting (LEO) satellites, which include specialized gravity-focused satellites like CHAMP, GOCE, and GRACE A/B. To create these models, SLR and HLSST data are integrated using variance component estimation techniques, resulting in a spatial resolution of approximately 1000–2000 km [13]. For this study, the v2-Kalman-filtered QF data from January 2003 to December 2018, provided by Weigelt [40], were utilized.

#### 2.2.5. Singular Spectrum Analysis (SSA) Coefficients

In the work conducted by Yi and Sneeuw [16], the authors successfully addressed the data gaps in both the GRACE and GRACE-FO missions using the Singular Spectrum Analysis (SSA) method. SSA is a relatively recent technique designed for the analysis of non-linear time-series data. It offers the advantage of generating straightforward and robust models that require less computational resources while maintaining high usability. The SSA approach involves breaking down a time-series into its constituent components, including the underlying trend, periodic variations, and random noise. This decomposition allows for a more detailed understanding of the data's structure and facilitates the filling-in of gaps in a systematic manner. Furthermore, the authors made the data they generated through this method available online.

In this study, the plan is to validate the gap-filling SHCs obtained through LRA using a comparison with SHCs from dataset sources. These datasets include the IGG, QF data, Swarm-ASU, and Swarm-COST-G and data derived from SSA. This validation process will allow us to assess the accuracy and reliability of the gap-filling SHCs obtained through LRA by comparing them with SHCs from other datasets and methods.

#### 2.3. Monsoon Linear Regression Analysis (LRA) Method

LRA is a statistical technique used which identifies the optimal line that best fits the data by determining the regression vector parameters  $x$  and  $y$  [41]. Linear regression is a modeling technique that assesses the relationship between a single independent variable and a dependent variable, employing a straight line for this purpose [42]. Both variables need to be quantitative in nature. Simple linear regression is considered a parametric test, which means it operates based on specific assumptions about the dataset [43,44]. These underlying assumptions include (1) Homoscedasticity: This condition assumes that the error magnitude in these predictions remains relatively consistent across the range of

values of the independent variable. In other words, the variability of errors should be uniform [45] and (2) Reliable Data Gathering Techniques: It is assumed that the observations in the dataset were collected using trustworthy and well-established statistical sampling methods [46].

In the study, the focus was on the 11-month data gap between the GRACE and GRACE-FO missions, spanning from July 2017 to May 2018. To estimate the missing observations, Linear Regression Analysis was applied. These assumptions are deemed appropriate for the GRACE data. Regression models can be used to forecast the dependent variable's value at specific values of the independent variable. However, it is essential to note that this method is applicable only within the range of observed data, and it cannot be used to predict values for future observation months.

#### 2.4. Gap-Filling Model Using LRA

Here, assume that  $l_i$  is the dependent variable, the observation vector  $l_i = (l_{i1}, l_{i2}, \dots, l_{in})$  for each existing month, and  $i = 1, \dots, m$ ,  $m$  is the number of available GRACE and GRACE-FO observation months (in this study  $m = 17$ ), while  $Q_i$  is the independent variable which represents the month's observation order through GRACE and GRACE-FO processing (i.e.,  $Q_i = i$ ). Here, estimation of the value for each missing month individually using the existing observation for the same month annually (all Januarys, all Februarys . . . , etc.) has been carried out. For instance, the  $Q_1$  value for the January 2003 column is equal 1, while  $Q_4$  for January 2006 is equal 4. The reason for this assumption is that the annual monthly observations have about the same climate conditions like rainfall and temperature. So, the linear regression equation will be in the following form:

$$\begin{aligned} l_1 &= x + yQ_1 + \varepsilon_1 \\ l_2 &= x + yQ_2 + \varepsilon_2 \\ &\vdots \\ l_m &= x + yQ_m + \varepsilon_m \end{aligned} \quad (3)$$

where  $x$  and  $y$  are the required regression coefficient vectors, when  $x = (x_1, x_2, \dots, x_n)$  and  $y = (y_1, y_2, \dots, y_n)$ , respectively, and  $n$  is the grid point number for every month ( $180 \times 360 = 64,800$  grids in  $1^\circ$  spatial resolution). In the case of SHCs,  $n$  is the number of spherical harmonic coefficients for each  $C_{nm}$  and  $S_{nm}$ . Specifically, if  $N$  represents the degree, then  $n$  can be calculated as  $n = (N + 1) \times (N + 2)/2$ .  $\varepsilon$  is a random error matrix.

Using the existing observations, the monthly  $x$  and  $y$  vectors are estimated, and their values are then used to estimate the missing months. It is assumed that  $L$  is the observation grid of the month, so the simple linear regression (SLR) form will be as follows:

$$\begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_m \end{bmatrix}_{m \times n} = \begin{bmatrix} x + yQ_1 \\ x + yQ_2 \\ \vdots \\ x + yQ_m \end{bmatrix}_{m \times n} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}_{m \times n} \quad (4)$$

$$\begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_m \end{bmatrix}_{m \times n} = \begin{bmatrix} 1 & Q_1 \\ 1 & Q_2 \\ \vdots & \vdots \\ 1 & Q_m \end{bmatrix}_{m \times 2} \beta_{2 \times n} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}_{m \times n} \quad (5)$$

where the design matrix is:

$$T_{m \times 2} = \begin{bmatrix} 1 & Q_1 \\ 1 & Q_2 \\ \vdots & \vdots \\ 1 & Q_m \end{bmatrix}_{m \times 2} \quad (6)$$

the vector of regression coefficients is expressed as

$$\beta_{2 \times n} = \begin{bmatrix} x \\ y \end{bmatrix}_{2 \times n} \quad (7)$$

and the vector of error is expressed as:

$$\varepsilon_{m \times n} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}_{m \times n} \quad (8)$$

Expressing the vector of regression observations as

$$L_{m \times n} = \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_m \end{bmatrix}_{m \times n} \quad (9)$$

then

$$L = T\beta + \varepsilon \quad (10)$$

solving Equation (10) based on least squared principle,  $\beta$  is as follows:

$$\beta = (T^T W T)^{-1} T^T W L \quad (11)$$

$W$  is the weight matrix that adjusts the contribution of each observation to the parameter estimation. The goal of  $W$  is to give more weight to observations with smaller variances.  $W$  is an  $m \times m$  matrix with  $(w_1, w_2, \dots, w_n)$  on the diagonal and 0 elsewhere.

$$W_{m \times m} = \begin{bmatrix} w_1 & & & \\ & w_1 & & \\ & & \ddots & \\ & & & w_m \end{bmatrix} \quad (12)$$

Then, to solve Equation (11),  $T^T T$  is as follows:

$$T^T W T = \begin{bmatrix} 1 & 1 & \dots & 1 \\ Q_1 & Q_2 & \dots & Q_m \end{bmatrix} \begin{bmatrix} w_1 & & & \\ & w_1 & & \\ & & \ddots & \\ & & & w_m \end{bmatrix} \begin{bmatrix} 1 & Q \\ 1 & Q_2 \\ \vdots & \vdots \\ 1 & Q_m \end{bmatrix} = \begin{bmatrix} m & \sum_{i=1}^m Q_i \\ \sum_{i=1}^m Q_i & \sum_{i=1}^m Q_i^2 \end{bmatrix} \quad (13)$$

$$(Q^T Q)^{-1} = \frac{1}{\det(Q^T Q)} \begin{bmatrix} \sum_{i=1}^m Q_i^2 & -\sum_{i=1}^m Q_i \\ -\sum_{i=1}^m Q_i & M \end{bmatrix} \quad (14)$$

$$(T^T T)^{-1} = \frac{1}{M \sum_{i=1}^m Q_i^2 - (\sum_{i=1}^m Q_i)^2} \begin{bmatrix} \sum_{i=1}^m Q_i^2 & -\sum_{i=1}^m Q_i \\ -\sum_{i=1}^m Q_i & M \end{bmatrix} \quad (15)$$

$$T^T L = \begin{bmatrix} \sum_{i=1}^m l_{(j,i)} \\ \sum_{i=1}^m Q_i l_{(j,i)} \end{bmatrix} \quad (16)$$

where  $j = 1, \dots, n$ . So, from Equation (11),  $\beta$  will be as follows:

$$\beta = \frac{1}{M \sum_{i=1}^m Q_i^2 - \left(\sum_{i=1}^m Q_i\right)^2} \begin{bmatrix} \sum_{i=1}^m Q_i^2 & -\sum_{i=1}^m Q_i \\ -\sum_{i=1}^m Q_i & M \end{bmatrix} \begin{bmatrix} \sum_{i=1}^m Q_i \\ \sum_{i=1}^m Q_i l_{(j,i)} \end{bmatrix} \quad (17)$$

$$\beta = \frac{1}{M \sum_{i=1}^m Q_i^2 - \left(\sum_{i=1}^m Q_i\right)^2} \begin{bmatrix} \sum_{i=1}^m Q_i^2 \sum_{i=1}^m l_{(j,i)} - \sum_{i=1}^m Q_i \sum_{i=1}^m Q_i l_{(j,i)} \\ M \sum_{i=1}^m Q_i l_{(j,i)} - \sum_{i=1}^m Q_i \sum_{i=1}^m l_{(j,i)} \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} \quad (18)$$

Then, the values of  $x_j$  and  $y_j$  will be as follows:

$$x_j = \frac{\sum_{i=1}^m Q_i^2 \sum_{i=1}^m l_{(j,i)} - \sum_{i=1}^m Q_i \sum_{i=1}^m Q_i l_{(j,i)}}{M \sum_{i=1}^m Q_i^2 - \left(\sum_{i=1}^m Q_i\right)^2} \quad (19)$$

$$y_j = \frac{M \sum_{i=1}^m Q_i l_{(j,i)} - \sum_{i=1}^m Q_i \sum_{j=1}^m l_{(j,i)}}{M \sum_{i=1}^m Q_i^2 - \left(\sum_{i=1}^m Q_i\right)^2} \quad (20)$$

After estimating the  $x$  and  $y$  vectors for every month, the missing month can be estimated by substituting the values of the  $x$  and  $y$  vectors in Equation (21) as follows:

$$H_{n \times 1} = x_{n \times 1} + y_{n \times 1} O \quad (21)$$

where  $H$  is the predicted vector for the missing month and  $O$  is the order of this month.

### 3. Comparison between LRA Model and Other Filling-In Models

To validate the accuracy of the model, the gravity models discussed in Section 2.2 are employed. To provide an overview of the SHCs using these gravity models, it is important to note that they can be broadly categorized into four distinct groups for the purpose of gap-filling datasets: (1) Datasets derived from other observations but with limited resolutions. An example is the time-varying gravity field models obtained by combining high–low satellite-to-satellite tracking (HLSST) and SLR techniques, as presented in the study by QF, Weigelt, 2019 [40]. (2) Fully data-driven approaches that rely on interpolations of GRACE observations. One such approach is the use of SSA for filling-in the gaps in GRACE data, as demonstrated in the research by Yi and Sneeuw, 2019 [16]. (3) Hybrid approaches that integrate data from SLR or other sources with GRACE. For instance, a hybrid modeling technique utilizes EOFs from GRACE solutions as base functions to recover time-varying gravity field models from SLR observations. This approach is described in the study conducted by IGG, Löcher and Kusche, 2021 [39]. (4) In addition to these three types of datasets, there are also data available from the Swarm satellite mission, processed by two different centers, ASU and COST-G [18,47].

## 4. Results and Discussion

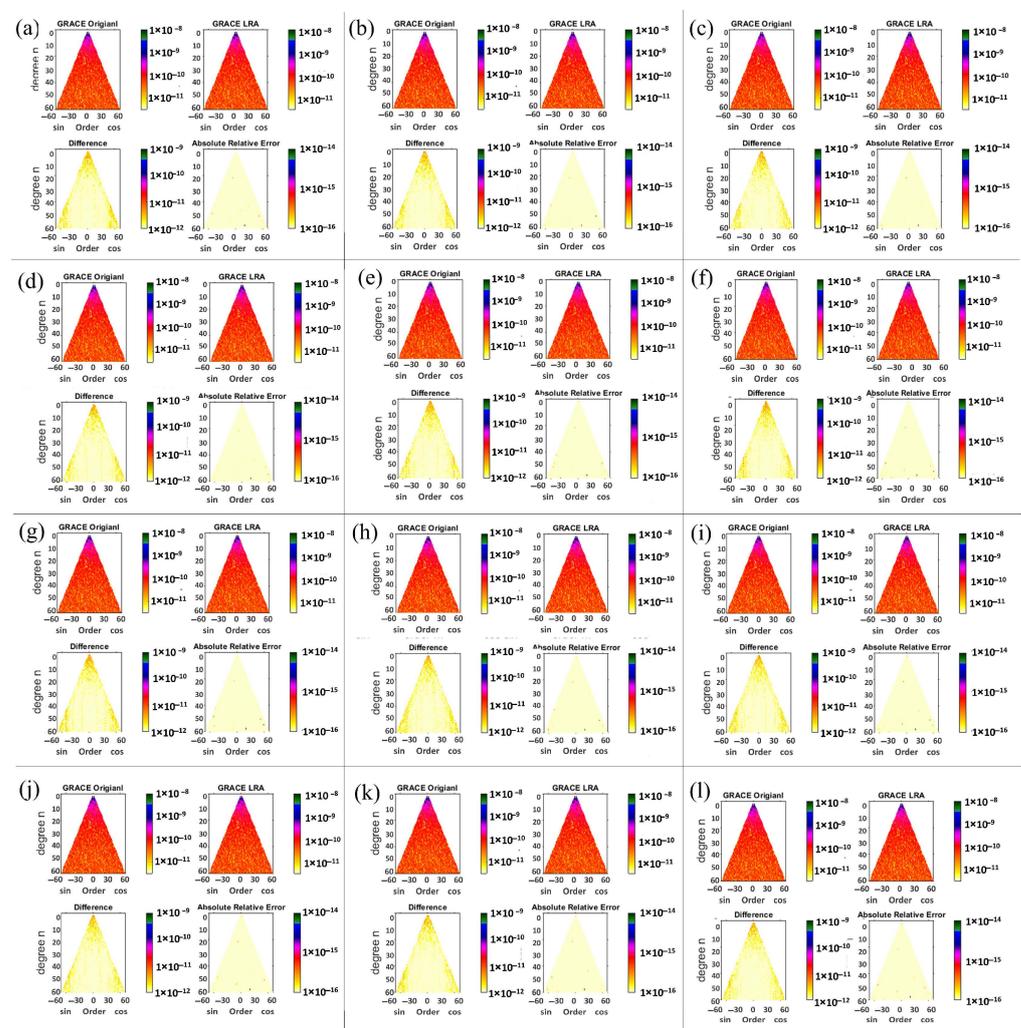
### 4.1. Spectral Domain: Artificial Gap in GRACE Era

#### 4.1.1. GRACE and LRA SHCs

To rigorously validate the LRA-based gap-filling approach in the spectral domain, an artificial data gap from January 2009 to December 2009 was intentionally created. SHCs were estimated within this period and compared directly with the original GRACE values for robust validation.

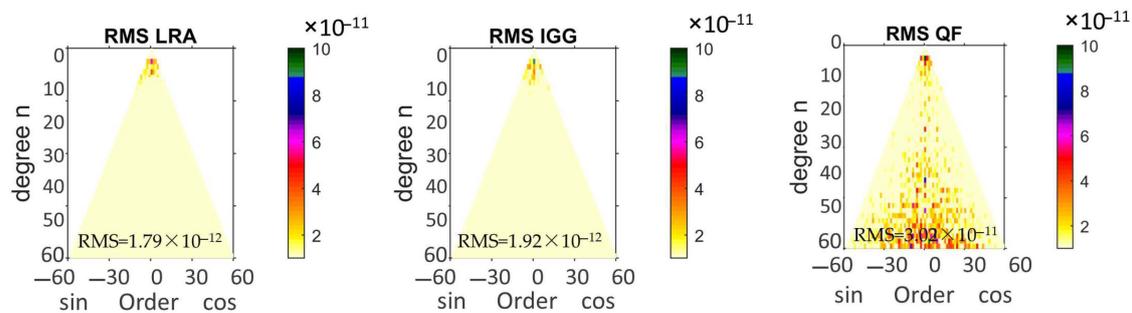
In these following sections, the five datasets mentioned above (three models + two Swarm) will be thoroughly examined in terms of their spectral characteristics. Their consistency will be evaluated, and their suitability for specific scenarios will be explored using an LRA model.

As depicted in Figure 3, the illustration presents a detailed comparison between the original GRACE-CSR SHCs and the coefficients generated by the LRA model for each month of the year 2009, denoted from (a) to (l). The figure is segmented into twelve boxes, one for each month, where the original month's SHCs are on the upper-left side and the corresponding LRA-generated coefficients are on the upper-right side. The subtraction of these values results in the data shown in the lower-left side of each box, illustrating the discrepancies between the GRACE and LRA SHCs. The observed disparities between the two datasets are minimal, demonstrating a high level of consistency between the GRACE and LRA model. Additionally, the absolute relative error for each month was calculated, and the results are displayed in the lower-right side of each sub-plot. Remarkably, the absolute relative error is nearly negligible across the board. This outcome underscores the remarkable accuracy of this model in effectively filling the data gaps.



**Figure 3.** The SHCs of GRACE-CSR and the LRA model for the period from January 2009 to December 2009 are presented in this figure. The figure is divided into 12 subfigures, labeled from (a–l), with each subfigure representing a specific month. Within each subfigure, four figures are shown in a clockwise arrangement. The upper-left figure displays the original GRACE-CSR SHCs, followed by the LRA SHCs in the virtual gap. The next figure shows the absolute relative error, and the final figure depicts the differences between the two sets of coefficients.

We evaluated the root mean square (RMS) for spherical harmonic coefficients (SHCs) up to degree/order 60 to validate correlations between GRACE and LRA, IGG, and QF. As depicted in Figure 4, during an artificial gap in 2009, we calculated the RMS. Notably, GRACE and LRA exhibited the smallest RMS ( $1.79 \times 10^{-12}$ ), indicating significant agreement. The IGG dataset, when compared to GRACE, showed a modest RMS of  $1.92 \times 10^{-12}$ . Conversely, the QF data had the lowest RMS ( $3.02 \times 10^{-11}$ ). The strong correlation between GRACE and LRA underscores their agreement.



**Figure 4.** The RMS between GRACE and LRA, IGG, and QF during an artificial gap in 2009.

#### 4.1.2. Validation between LRA, GRACE, QF, and IGG

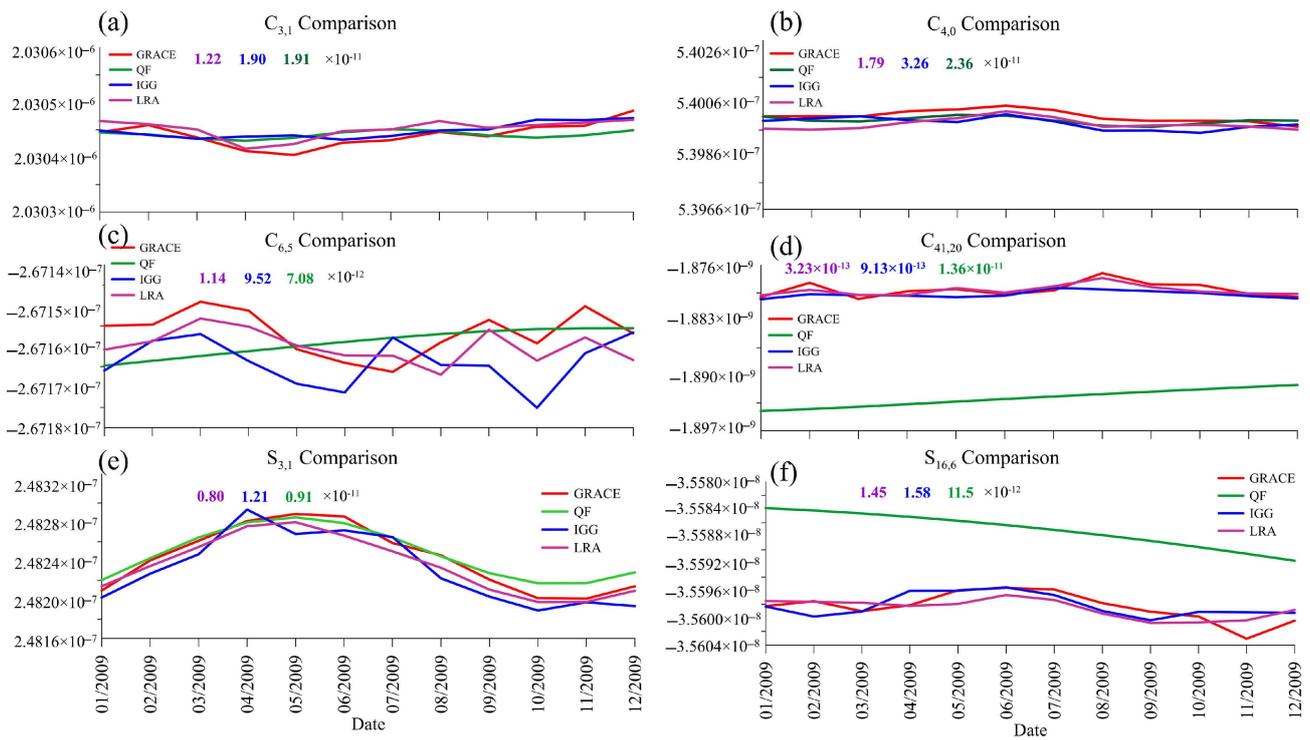
In this comprehensive analysis, a time-series examination from January 2009 to December 2009 was undertaken. The objective was a meticulous comparison of datasets, including GRACE CSR, QF, IGG, and LRA-derived SHCs. This comparative assessment encompassed the scrutiny of a diverse set of SHCs, ranging from low to high degrees, such as  $C_{3,1}$ ,  $C_{4,0}$ ,  $C_{6,5}$ ,  $C_{41,20}$ ,  $S_{3,1}$ , and  $S_{16,6}$ .

As delineated in Figure 5, the time-series data visually present the outcomes of this rigorous comparison, revealing patterns of consistency or inconsistency. The correlation coefficient values between each dataset and the LRA-derived coefficients are prominently displayed at the top of each subfigure, representing GRACE CSR, QF, IGG, and LRA SHCs, respectively. The overarching findings indicate a noteworthy level of overall consistency among the datasets. Particularly striking is the high degree of coherence observed between the IGG and LRA SHCs with the GRACE CSR coefficients, in contrast to the QF model. This discrepancy with the QF model may arise from the significant reliance of the IGG data on GRACE/GRACE-FO sources. At lower degrees (up to degree 4), all datasets exhibit similar behavior. However, beyond this degree, they generally exhibit comparable patterns, except for the QF coefficients. Notably, in the case of  $C_{41,20}$  and  $S_{16,6}$ , the QF time-series shows similar averages to other datasets but distinct behavior.

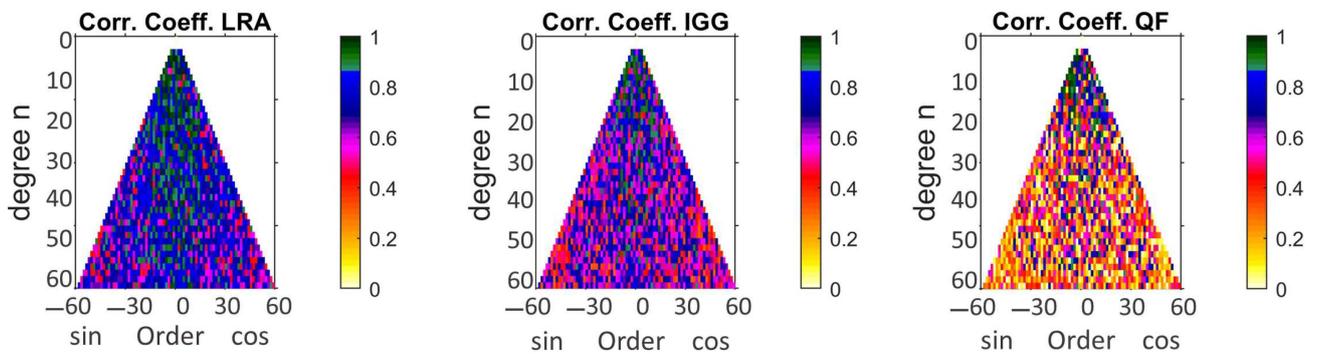
In contrast, the LRA gap-filling approach exhibits higher consistency with the GRACE coefficients across various SHCs, as evident in the correlation coefficients. For instance, in  $S_{31}$  and  $C_{31}$ , the correlation coefficients reach 0.99 and 0.89 between GRACE and LRA, compared to 0.98 and 0.75 for IGG and 0.95 and 0.45 for QF. Similarly, for  $C_{4,0}$ ,  $S_{16,6}$ , and  $C_{41,20}$ , LRA demonstrates greater consistency with correlation coefficients of 0.74, 0.72, and 0.69, respectively, compared to other models with coefficients of 0.64, 0.66, and 0.66, respectively.

We computed these correlation coefficients for the entire spherical harmonic coefficients (SHCs) up to degree/order 60 in order to validate the correlations between GRACE and LRA, IGG, and QF. In 2009, we calculated the correlation coefficients during an artificial gap.

According to Figure 6 and Table 2, GRACE and LRA showed the greatest correlation coefficient (0.8258), indicating significant agreement. With GRACE, the IGG dataset showed a modest correlation coefficient of 0.6763. The QF data, on the other hand, had the lowest correlation coefficient, at 0.5428. The high correlation between GRACE and LRA suggests strong agreement between these two datasets.



**Figure 5.** A time-series comparison is conducted among four types of data: GRACE-CSR, QF, IGG, and LRA. The figure consists of six subfigures, each focusing on a specific SHC comparison. Subfigure (a) compares  $C_{3,1}$ , subfigure (b)  $C_{4,0}$ , subfigure (c)  $C_{6,5}$ , subfigure (d)  $C_{41,20}$ , subfigure (e)  $S_{3,1}$ , and subfigure (f) compares  $S_{16,6}$ . The correlation coefficient values between GRACE-CSR, QF, and IGG with LRA SHCs are provided above each corresponding subfigure.



**Figure 6.** Correlation coefficients between GRACE and geodetic data sources (LRA, IGG, and QF) during an artificial gap in 2009.

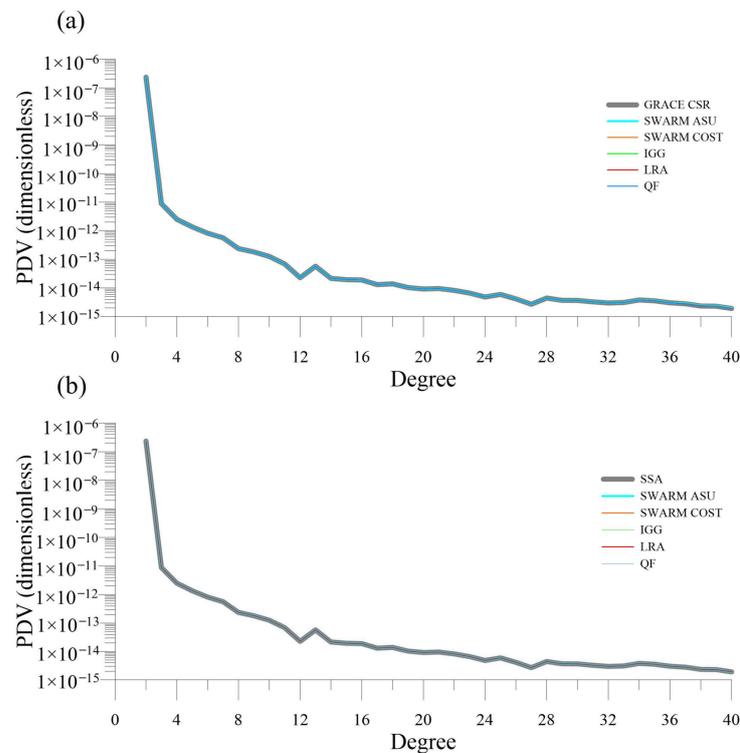
**Table 2.** Correlation coefficients between GRACE and LRA, IGG, and QF.

Source	Correlation Coefficients
LRA	0.8258
IGG	0.6763
QF	0.5428

#### 4.2. Potential Degree Variance (PDV)

In December 2013, the Swarm mission released products, which was a significant development. Coefficients for that month were estimated using the LRA approach. A comparative assessment included GRACE-CSR, Swarm (ASU and COST-G), IGG, QF, and

the LRA gap-filling model. The evaluation focused on PDV. For a fair comparison, all coefficients for all datasets were standardized by truncating them up to degree 40, aligning with Swarm products. Figure 7a illustrates the results of this comparison, revealing that the PDVs for all products, including the LRA model, exhibit a high degree of consistency with one another.



**Figure 7.** Potential degree variance (PDV). (a) PDV in December 2013 using GRACE CSR, Swarm ASU and Swarm COST-G, IGG, LRA, and QF, (b) PDV in July 2014 using SSA, Swarm ASU and Swarm COST-G, IGG, LRA, and QF.

We opted to analyze a specific month, July 2014, which featured a data gap in GRACE but also included data from the Swarm mission. In this analysis, the LRA gap-filling approach was employed to estimate missing observations for this month. Subsequently, a comparative assessment was conducted, contrasting this estimated data with datasets obtained from Swarm (ASU and COST-G), IGG, QF, and SSA data, as documented in the work by Yi and Sneeuw [16].

As portrayed in Figure 7b, this comparison was focused on assessing the potential degree variance (PDV) among these datasets, with all coefficients truncated up to degree 40 to align with Swarm data standards. The outcomes of this evaluation underscore a remarkable level of consistency among all products, including the LRA model, in terms of PDV. In December 2013 (as shown in Figure 7a), the aim was to determine which product closely aligns with GRACE values. To achieve this, the variance for the PDV differences between GRACE-CSR and the other products was calculated. The resulting variances were as follows: Swarm ASU ( $4.47 \times 10^{-25}$ ), Swarm COST-G ( $4.35 \times 10^{-25}$ ), IGG ( $1.65 \times 10^{-29}$ ), LRA ( $9.27 \times 10^{-34}$ ), and QF ( $1.22 \times 10^{-29}$ ). Among these, the LRA gap-filling approach exhibited the smallest variance, with a value of  $9.27 \times 10^{-34}$ .

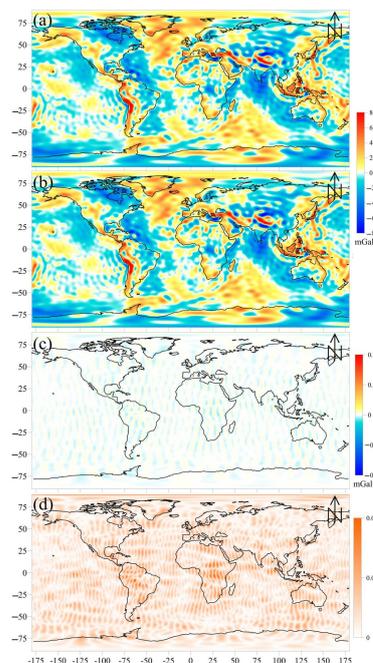
#### 4.3. Gravity Anomaly (GA)

In geodesy, gravity anomalies are used to define the figure of the Earth, notably the geoid (the equipotential surface of the Earth's gravity field that corresponds most closely to the mean sea level). They are used in geodesy in three main areas: (1) to determine the anomalous potential  $T$ , the geoid's undulation  $N$ , and the vertical deflections

globally and locally [48]; (2) to determine the vertical crustal deformation by studying the variation in the gravitational field over time [49]; and (3) to adjust various geodetic readings, including leveling observations [50]. The GA or  $\Delta g$  is expressed as given by Heiskanen and Moritz [48]:

$$\Delta g(r, \varphi, \lambda) = \frac{km}{r^2} \sum_{n=2}^{\infty} (n-1) \left(\frac{a}{r}\right)^n \sum_{m=0}^n (\bar{C}_{nm} \cos(m\lambda) + \bar{S}_{nm} \sin(m\lambda)) \bar{P}_{nm}(\sin \varphi) \quad (22)$$

where  $\Delta g$  is the gravity anomaly measured by mGal,  $km$  is the geocentric gravitational constant of the Earth,  $r$  is the geocentric radius,  $\varphi$  is the geocentric latitude,  $\lambda$  is the geodetic longitude,  $a$  is the equatorial radius, and  $\bar{P}_{nm}(\sin \varphi)$  is the fully normalized associated Legendre function. The GA for the same month for both the GRACE-CSR and LRA gap-filling approach shows consistency and agreement between them (Figure 8a,b). Figure 8c shows the difference between the global GA for both the GRACE-CSR and LRA gap-filling approach, while Figure 8d shows the RMS. The difference map (Figure 8c) illustrates that the LRA method closely approximates GRACE-CSR, exhibiting minimal discrepancies. Additionally, the RMS value (Figure 8d) indicates low variability, emphasizing the high accuracy of LRA-based gravity anomaly predictions. These findings underscore the reliability of the LRA method for predicting gravity anomalies.



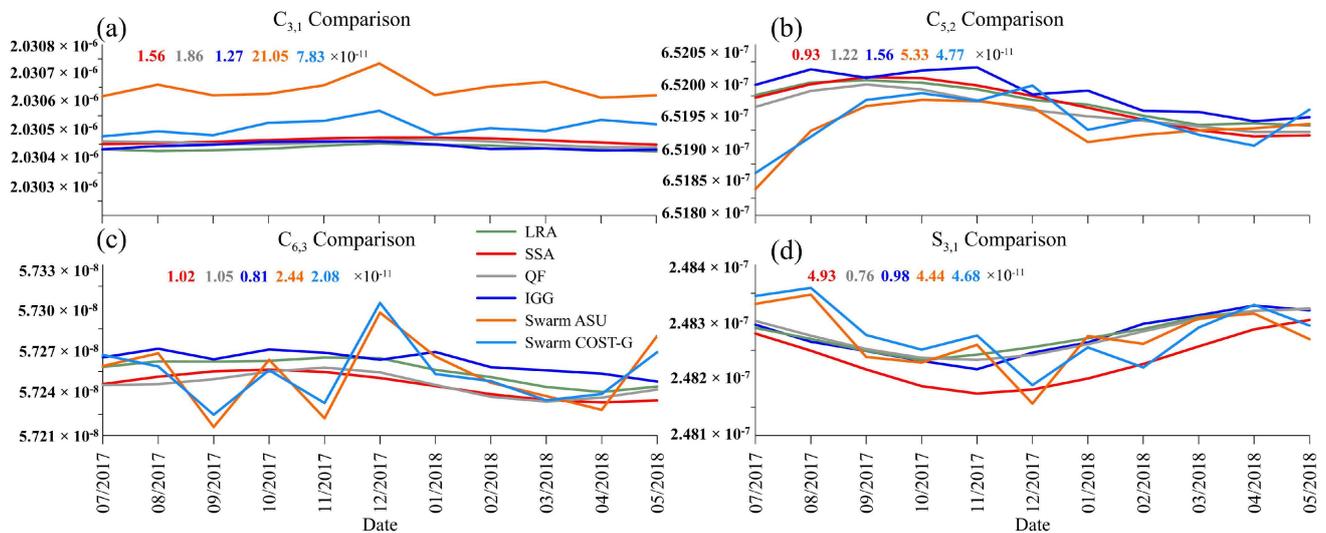
**Figure 8.** Gravity anomaly in December 2013. The gravity anomalies for GRACE-CSR- and LRA-derived SHCs are shown in (a,b), respectively, while the difference between them is shown in (c). In addition, the RMS is shown in (d).

#### 4.4. Validation between LRA and GRACE, QF, IGG, and Swarm (GRACE Gap)

Following the validation process, involving comparing the effectiveness of the LRA gap-filling approach with both the GRACE data and other gap-filling models during the GRACE era, the LRA gap-filling model has been employed to bridge the data gap that occurred between GRACE and GRACE-FO from July 2017 to May 2018.

In this phase of the study, a thorough and extensive comparison was conducted, encompassing the LRA-derived SHCs alongside datasets from IGG, QF, and Swarm mission data (ASU and COST-G) and SSA for data filling purposes. Figure 8 illustrates a time-series comparison that involves the examination of random SHCs, including  $C_{3,1}$ ,  $C_{5,2}$ ,  $C_{6,3}$ , and  $S_{3,1}$ . The outcomes of this analysis are presented in Figure 9, offering a visualization of the time-series data and highlighting instances of consistent or inconsistent series comparisons.

At the top of each subfigure within the figure, the correlation coefficient values between each dataset and the LRA-derived SHCs are prominently displayed. These subfigures represent SSA, QF, IGG, Swarm ASU, and Swarm-COST-G with LRA, respectively.

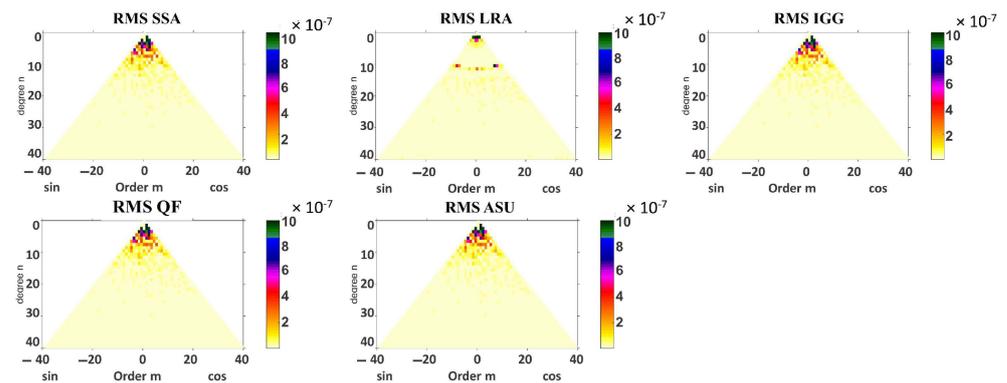


**Figure 9.** A time-series comparison is conducted among four types of data: SSA, QF, IGG, Swarm (ASU and COST-G), and LRA. The figure consists of four subfigures, each focusing on a specific SHC comparison. Subfigure (a) compares  $C_{3,1}$ , subfigure (b)  $C_{5,2}$ , subfigure (c)  $C_{6,3}$ , and subfigure (d)  $S_{3,1}$ . The correlation coefficient values between SSA, QF, IGG, Swarm ASU, and Swarm-COST-G with LRA SHCs are provided above each corresponding subfigure.

Despite some higher volatility observed in the Swarm SHCs' time-series compared to other datasets; Swarm generally demonstrates a noteworthy degree of consistency on average. It is noteworthy that a higher level of consistency is observed between SSA, QF, and IGG with LRA SHCs when compared to the Swarm data. This could potentially be attributed to the SSA and IGG data having a substantial derivation from GRACE/GRACE-FO sources.

The correlation coefficient values consistently exceed 0.89 for all coefficients between SSA and LRA, indicating a maximum level of compatibility since both datasets are extracted from GRACE. Conversely, in the case of IGG and QF, the correlation coefficient consistently exceeds 0.86 for all SHCs except  $C_{3,1}$ , suggesting a considerable degree of association with the GRACE data.

In this analysis, we estimate the RMS values for several gravity field models during the GRACE/GRACE-FO gap from July 2017 to May 2018 (Figure 10). The following models are considered: LRA, SSA, IGG, QF, and ASU. LRA exhibits the smallest RMS values across all degrees, particularly in the lowest degrees (up to degree-10). SSA shows a more dispersed pattern of RMS values. IGG and QF exhibit higher RMS values compared to LRA. ASU demonstrates a spread similar to SSA but denser. The RMS values for ASU are generally higher than those for LRA.

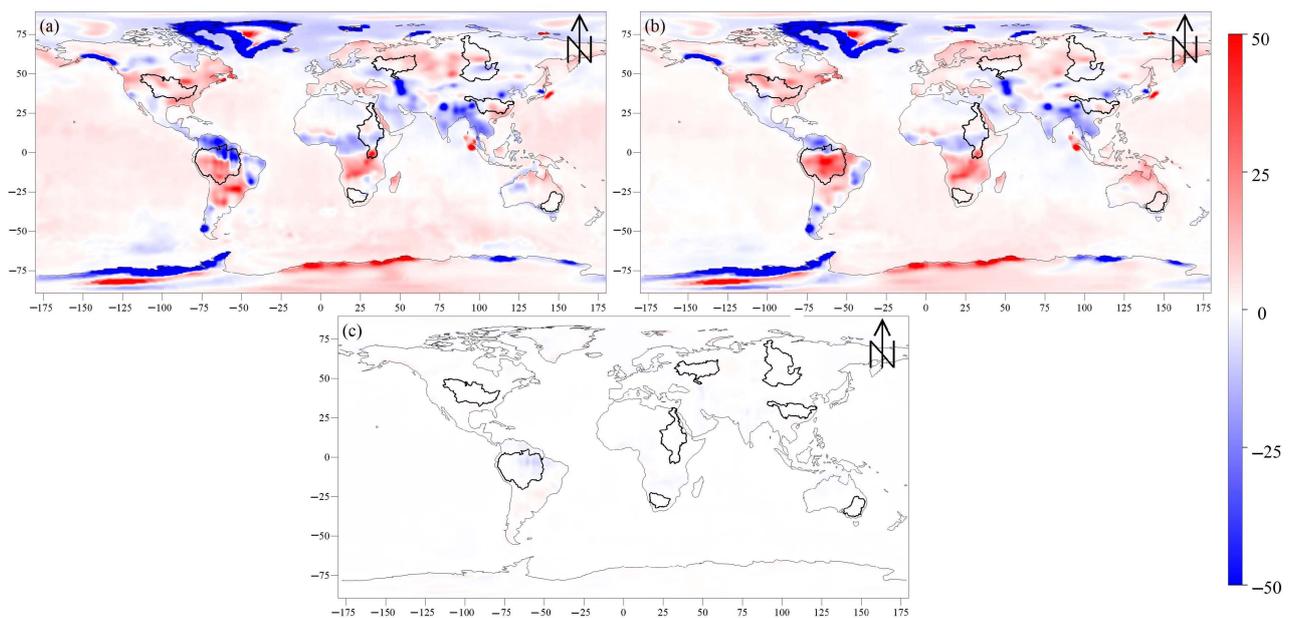


**Figure 10.** The RMS for 5 SHC products, SSA, LRA, IGG, QF, and ASU, as triangles during the GRACE/GRACE-FO gap from July 2017 to May 2018.

## 5. Mascon Data Using LRA Gap-Filling Model

### 5.1. Mass Change in the Selected Basins

The LRA gap-filling model is versatile and can also be applied to the estimation of Equivalent Water Height (EWH) as a mascon data type. To evaluate the performance of the gap-filling approach, assessments were conducted in eight well-distributed geographical regions, as illustrated in Figure 11. These eight basins span across all continents.



**Figure 11.** Predictions and observations. (a) The GRACE-derived TWS in January 2016, (b) the predicted TWS in the same month using LRA, (c) the difference between both of them.

To validate the effectiveness of the method in the context of EWH estimation, a virtual data gap was simulated during the GRACE era, specifically in January 2016. During this simulation, the LRA-derived EWH was employed and compared with GRACE-EWH obtained from mascon data. The results, depicted in Figure 11, showcase the original GRACE data alongside the EWH values generated using the LRA gap-filling approach, along with the differences between them. It is notable that while there may be slightly larger discrepancies in the Amazon basin compared to other regions, the differences in a global context tend to approach minimum values.

Additionally, Table 3 presents the standard deviation values for all eight basins, facilitating a comparative analysis. These standard deviation values demonstrate minimal

differences between the results obtained using GRACE data and the LRA gap-filling approach for all basins.

**Table 3.** The TWS standard deviation for the 8 basins in January 2016 using GRACE, LRA, and the subtraction results of GRACE-LRA.

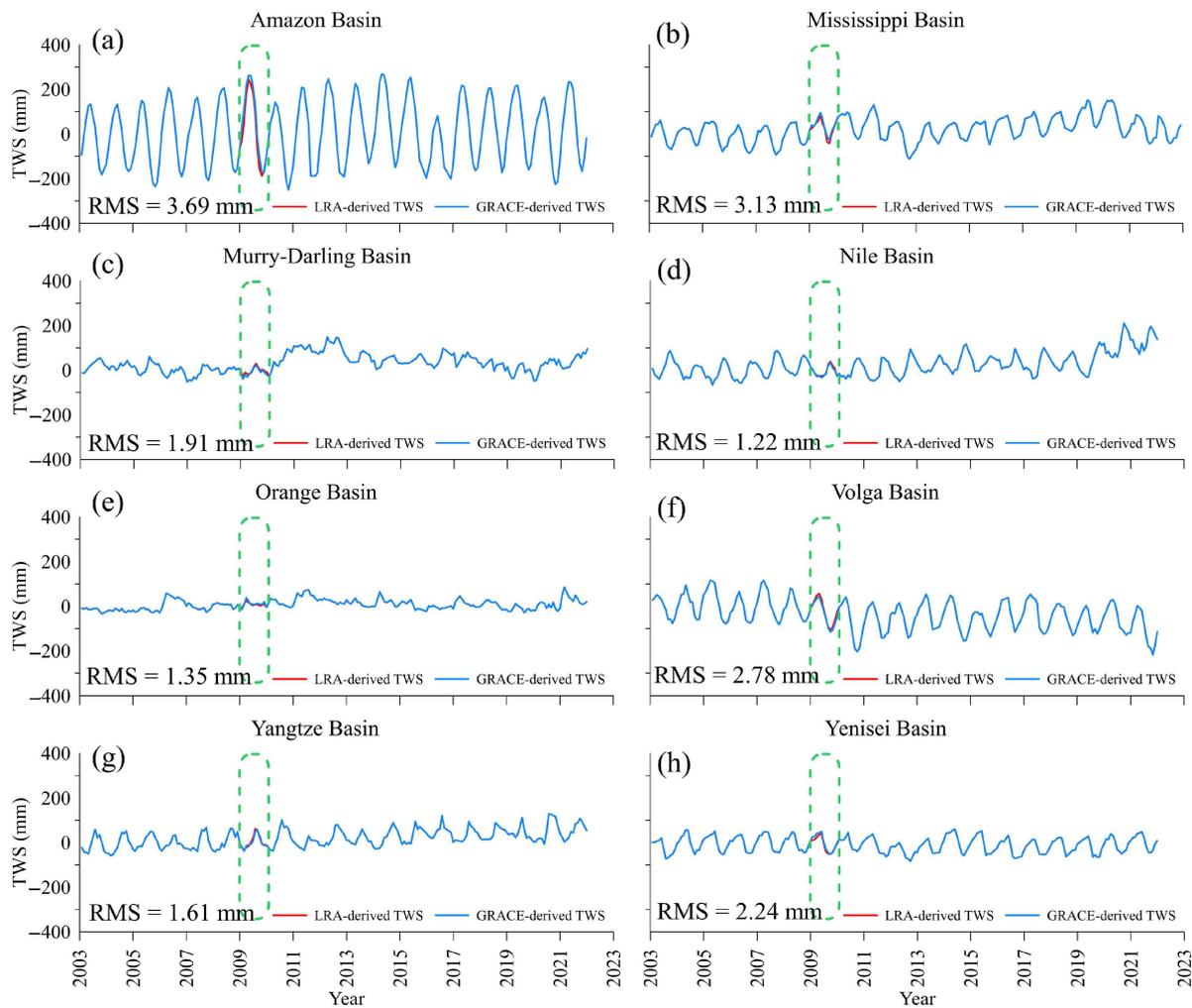
Basin	Terrestrial Water Storage (mm)		
	GRACE	LRA	GRACE-LRA
Amazon	5.53	6.48	5.14
Mississippi	1.18	1.53	0.78
Murry-Darling	0.42	0.64	0.45
Nile	1.82	1.34	0.64
Orange	0.41	0.54	0.58
Volga	1.69	1.05	0.96
Yangtze	1.62	0.97	1.05
Yenisei	1.31	1.08	0.65

Furthermore, as part of the analysis, an artificial data gap was deliberately introduced into the GRACE dataset, specifically during the year 2009. We analyzed the TWS time-series for eight river basins and computed the root mean square (RMS) difference between GRACE and the LRA (Figure 12). The RMS values, which reflect variability, remained small across all basins relative to the TWS range. Specifically, the Amazon basin exhibits significant seasonal fluctuations, with an RMS of 3.69 mm, indicating high variability. The Mississippi basin shows fluctuations that are less pronounced than the Amazon, with an RMS of 3.13 mm. In the Murray Darling, Orange, and Yangtze basins, we observe moderate fluctuations, with RMS values below 2 mm. The Volga and Nile basins exhibit significant fluctuations over time, with RMS values of 2.78 mm and 1.22 mm, respectively. Finally, the Yenisei basin shows present but less pronounced fluctuations compared to the Amazon or Volga basin, with an RMS of 2.24 mm.

The LRA gap-filling approach was then utilized to estimate the EWH for each of the 12 months within that year. Subsequently, the mean ( $M$ ), standard deviation ( $\sigma$ ), and variance for the EWH values during this period were calculated. Remarkably, when comparing the mean EWH values obtained from both the GRACE dataset and the LRA approach for all months in 2009 (Table 4), a high degree of similarity was observed. This consistency also extended to the standard deviation and variance values, which exhibited matching characteristics. This convergence of numerical results strongly indicates that the LRA approach is a reliable method for filling gaps in GRACE data during the GRACE era.

**Table 4.** The global mean, standard deviation, and variance values for EWH computed for each month throughout the year 2009. The values are presented for both the GRACE CSR mascon data and the EWH derived through the LRA gap-filling approach.

Month	Mean ( $M$ )		Standard Deviation ( $\sigma$ )		Variance ( $1/\sigma^2$ )	
	GRACE	LRA	GRACE	LRA	GRACE	LRA
January	−0.653	−0.701	6.415	6.869	0.0243	0.0212
February	−0.447	−0.526	6.817	7.071	0.0215	0.0200
March	−0.580	−0.483	7.618	7.920	0.0172	0.0159
April	−0.396	−0.348	7.819	7.593	0.0164	0.0173
May	0.018	0.164	8.098	8.117	0.0152	0.0152
June	0.315	0.316	7.819	8.070	0.0164	0.0154
July	0.126	−0.089	7.902	8.042	0.0160	0.0155
August	−0.134	−0.184	9.149	9.281	0.0119	0.0116
September	−0.290	−0.217	9.698	9.545	0.0106	0.0110
October	−0.225	−0.506	9.666	9.441	0.0107	0.0112
November	−0.556	−0.564	9.542	9.158	0.0110	0.0119
December	−0.707	−0.734	8.949	9.218	0.0125	0.0118



**Figure 12.** Time-series of total water storage (TWS) for each river basin, including the root mean square (RMS). The green rectangle denotes an artificial gap. (a) Amazon, (b) Mississippi, (c) Murray-Darling, (d) Nile, (e) Orange, (f) Volga, (g) Yangtze, and (h) Yenisei River basins.

Then, the weighted mean  $\bar{X}_w$  for the estimated EWH values during 2009 was estimated, along with the standard deviation of the weighted mean  $S_{\bar{M}}$ . The weighted mean and standard deviation of the weighted mean can be calculated as follows [51]:

$$\bar{X}_w = \frac{\sum w\bar{X}}{\sum w} \quad (23)$$

$$S_{\bar{M}} = \sqrt{\frac{\sum wv^2}{(n-1)\sum w}} \quad (24)$$

where  $w$  is the variance ( $1/\sigma^2$ ), and  $n$  is the number of months. The residuals,  $v$ , can be calculated as follows:

$$v = \bar{X}_w - \bar{X} \quad (25)$$

The weighted mean provides a way to calculate the average when different data points have varying levels of importance or significance, while the standard deviation represents the measure of the variability or spread of the weighted mean [52]. Table 5 presents the weighted mean  $\bar{X}_w$  and standard deviation of the weighted mean  $S_{\bar{M}}$  during the year 2009. Notably, while the  $\bar{X}_w$  value was  $-0.308$  when calculated using GRACE data, it slightly differed at  $-0.329$  when computed using the LRA gap-filling approach. Similarly, for

$S_{\overline{M}}$ , it registered at 1.047 with GRACE but increased to 1.149 when employing the LRA gap-filling approach.

**Table 5.** The global weighted mean and standard deviation of the weighted mean during 2009 comparing results obtained from GRACE CSR mascon data with those from the LRA gap-filling approach.

Source	Weighted Mean	Standard Deviation of the Weighted Mean
GRACE/GRACE-FO	−0.308	1.047
LRA	−0.329	1.149

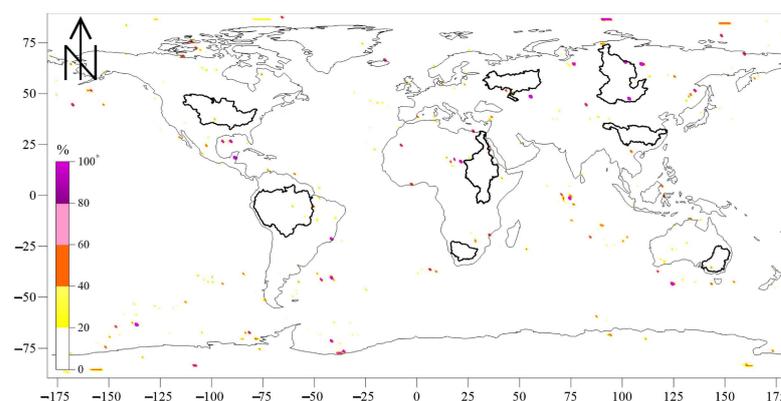
The standard deviation of the weighted mean offers insights into the expected variation in the weighted mean across diverse datasets. The convergence of these values between GRACE and the LRA approach underscores the convenience and suitability of this approach as a replacement for GRACE during data gaps.

### 5.2. Mean Percentage Error (MPE) and Coefficient of Determination ( $R^2$ )

To assess the model, various error metrics were calculated here to quantify the differences between the LRA gap-filling model and the GRACE observations. The estimated error metrics include the Mean Percentage Error (MPE) and R-squared (coefficient of determination), which measures the proportion of variance in the observed data that is explained by the model's predictions. The Mean Percentage Error (MPE) is used to calculate the percentage difference between each observed value and its corresponding predicted value, while R-squared measures the proportion of variance in the observed values that can be explained by the predicted values. It ranges from 0 to 1, where higher values indicate a better fit. The Mean Percentage Error (MPE) can be estimated as follows:

$$MPE = \frac{\sum [\text{observed}(GRACE) - \text{predicted}(LRA)]}{\sum [\text{observed}(GRACE)]} \times 100 \quad (26)$$

Choosing March 2016 as a random new virtual gap, the EWH observations were estimated using the LRA approach during this month and compared with GRACE-EWH observations. Figure 13 illustrates the MPE globally, indicating that the MPE is close to zero except for small dots in the Atlantic and Indian oceans.



**Figure 13.** The Mean Percentage Error (MPE) in the estimation of EWH in March 2016, comparing results obtained from GRACE CSR mascon data with those from the LRA gap-filling approach.

In estimating the Mean Percentage Error (MPE), the estimation can be made using data from a single month. However, for estimating the coefficient of determination ( $R^2$ ), data spanning multiple months are employed to ensure a more comprehensive and discernible comparison. To facilitate this comparison, an artificial data gap was intentionally

introduced during the year 2016. Subsequently, the LRA gap-filling approach was utilized to generate new observations of EWH, which were then compared to the original EWH values derived from the GRACE data. Notably,  $R^2$  values were calculated for the entire year, with the exception of the months of April, September, and October. These months were omitted because they were not covered by GRACE observations, rendering any comparison unfeasible.

Figure 14 visually represents the  $R^2$  values throughout the year 2016. The results reveal a high level of agreement, with values exceeding 98% in January, March, May, and December. Additionally, there is very good agreement, with values greater than or equal to 94%, in February, July, and August. However, the agreement in June and November is slightly lower, at 91% and 85%, respectively. This is partly due to the limited overlap between data sources in these months, with only a small portion of the observations aligning with the normal November data (from day 305 or 306 to 334 in most month, while the November 2016 observations were from 319 to 345). This means there are only 11 days common between them. In June, for instance, there are four missing EWH observation months, in addition to the month aiming to be estimated, resulting in 75% of the data being available for comparison.

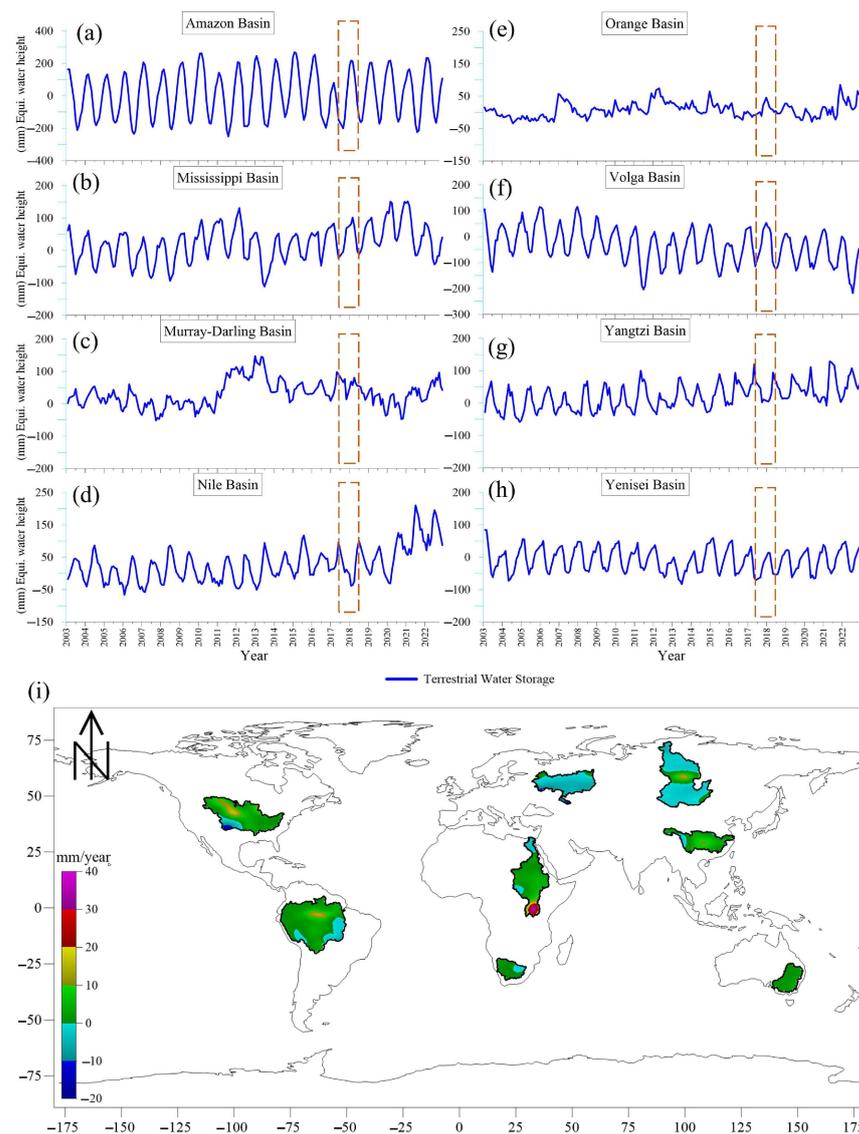


**Figure 14.** The coefficient of determination ( $R^2$ ) in the estimation of EWH during 2016, comparing results obtained from GRACE CSR mascon data with those from the LRA gap-filling approach.

The consistently high levels of agreement observed between GRACE and the LRA gap-filling approach provide us with confidence in the reliability of this method for various applications involving GRACE data.

### 5.3. TWS Trend

After successfully estimating the Equivalent Water Height using the LRA gap-filling approach to represent TWS, the method was applied to eight geographically diverse basins worldwide. The aim was to estimate TWS trends and observe TWS fluctuations over the period from 2002 to 2022 (Figure 15; Table 6). It is worth noting that only 14% of the months had no available observations, suggesting that the impact of data gap filling on the analysis is expected to be minimal.



**Figure 15.** TWS time-series for GRACE from April 2002 to February 2022. The figure is divided into nine images, with eight TWS time-series of the Amazon, Mississippi, Murray Darling, Nile, Orange, Volga, Yangtze, and Yenisei River basins (a–h). The GRACE TWS is represented by the blue line, while the orange rectangle represents the GRACE/GRACE-FO 11-month gap. Subfigure (i) represents the spatial TWS trend.

**Table 6.** TWS trend in 8 basins with and without using the LRA gap-filling approach.

Basin	TWS Trend (mm/year)	
	Using LRA	Without LRA
Amazon	$2.17 \pm 0.15$	$2.21 \pm 0.16$
Mississippi	$3.43 \pm 0.22$	$3.65 \pm 0.21$
Murry Darling	$1.74 \pm 0.13$	$1.75 \pm 0.13$
Nile	$4.94 \pm 0.25$	$5.06 \pm 0.23$
Orange	$0.94 \pm 0.36$	$0.94 \pm 0.37$
Volga	$-4.47 \pm 0.30$	$-4.53 \pm 0.28$
Yangtze	$3.49 \pm 0.32$	$3.52 \pm 0.30$
Yenisei	$-0.56 \pm 0.05$	$-0.53 \pm 0.05$

Figure 15 illustrates the seasonal time-series of TWS for these basins, along with the spatial trend derived from the mascon data. Table 6 presents the TWS trends with and without the inclusion of LRA gap-filling data.

The time-series include both the original TWS values based on the original GRACE grids and the estimated values obtained through this new gap-filling method. The orange rectangle in the figure highlights the period of the GRACE/GRACE-FO data gap. The estimated values align closely with the existing data, reinforcing the confidence in the effectiveness of this gap-filling method between GRACE and GRACE-FO. However, differences in trends were observed, which could be as large as 0.22 mm/year in certain areas (Figure 15i; Table 6). Importantly, these variations in trends are not a result of extreme values introduced by the method but are primarily attributable to the uneven sampling of the original time-series.

By comparing trends before and after gap filling, differences of several millimeters per year could be noted. For example, in the Amazon basin, the TWS trend is estimated at  $2.17 \pm 0.15$  mm/year using LRA, whereas it is  $2.21 \pm 0.16$  mm/year without it. Across all basins, positive trends are observed, except for the Volga and Yenisei basins, where negative trends are attributed to ice melting in high-latitude basins in the northern hemisphere.

Nonetheless, it is important to highlight that the LRA gap-filling approach has proven to be highly effective in bridging the data gap between GRACE and GRACE-FO. Considering its capability in filling gaps, it is natural to question whether this approach could be adapted for predictive purposes, whether looking back retrospectively or forward prospectively. From a technical standpoint, the process involves introducing data gaps at the start or end of a sequence, mimicking prediction scenarios. However, it is essential to acknowledge the substantial uncertainty linked to such purely mathematical extrapolation. Given these uncertainties, this study refrains from pursuing this application at its current stage.

## 6. Conclusions

In this study, a novel strategy based on the LRA gap-filling method is proposed to address missing data in the GRACE missions. The LRA gap-filling method is applied to fill gaps within the GRACE missions and the 11-month gap between the missions, along with a two-month gap in GRACE-FO. The LRA gap-filling method is implemented in two types of GRACE data, SHCs and mascon. In the SHC data type, five comparisons were conducted, including (1) an artificial gap inside the GRACE era compared with GRACE data, (2) random SHCs compared with other data types (IGG, QF, and GRACE), (3) estimation of EWH during December 2013 and a comparison with various datasets, (4) estimation of SHCs during the GRACE/GRACE-FO 11-month gap and comparison with the other datasets, and (5) estimation of gravity anomaly for both original and created data for December 2013. In the mascon data type, comparisons involved estimating EWH during an artificial gap in January 2016 and comparing it with GRACE-EWH, calculating statistics for the 12 months of 2009, and estimating the mean percentage error and coefficient of determination. Finally, the TWS trend was estimated from April 2002 to February 2022. The results demonstrate a high degree of consistency and accuracy when comparing LRA-derived data with original GRACE observations and other gap-filling models. These findings validate the LRA approach as a reliable tool for data interpolation, particularly during data gaps. The limitations of the LRA filling approach are that it is only relevant within the range of observed data and it cannot be utilized for predicting values for future or past observation periods. And it only reflects the resolution of the original data; it does not improve roughness resolution if it is available. In conclusion, this study underscores the significance of the LRA gap-filling approach in ensuring the continuity of critical Earth observation data from the GRACE and GRACE-FO missions. By filling data gaps, this method enables researchers to gain valuable insights into various geophysical and hydrological processes, ultimately advancing our understanding of climate change and Earth's dynamic systems.

**Author Contributions:** Conceptualization, H.A.M. and J.J.; methodology, H.A.M.; software, H.A.M.; validation, H.A.M., J.J. and W.S.; writing—original draft preparation, H.A.M.; writing—review and editing, J.J. and W.S.; visualization, W.S.; supervision, W.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundations of China (No. 41721003, 42030105, 41874023) Fundamental Research Funds for the Central Universities, and the Open Fund of Wuhan Gravitation and Solid Earth Tides National Observation and Research Station (WHYWZ202202).

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author and/or on the corresponding websites in the Acknowledgment section. CSR RL06 GRACE/GRACE-FO Mascon Solutions can be downloaded from [https://www2.csr.utexas.edu/grace/RL06\\_mascons.html](https://www2.csr.utexas.edu/grace/RL06_mascons.html) (accessed on 15 November 2023), GRACE CSR SHCs can be found at <https://www2.csr.utexas.edu/grace/RL061LRI.html/> (accessed on 16 November 2023), and GRACE-FO CSR SHCs data can be downloaded from “<ftp://isdcftp.gfz-potsdam.de/grace-fo/Level-2/CSR/RL06/>” (accessed on 16 November 2023). The Swarm ASU and COST-G data are available for download at the following links, respectively, “<http://www.asu.cas.cz/~bezdek/vyzkum/geopotencial/index.php>” (accessed on 16 November 2023) and “<https://earth.esa.int/eogateway/missions/swarm/data>” (accessed on 16 November 2023). The IGG data are available at “<https://doi.org/10.22000/357>” (accessed on 16 November 2023). The SSA data are available at “<https://doi.org/10.18419/darus-807>” (accessed on 16 November 2023).

**Acknowledgments:** The authors extend their gratitude to the Supercomputing Center of Wuhan University for generously providing access to their supercomputing system, which was instrumental in conducting the numerical calculations for this research. Special thanks are also due to Löcher and Kusche for granting access to the IGG data, as well as to Shuang Yi and Nico Sneeuw for sharing their valuable SSA data. Furthermore, the authors express their appreciation to Matthias Weigelt for providing access to the QF data utilized in this study.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Kornfeld, R.P.; Arnold, B.W.; Gross, M.A.; Dahya, N.T.; Klipstein, W.M.; Gath, P.F.; Bettadpur, S. GRACE-FO: The gravity recovery and climate experiment follow-on mission. *J. Spacecr. Rocket.* **2019**, *56*, 931–951. [\[CrossRef\]](#)
- Frappart, F.; Ramillien, G. Monitoring groundwater storage changes using the Gravity Recovery and Climate Experiment (GRACE) satellite mission: A review. *Remote Sens.* **2018**, *10*, 829. [\[CrossRef\]](#)
- Tapley, B.D.; Bettadpur, S.; Ries, J.C.; Thompson, P.F.; Watkins, M.M. GRACE measurements of mass variability in the Earth system. *Science* **2004**, *305*, 503–505. [\[CrossRef\]](#) [\[PubMed\]](#)
- Landerer, F.W.; Flechtner, F.M.; Save, H.; Webb, F.H.; Bandikova, T.; Bertiger, W.I.; Bettadpur, S.V.; Byun, S.H.; Dahle, C.; Dobslaw, H.; et al. Extending the global mass change data record: GRACE Follow-On instrument and science data performance. *Geophys. Res. Lett.* **2020**, *47*, e2020GL088306. [\[CrossRef\]](#)
- Mohasseb, H.A.; Shen, W.; Jiao, J.; Wu, Q. Groundwater Storage Variations in the Main Karoo Aquifer Estimated Using GRACE and GPS. *Water* **2023**, *15*, 3675. [\[CrossRef\]](#)
- Kusche, J.; Klemann, V.; Bosch, W. Mass distribution and mass transport in the Earth system. *J. Geodyn.* **2012**, *59*, 1–8. [\[CrossRef\]](#)
- Munagapati, H.; Tiwari, V.M. Spatio-temporal patterns of mass changes in himalayan glaciated region from EOF analyses of GRACE Data. *Remote Sens.* **2021**, *13*, 265. [\[CrossRef\]](#)
- Ghobadi-Far, K.; Šprlák, M.; Han, S.-C. Determination of ellipsoidal surface mass change from GRACE time-variable gravity data. *Geophys. J. Int.* **2019**, *219*, 248–259. [\[CrossRef\]](#)
- Liu, R.; Li, J.; Fok, H.S.; Shum, C.; Li, Z. Earth surface deformation in the north China plain detected by joint analysis of GRACE and GPS data. *Sensors* **2014**, *14*, 19861–19876. [\[CrossRef\]](#)
- Śliwińska, J.; Wińska, M.; Nastula, J. Preliminary estimation and validation of polar motion excitation from different types of the grace and grace follow-on missions data. *Remote Sens.* **2020**, *12*, 3490. [\[CrossRef\]](#)
- Pie, N.; Bettadpur, S.V.; Tamisiea, M.; Krichman, B.; Save, H.; Poole, S.; Nagel, P.; Kang, Z.; Jacob, G.; Ellmer, M.; et al. Time variable Earth gravity field models from the first spaceborne laser ranging interferometer. *J. Geophys. Res. Solid Earth* **2021**, *126*, e2021JB022392. [\[CrossRef\]](#)
- Li, W.; Wang, W.; Zhang, C.; Wen, H.; Zhong, Y.; Zhu, Y.; Li, Z. Bridging terrestrial water storage anomaly during GRACE/GRACE-FO gap using SSA method: A case study in China. *Sensors* **2019**, *19*, 4144. [\[CrossRef\]](#)
- Qian, A.; Yi, S.; Li, F.; Su, B.; Sun, G.; Liu, X. Evaluation of the Consistency of Three GRACE Gap-Filling Data. *Remote Sens.* **2022**, *14*, 3916. [\[CrossRef\]](#)
- Zhang, B.; Yao, Y.; He, Y. Bridging the data gap between GRACE and GRACE-FO using artificial neural network in Greenland. *J. Hydrol.* **2022**, *608*, 127614. [\[CrossRef\]](#)

15. Zhang, X.; Li, J.; Dong, Q.; Wang, Z.; Zhang, H.; Liu, X. Bridging the gap between GRACE and GRACE-FO using a hydrological model. *Sci. Total Environ.* **2022**, *822*, 153659. [[CrossRef](#)]
16. Yi, S.; Sneeuw, N. Filling the data gaps within GRACE missions using singular spectrum analysis. *J. Geophys. Res. Solid Earth* **2021**, *126*, e2020JB021227. [[CrossRef](#)]
17. Lenczuk, A.; Klos, A.; Bogusz, J. Studying spatio-temporal patterns of vertical displacements caused by groundwater mass changes observed with GPS. *Remote Sens. Environ.* **2023**, *292*, 113597. [[CrossRef](#)]
18. Mohasseb, H.; Abd-Elmotaal, H.A.; Shen, W. Validation of Using SWARM to Fill-in the GRACE/GRACE-FO Gap: Case Study in Africa. In Proceedings of the EGU General Assembly Conference Abstracts 2021, EGU21-2723, Virtual, 19–30 April 2021; Available online: <http://www.doi.org/10.5194/egusphere-egu21-2723> (accessed on 31 December 2023).
19. Richter, H.M.P.; Lück, C.; Klos, A.; Sideris, M.G.; Rangelova, E.; Kusche, J. Reconstructing GRACE-type time-variable gravity from the Swarm satellites. *Sci. Rep.* **2021**, *11*, 1–14. [[CrossRef](#)] [[PubMed](#)]
20. Forootan, E.; Schumacher, M.; Mehrnegar, N.; Bezděk, A.; Talpe, M.J.; Farzaneh, S.; Zhang, C.; Zhang, Y.; Shum, C.K. An iterative ICA-based reconstruction method to produce consistent time-variable total water storage fields using GRACE and Swarm satellite data. *Remote Sens.* **2020**, *12*, 1639. [[CrossRef](#)]
21. Wang, F.; Shen, Y.; Chen, Q.; Wang, W. Bridging the gap between GRACE and GRACE follow-on monthly gravity field solutions using improved multichannel singular spectrum analysis. *J. Hydrol.* **2021**, *594*, 125972. [[CrossRef](#)]
22. da Encarnacao, J.T.; Visser, P.; Arnold, D.; Bezdek, A.; Doornbos, E.; Ellmer, M.; Guo, J.; Ijssel, J.v.D.; Iorfida, E.; Jäggi, A.; et al. Description of the multi-approach gravity field models from Swarm GPS data. *Earth Syst. Sci. Data* **2020**, *12*, 1385–1417. [[CrossRef](#)]
23. Zhong, B.; Li, X.; Chen, J.; Li, Q.; Liu, T. Surface Mass Variations from GPS and GRACE/GFO: A Case Study in Southwest China. *Remote Sens.* **2020**, *12*, 1835. [[CrossRef](#)]
24. Matsuo, K.; Chao, B.F.; Otsubo, T.; Heki, K. Accelerated ice mass depletion revealed by low-degree gravity field from satellite laser ranging: Greenland, 1991–2011. *Geophys. Res. Lett.* **2013**, *40*, 4662–4667. [[CrossRef](#)]
25. Bloßfeld, M.; Müller, H.; Gerstl, M.; Štefka, V.; Bouman, J.; Göttl, F.; Horwath, M. Second-degree Stokes coefficients from multi-satellite SLR. *J. Geod.* **2015**, *89*, 857–871. [[CrossRef](#)]
26. Sośnica, K.; Jäggi, A.; Meyer, U.; Thaller, D.; Beutler, G.; Arnold, D.; Dach, R. Time variable Earth’s gravity field from SLR satellites. *J. Geod.* **2015**, *89*, 945–960. [[CrossRef](#)]
27. Uz, M.; Atman, K.G.; Akyilmaz, O.; Shum, C.; Keleş, M.; Ay, T.; Tandoğdu, B.; Zhang, Y.; Mercan, H. Bridging the gap between GRACE and GRACE-FO missions with deep learning aided water storage simulations. *Sci. Total Environ.* **2022**, *830*, 154701. [[CrossRef](#)]
28. Chu, J.; Su, X.; Jiang, T.; Qi, J.; Zhang, G.; Wu, H. Filling the gap between GRACE and GRACE-FO data using a model integrating variational mode decomposition and long short-term memory: A case study of Northwest China. *Environ. Earth Sci.* **2023**, *82*, 1–16. [[CrossRef](#)]
29. Gyawali, B.; Ahmed, M.; Murgulet, D.; Wiese, D.N. Filling Temporal Gaps within and between GRACE and GRACE-FO Terrestrial Water Storage Records: An Innovative Approach. *Remote Sens.* **2022**, *14*, 1565. [[CrossRef](#)]
30. Sun, Z.; Long, D.; Yang, W.; Li, X.; Pan, Y. Reconstruction of GRACE data on changes in total water storage over the global land surface and 60 basins. *Water Resour. Res.* **2020**, *56*, e2019WR026250. [[CrossRef](#)]
31. Li, F.; Kusche, J.; Rietbroek, R.; Wang, Z.; Forootan, E.; Schulze, K.; Lück, C. Comparison of data-driven techniques to reconstruct (1992–2002) and predict (2017–2018) GRACE-like gridded total water storage changes using climate inputs. *Water Resour. Res.* **2020**, *56*, e2019WR026551. [[CrossRef](#)]
32. Lenczuk, A.; Weigelt, M.; Kosek, W.; Mikocki, J. Autoregressive Reconstruction of Total Water Storage within GRACE and GRACE Follow-On Gap Period. *Energies* **2022**, *15*, 4827. [[CrossRef](#)]
33. Karimi, H.; Iran-Pour, S.; Amiri-Simkooei, A.; Babadi, M. A gap-filling algorithm selection strategy for GRACE and GRACE Follow-On time series based on hydrological signal characteristics of the individual river basins. *J. Geod. Sci.* **2023**, *13*, 2023. [[CrossRef](#)]
34. Wiese, D.N.; Landerer, F.W.; Watkins, M.M. Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution. *Water Resour. Res.* **2016**, *52*, 7490–7502. [[CrossRef](#)]
35. Save, H.; Bettadpur, S.; Tapley, B.D. High-resolution CSR GRACE RL05 mascons. *J. Geophys. Res. Solid Earth* **2016**, *121*, 7547–7569. [[CrossRef](#)]
36. Wahr, J.; Molenaar, M.; Bryan, F. Time variability of the Earth’s gravity field: Hydrological and oceanic effects and their possible detection using GRACE. *J. Geophys. Res. Solid Earth* **1998**, *103*, 30205–30229. [[CrossRef](#)]
37. Chen, J.; Tapley, B.; Tamisiea, M.E.; Save, H.; Wilson, C.; Bettadpur, S.; Seo, K.-W. Error Assessment of GRACE and GRACE Follow-On Mass Change. *J. Geophys. Res. Solid Earth* **2021**, *126*, e2021JB022124. [[CrossRef](#)]
38. Swenson, S.; Famiglietti, J.; Basara, J.; Wahr, J. Estimating profile soil moisture and groundwater variations using GRACE and Oklahoma Mesonet soil moisture data. *Water Resour. Res.* **2008**, *44*, W01413. [[CrossRef](#)]
39. Löcher, A.; Kusche, J. A hybrid approach for recovering high-resolution temporal gravity fields from satellite laser ranging. *J. Geod.* **2020**, *95*, 6. [[CrossRef](#)]
40. Weigelt, M. *Time Series of Monthly Combined HLSST and SLR Gravity Field Models to Bridge the Gap between GRACE and GRACE-FO: QuantumFrontiers\_HLSST\_SLR\_COMB2019s*; GFZ Data Services; GFZ: Potsdam, Germany, 2019. [[CrossRef](#)]
41. Gillard, J. An overview of linear structural models in errors in variables regression. *REVSTAT-Stat. J.* **2010**, *8*, 57–80. [[CrossRef](#)]

42. Tranmer, M.; Elliot, M. Multiple linear regression. *Cathie Marsh Cent. Census Surv. Res. (CCSR)* **2008**, *5*, 1–5.
43. Casson, R.J.; Farmer, L.D. Understanding and checking the assumptions of linear regression: A primer for medical researchers. *Clin. Exp. Ophthalmol.* **2014**, *42*, 590–596. [[CrossRef](#)] [[PubMed](#)]
44. Zou, K.H.; Tuncali, K.; Silverman, S.G. Correlation and simple linear regression. *Radiology* **2003**, *227*, 617–628. [[CrossRef](#)] [[PubMed](#)]
45. Laurent, R.T.S. *Understanding Regression Assumptions*; Taylor & Francis: Abingdon, UK, 1994.
46. Berry, W.D. *Understanding Regression Assumptions*; Sage: New York, NY, USA, 1993. [[CrossRef](#)]
47. Cui, L.; Song, Z.; Luo, Z.; Zhong, B.; Wang, X.; Zou, Z. Comparison of terrestrial water storage changes derived from GRACE/GRACE-FO and Swarm: A case study in the Amazon River Basin. *Water* **2020**, *12*, 3128. [[CrossRef](#)]
48. Heiskanen, W.A.; Moritz, H. *Physical Geodesy*; Springer: Berlin/Heidelberg, Germany, 1967. [[CrossRef](#)]
49. Vanicek, P.; Krakiwsky, E.J. *Geodesy: The Concepts*; Elsevier: Amsterdam, The Netherlands, 2015.
50. Tscherning, C. Determination of sea surface topography from satellite radar altimetry [ocean geoid, ocean circulation, altimeter]. In *Conference on Satellite Based Navigation and Remote Sensing of the Sea, Copenhagen (Denmark), 4 Mar 1980*; DDNIUGG: Ottawa, ON, Canada, 1980.
51. Wolf, P.R.; Ghilani, C.D. *Adjustment Computations Spatial Data Analysis*; John Wiley: Hoboken, NJ, USA, 2006.
52. Ghilani, C.D. *Adjustment Computations: Spatial Data Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2017.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.