Characterization of the hydro-geological regime of Yangtze River basin using remotely-sensed and modeled products


PII: S0048-9697(20)30864-0
DOI: https://doi.org/10.1016/j.scitotenv.2020.137354
Reference: STOTEN 137354

To appear in: Science of the Total Environment

Received date: 11 November 2019
Revised date: 29 January 2020
Accepted date: 14 February 2020


This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier.
Characterization of the hydro-geological regime of Yangtze River basin using remotely-sensed and modeled products

V.G. Ferreira\textsuperscript{a,b,}\textsuperscript{*}, B. Yong\textsuperscript{a,b}, M.J. Tourian\textsuperscript{c}, C.E. Ndehedehe\textsuperscript{d}, Z. Shen\textsuperscript{b}, K. Seitz\textsuperscript{e}, R. Dannouf\textsuperscript{b}

\textsuperscript{a}State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, 210098, China
\textsuperscript{b}School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China
\textsuperscript{c}Institute of Geodesy, University of Stuttgart, Stuttgart, Germany
\textsuperscript{d}Australian Rivers Institute and Griffith School of Environment & Science, Griffith University, Nathan, Queensland 4111, Australia
\textsuperscript{e}Geodetic Institute of Karlsruhe, Karlsruhe Institute of Technology, Karlsruhe 76128, Germany

Abstract

The hydrology of the Third Pole, Asia’s freshwater tower, has shown considerable sensitivity to the impacts of climate change and human interventions, which affect the headwaters of many rivers that originate therein. For example, the Yangtze River has its basin (YRB) experiencing wetness of terrestrial water storage (TWS), which rainfall seems to be the primary source as inferred from the previous studies. Consequently, it is crucial to understand the contributions of each TWS’s sub-domain - i.e., groundwater (GWS); total water content (TWC) stored as soil moisture, ice/snow, and canopy; and the surface water (SWS) storages - on YRB’s wetness. Hence, SWS, from altimetry and imagery satellites, and TWC, from Global Land Data Assimilation System, are inverted considering the same basis function as for TWS from the Gravity Recovery and Climate Experiment, which account for the differences in the resolutions inherent in each product. Furthermore, a “tie-in signal approach is used to fit the temporal patterns of GWS, TWC, and SWS to TWS (i.e., the observations). Results show improvements in the reconstructed GWS series concerning standard deviation, correlation coefficient, and NashSutcliffe efficiency of 22\%, 27\%, and 120\%, respectively, regarding the use of the TWS-budget equation. The reconstructed time series of GWS, TWC, and SWS present an increase of 1.76, 2.69, and 0.14 mm per year (mm/yr) and that YRB loses water stored at its aquifers 55\% of the time (regarding 2003-2016 period) based on the quantile function of storage (QFS). The QFS’s slope shows that TWS has a fast and small storage potential w.r.t. inland waters and soil moisture reflect the dryness impacting TWS first. Despite the evidence of an increase of 19.05 mm/yr in annual precipitation, which seems to explain the bulk in TWS, further investigation to characterize controls on TWS memory within YRB still necessary.

\textsuperscript{*}Corresponding author
Email address: vagnergf@hhu.edu.cn (V.G. Ferreira)
1. Introduction

The Third Pole is the water tower of Asia and encompasses the Himalaya-Hindu Kush mountain ranges and the Tibetan Plateau (Zhang et al., 2019b). It is the headwaters of the source of the ten large Asian river systems (i.e., the Amu Darya, Indus, Ganges, Brahmaputra, Irrawaddy, Salween, Mekong, Yangtze, Yellow River, and Tarim) fed by melt-waters (Gao et al., 2019). Following Yao et al. (2012), the Third Pole was identified as one of the most sensitive regions to the impacts of climate change, which are already visible. The Tibetan Plateau is known to be vulnerable to warming due to anthropogenic activities, and Zhang et al. (2017) have reported on the intensification of the region’s hydrologic cycle during the recent decades. For example, the Tibetan Plateau’s glaciers are melting fast, and in the past half-century about 82% have retreated and in the last decade 10% of the plateau’s permafrost regions have degraded (cf. Qiu, 2008). Consequently, water availability in river systems due to the shrinking glaciers is expected to be influenced by the ongoing global climate change, as pointed out by Kaser et al. (2010). The Indus and Brahmaputra basins heavily rely on the meltwaters whereas such hydrological process plays only a modest role for the Ganges, Yellow, and Yangtze basins (Immerzeel et al., 2010). For instance, glacier runoff accounted for approximately 11% of the total discharge of the Yangtze River basin (YRB), that is, concerning the period of 1961-2000 (Shiyin et al., 2009). The groundwater resources in the Tibetan Plateau, where the source of the Yangtze River is located, are controlled by permafrost, topography, geomorphology, tectonics, and surface hydrology (cf. Xiang et al., 2016) as well as the development of water resources. Furthermore, groundwater occurrences in YRB make it a valuable supplement to surface water, especially throughout the Yangtze River Valley, downstream of the Three Gorges Dam (TGD). Hence, understanding the groundwater changes over this region plays a central role since several rivers in the area are effluent rivers (gaining stream) and, consequently, promotes the regional economy.

Groundwater quantity in YRB is not an issue (see, e.g., Jiang et al., 2016; Feng et al., 2018), but there are reports of high concentrations of arsenic (As) in Jianghan Plain, an inland sedimentary basin, in central YRB (Schaefer et al., 2017; Deng et al., 2018). The concentration of As is generally associated at locations with heavy groundwater abstraction (e.g., Smith et al., 2018) as well as shifts in hydrology resulting in seasonally variable As concentrations in groundwater (e.g., Schaefer et al., 2017). However, in-depth investigation of the spatial distribution of groundwater resources in parts of the Third Pole region and, consequently, its portion over YRB, is heightened by the physiography of the area and the fact that only a few well stations exist (Xiang et al., 2016), which imposes challenges. Given that most groundwater use is for irrigation, major groundwater aquifers can be exposed to salinity and other
negative impacts due to fluctuations in the water table thus making groundwater resources difficult to assess and manage. A large-scale quantitative assessment of groundwater is therefore required to support sustainable groundwater extraction limits and manage the impacts of extraction. Consequently, multi-source approach based on remotely-sensed and modeled datasets is a viable resource to monitor groundwater changes as investigated by, for example, Rodell et al. (2009). In such a case, the Gravity Recovery and Climate Experiment (GRACE) mission (Tapley et al., 2004) plays a key role since the GRACE observations are uniquely inverted to the terrestrial water storages (TWS) given that the other mass variations (i.e., ocean and atmosphere) that contribute to the temporal variations of the Earth’s gravity field have been reduced (e.g., Wouters et al., 2014). Past studies that used GRACE data over YRB focused on the overall use of TWS in the context of terrestrial water budget (see, e.g., Ferreira et al., 2013; Long et al., 2015; Chen et al., 2019) and its analysis in drought studies (see, e.g., Zhang et al., 2015, 2016; Sun et al., 2018; Zhang et al., 2019a). However, only few studies have been dedicated to monitor and investigate groundwater resources in YRB using GRACE datasets. For instance, Xiang et al. (2016) reported an overall increase in the groundwater storage over the source of Yangtze River (it originates from the Tongtian River, see Section 2.1) with a rate of 1.86 ±1.69 Gigatons per year (Gt/yr) during the period 2003-2009. During the same period, groundwater storage in the Inner Tibetan Plateau increased by 5.01 ±1.59 Gt/yr (Zhang et al., 2017). Over YRB’s broad domain, groundwater gain of 3.34 Gt/yr (significant at 95% confidence level) has been reported in a study carried out by Jiang et al. (2016) considering the same period. More recently, Wang et al. (2019) reported an increase of approximately 9.7 ±7.7 Gt/yr based on in-situ data covering the time span of 2005-2011.

Regarding changes in TWS, its increase in the whole YRB, at a rate of 5.47 Gt/yr, has been reported while precipitation increased at a rate of 0.94 Gt/yr during 2003-2014 (cf. Sun et al., 2017). Furthermore, Rodell et al. (2018) reported on the overall increase of TWS over eastern central China (with a surface area of about 6.57×10^5 km^2), a region that coincides with the portion of the Third Pole known as the Three Rivers source region. They noted an apparent increase of 7.8 ±1.6 Gt/yr over the period between Apr 2002 and Mar 2016 (14-year period). Moreover, they attributed the observed rise in TWS to the number of dam construction and subsequent reservoir filling over the region. Nevertheless, most of existing hydropower over YRB is located at its middle and lower reaches (Zarfl et al., 2015), which is relatively far from eastern central China. Indeed, YRB will face the highest dam construction activity in Asia, as reported by Zarfl et al. (2015). For example, considering dam with a capacity larger than 1.0 Megawatt, YRB will have additional 140 dams considering the ones under construction (95) and planned (45) (see Zarfl et al., 2015, for more details on this topic). Furthermore, approximately 14% of eastern central China is under irrigation – as summarized by Rodell et al. (2018) based on the findings of Salmon et al. (2015) – which might induce infiltration from canal and paddy rice fields’ seepage. Due to the filling of the dams, water seeps underground and elevates the water table (e.g., Chao et al., 2008).
Within YRB, the Three Gorges Reservoir (TGR), has a gravitational signature in GRACE’s observations (cf. Wang et al., 2011) and seepage response due to TGR’s water filling has been investigated (cf. Wang et al., 2019). For example, a mass gain of 3.85 Gt/yr for TGR based on GRACE data has recently been reported (Wang et al., 2019). Whereas this figure is only 2.29 Gt/yr based on in-situ water level, the difference of about 1.56 Gt/yr could be mainly due to seepage, that is, approximately 41% the GRACE derived TGR’s mass changes.

As previously mentioned, considering the complex physiography of YRB (especially the portion belonging to the Third Pole), in-situ datasets are limited in quantity, and, to make things worse, access to them is a Sisyphean task. Thus, one could rely on hydrologic models, which generally depend on the ground-based datasets to provide meaningful results. However, available hydrologic models, such as the WaterGAP Global Hydrology Model (WGHM, Dll et al., 2003), face challenges to predict groundwater storage over the Third Pole (e.g., Xiang et al., 2016). Specifically, over YRB, the outputs from WGHM are in weak agreement with the in-situ measurements (cf. Wang et al., 2019), probably due to the complex interactions between surface water hydrology and the aquifer. Consequently, in this contribution, the question of groundwater variations is addressed over YRB by applying a novel combination of satellite gravimetry, imagery, and altimetry with modeled water contents based on hydrological modeling. To start with, the TWS-budget equation (see, e.g., Rodell et al., 2009):

\[ tT_n = tG_n + tS_n + tI_n + tR_n + tC_n, \]

shows that the groundwater storage \((G)\) can be isolated from the GRACE-derived terrestrial water storage \((T)\) if soil moisture \((S)\), ice and snow \((I)\), inland surface reservoirs \((R)\), and canopy water \((C)\) storages are available. In Eq. (1), the sub-scripts \(t\) and \(n\) stand for the number of rows, representing the epochs (months), and the columns, representing the bins of the geographical locations, respectively. However, many authors that have used such approach have not considered altogether the resolution of the different data products and a proper signal separation procedure. Regarding to the former, it is considered the point-mass modeling (e.g., Baur & Sneeuw, 2011) as the basis function to invert the gravitational potential due to each sub-domain of TWS appearing in Eq. (1) - except groundwater storage (GWS), which is unknown - into equal-area panels covering the entire Earth’s surface. Regarding to the latter, a signal separation approach is considered to estimate GWS (i.e., \(G\) in Eq. 1) using remotely-sensed and modeled datasets, which has its roots in the previous contributions of Schmeer et al. (2012) and Forootan et al. (2014).

Under such a framework, the main objective of this study is to understand the hydro-geological regime of YRB in response to TWS and precipitation changes following the analysis in the inter- and intra-annual time scales. The main justification for this study relies on the simple fact that there is no
previous work dealing with the quantitative TWS-budget equation and its bulk in YRB. To accomplish this, it is introduced a method to derive groundwater storage fields by adequately considering the nature of different datasets, which include GRACE (Tapley et al., 2004), Earth Observation (EO) satellites using imagery, altimetry (e.g., Crétaux et al., 2011) and precipitation (Huffman et al., 2007) sensors, and modeled soil moisture and ice/snow (when applicable) storages from Global Land Data Assimilation Systems (GLDAS, Rodell et al., 2004). The method is referred here to as a “tie-in signal approach”, i.e., EO- and model-based products are fitted to the GRACE-derived TWS in lieu of isolating them from TWS.

The remainder of this study is structured as follows: Section 2 presents the regional setting of the study area, followed by a description of the datasets and methods in Section 3. The results and discussion are presented in Section 4 followed by the conclusions of the study in Section 5.

2. Regional Setting

2.1. Geography

The Tibetan Plateau (Figure 1), also known as the “Roof of the World” in a metaphorical description, is regarded as the water source of China as well as the south and southeast Asia regions. For instance, the Yangtze River (Chang Jiang), and many others, such as the Mekong River and Yellow River, originates in the Inner Tibet (Figure 1). Consequently, the Tibetan Plateau provides high-quality water for China through rivers. For instance, the Yangtze River takes its first drop from the Jianggendiru Glacier, on the southwest side of Geladandong Peak, located in southwestern Qinghai. The Yangtze River ranks as the third-largest river on Earth by discharge, which disembogues in the East China Sea with an average discharge of $34 \times 10^3$ m$^3$/s after a journey of about $6.3 \times 10^3$ km (e.g., Ferreira et al., 2013). The source of the Yangtze River is the Tongtian River, which has three sources, namely the Tuotuo River (the actual source), the Dangqu River (as the southern source), and the Chumar River (as the northern source) as further explained by Chen (2020a). The TGR, the most important anthropogenic features in YRB (e.g., Yang et al., 2015), controls the Yangtze River at a length of 2.3 km and a height of 185 m. The drainage of YRB is approximately $1.8 \times 10^6$ km$^2$ and its upper reach (the headwaters) extends from its most western part at Tuotuohe (34.09$^\circ$N, 92.91$^\circ$E) to Yichang (30.69$^\circ$N, 111.29$^\circ$E) occupying about 55% of the total surface of the basin. This geomorphic region presents a deep cutting plateau area identified in Figure 1 by the isoline of elevation exceeding 2000 masl. The middle reach extends from Yichang to Hukou (29.65$^\circ$N, 116.28$^\circ$E), and the lower range extends up to the Yangtze River’s mouth at East China Sea (32.60$^\circ$N, 94.51$^\circ$E), near Shanghai. Furthermore, the middle and lower reaches of the Yangtze River include the Jianghan Plain (downstream of TGD), the Dongting Lake, the Poyang Lake,
the Lake Tai, and the Delta of the Yangtze River of which topography varies from west to east by the
gradient from 40 m to 2 m.

Figure 1: The geographical location of the Third Pole, denoted by areas with an elevation exceeding 2000 masl (cf.
Immerzeel et al., 2010) as indicated by the opacity in red (here defined using the ETOPO1 global relief). The region defined
in yellow depicts YRB’s geographical domain. The numbers within the circles denote the glaciers as (1) Central Himalaya,
(2), East Himalaya, (3) South and East Tibet, (4) Hengduan Shan, (5) East Kunlun and Inner Tibet, and (6) Qilian Shan
(Xiang et al., 2016). The main hydro-geological provinces within YRB are depicted by the colored polygons, in which the
colors are associated with the annual recharges (WHYMAP, 2008). The dotted-lines in red depict the limits of the upper,
middle, and lower reaches of YRB. The red triangles depict the approximated location of the in-situ groundwater storage
as digitally reconstructed from Wang et al. (2019). The scale refers to the center of the map in Mercator projection.

The Asian atmospheric circulation, hydrological and geological settings, topography, surface waters,
and developments of water resources govern the groundwater occurrences in YRB (cf. Chen, 2020c). The
groundwater system in YRB is primarily characterized as a sub-tropical humid system in the hilly area
and plain of the Yangtze River, and the secondary system is in the hilly area of middle and lower reaches
of Yangtze River (CGS, 2012). The Yangtze River Plain starts from Yichang and includes the Huaiyang
mountains, the fold system in South China. The north of Yangtze River is the lower plains of Quaternary
unconsolidated sediment distribution, mainly carbonate rocks to the west and metamorphic rocks and
igneous rocks mixed distribution to the east areas. The deposits in the middle and lower reaches of YRB
are mainly the Quaternary river, lake, and marine clay of which the accumulative thickness of soft soil
ranges from 3 to 6 m.

Figure 1 shows the primary groundwater environments in YRB compiled from the Groundwater
Resources Map of the World 1:25,000,000 provided by the World-wide Hydrogeological Mapping and
Assessment Programme (WHYMAP, 2008). The areas in blue are used for large and rather uniform
groundwater basins, usually in large sedimentary basins that may offer excellent conditions for groundwa-
ter exploitation. This coincides with the Pleistocene aquifer of the central YRB, precisely, the Jianghan
Plain, where groundwater mining takes place. The regions in green present complex hydro-geological
structures in heterogeneous folded or faulted regions where highly productive aquifers, including karst aquifers, may occur and zones of high yielding aquifers identified. The other areas in the basin generally present limited groundwater resources in local and shallow aquifers. The overall distribution of groundwater resources in YRB is characterized as follows: groundwater occurrences are higher in the middle and lower reaches, which distribution is higher in the three lakes area, relatively to the upper reaches (e.g. Chen, 2020c). Furthermore, groundwater recharge in YRB takes 97.7% from precipitation and 2.3% from surface waters, whereas in mountainous regions this figure is 100.0% from precipitation (cf. Chen, 2020c).

There are 1332 ice-filled regions in YRB (glaciers, code 5K) with a total surface area of approximately $1.9 \times 10^3 \text{ km}^2$ consisting of both, continental and marine temperate glaciers (e.g. Chen, 2020c). However, these figures have been updated by a new release of the Chinese glacier inventory (second), and there are 1,378 glaciers covering a total area of 1541.3 km$^2$ with an uncertainty of $\sim 3.2\%$ (cf. Guo et al., 2015). The area of the glaciers in YRB has been retreated by approximately 12% over the period from 1977 to 2009, which could be attributed to the global warming in the last decades since there is a significant correlation between mass balance and temperature (cf. Yuanqing et al., 2008). For instance, the glaciers in the Gongga Mountain region (29.6°N, 101.9°E), located at south-eastern margin of Qinghai-Tibetan Plateau and the north bank of Yangtze River, have declined 11.3% during 1966-2009 (cf. Pan et al., 2012). This has been hampered by the fact that supraglacial debris coverage in the Gongga Mountain is also contributing to the rates of mass loss since glaciers at lower elevations seem to lose more mass than at higher altitude as shown by the analysis of Cao et al. (2019).

2.2. Climate

Located at sub-tropical and temperate climate zones, YRB has mean annual precipitation generally higher than 1000 mm, which is associated with very favorable natural conditions, and, consequently, the basin is the bed of many rivers and lakes (Figure 1). Precipitations increase typically from northwest to southeast providing an uneven distribution due to the monsoon activities through the basin (e.g., Chen, 2020b). The mean annual air temperature is approximately 14°C, and average yearly evaporation is higher than 850 mm. The annual precipitation is higher in YRB’s portion at the south of the Yangtze River in comparison with the north region, and it decreases from southeast to the northwest areas of the basin. Considering the Upper, Middle, and Lower reaches of YRB (Figure 1), with catchment areas occupying approximately 55%, 29%, and 16% of the whole basin, the respective annual average precipitation accounts for 897, 1255, and 1455 mm while the annual average potential evapotranspiration is 858, 836, and 894 mm (cf. Zhang et al., 2016). Considering the Thornthwaite annual moisture index (e.g., Willmott & Feddema, 1992), the Upper, Middle, and Lower reaches of YRB would present 1.05, 1.50, and 1.63, respectively, indicating that moisture supplies (precipitation) over these regions exceed
the atmospheric demands (evapotranspiration) by approximately 5%, 50%, and 47%. (The snow-melt over snow-covered environs were not considered.)

Furthermore, several drought events have been reported over YRB (see, e.g., Zhang et al., 2015, 2016; Sun et al., 2018; Zhang et al., 2019a). Historical records have shown that YRB has faced droughts once every 1.8 years on average in which severe droughts ravage YRB every 7.8 years (cf. Chen, 2020c). The overall cycle of 2-8 years period seems to be related to the cycle of El Niño/Southern Oscillation (ENSO) index series. Indeed, the relationship between droughts/floods presents teleconnections with La Niña/El Niño (cold and warm phases of ENSO, respectively) events over the lower and middle reaches of YRB and, conversely, the relationships are reversed over the upper reaches of YRB (cf. Tong et al., 2006). For instance, Zhang et al. (2015) have found a significant correlation between El Niño events with high TWS occurrences and, conversely, La Niña with low TWS with a phase lead of 7 to 8 months of which the most impacts are seen over the Lower YRB in comparison to the Upper YRB.

3. Material and Methods

3.1. Datasets

3.1.1. Gravitational potential (synthesized at GRACE’s altitude)

The Stokes coefficients provided by the Center for Space Research (CSR) of the University of Texas, USA up to a maximum degree of 96 ($n_{\text{max}} = 96$) were considered. The degree-one coefficients were replaced by the results provided by Swenson et al. (2008), because GRACE gravity solutions did not provide those coefficients, which represent the changes in the geocenter due to mass redistribution in the Earth system. Likewise, the GRACE-derived, degree-two coefficients present relatively high uncertainties and so they were replaced by the values derived from satellite laser ranging provided by Cheng & Tapley (2004). They are also based on RL06, which uses new background models. A glacial isostatic adjustment (GIA) model of secular trends (Geruo et al., 2013; Peltier, 2004) was removed from the secular changes in the Stokes coefficients. However, signals due to large earthquakes were not considered; hence, the analysis of water mass gain/loss must be treated with caution in regions near large earthquakes. Noteworthy, the coefficients were not filtered as it would be generally necessary while using them to estimate mass changes (cf. Swenson & Wahr, 2006). This is supported by the studies carried out by Forsberg & Reeh (2007) and Baur & Sneeuw (2011), which neither smoothing nor filtering were applied to the Stokes coefficients. Moreover, Jacob et al. (2012) did not apply a “destripe” step in their mascon solutions based on the Stokes coefficients, hence improving the sample characteristics of the estimated mascons.

The residual Stokes's coefficients, $\Delta \hat{C}$ and $\Delta \hat{S}$, after removing their respective long-term mean, can...
be used in
\[
\delta V(r, \varphi, \lambda, t) = \frac{GM}{r} \sum_{n=0}^{n_{\text{max}}} \left( \frac{R}{r} \right)^{n+1} \frac{1}{1 + k_n} \times \sum_{m=0}^{n} \left[ \Delta \bar{C}_{nm}(t) \cos(m\lambda) + \Delta \bar{S}_{nm}(t) \sin(m\lambda) \right] \bar{P}_{nm}(\sin \varphi),
\]

(2)

to compute the band-limited gravitational potential (\(\delta V\)) at the GRACE’s orbit (approximately 500 km in height). In Eq. (2), \(GM\) is the geocentric gravitational constant of the Earth, \(R\) is the mean radius of the Earth (6,378 km), \(r\) is the geocentric distance of the GRACE’s orbit (approx. \(R + 500\) km), \(n\) and \(m\) are the degree and order, \(k_n\) are the degree-dependent load Love numbers accounting for the contribution of the solid Earth (Wahr et al., 1998), and \(\lambda\) and \(\varphi\) are the longitudes and latitudes. The use of the residual Stokes’s coefficients in Eq. (2) results in the monthly variations of the gravitational potential (\(m^2 s^{-2}\)), which can be used to estimate the variations in surface densities over a particular region by applying inverse modeling (e.g., Ramillien et al., 2011).

3.1.2. Model for continental hydrology

3.1.2.1. The GLDAS water content.

The GLDAS reanalysis, which consists of a series of land surface state (e.g., soil moisture and surface temperature) and flux (e.g., evapotranspiration and sensible heat flux) products (Rodell et al., 2004), was employed to model the soil moisture, ice and snow storages and canopy water. GLDAS incorporates both ground- and space-based observation systems to produce optimal estimates of land surface state and flux. GLDAS products consist of four land models, namely, Community Land Model (CLM), Mosaic, Noah, and Variable Infiltration Capacity (VIC). Here, it was considered the Noah model (Ek et al., 2003) of the GLDAS Version 2.1 (GLDAS-2.1), with a spatial resolution of 0.25 arc-degree and a temporal resolution of one month (Hiroko et al., 2016).

The Noah land surface model in GLDAS provides continental water storages composed by the soil moisture in layers totaling 2 meters depth, accumulated snow, and plant canopy surface water in which the sum of these compartments is named here as total water content (TWC). This means that while using Eq. (1), or its modification to be presented in Section 3.2.4, to estimate \(G\), that is, groundwater storage, any mass variation below the 2 meters depth will be regarded as groundwater. In Section 2.1 it has been reported that the soft soil layers in YRB varies from 3-6 m, hence, distinctions between shallow groundwater and soil moisture bellow to 2 m will be neglected. Due to the unreliable Greenland forcing data and a lack of a glacier/ice sheet model, snow water equivalent accumulates indefinitely over Greenland and a few other Arctic locations and they were masked out.

3.1.2.2. The WaterGAP groundwater content.
The groundwater fields simulated from the Water Global Assessment and Prognosis (WaterGAP) model (e.g. Schmied et al., 2014) were used for assessment and interpretation purposes of estimated GWS based on GRACE data. It was considered the new version of the WaterGAP, version 2.2d, which is calibrated and forced with a combination of WFD and WFDEI, with monthly precipitation scaled to GPCC (WFDEI-GPCC). The monthly groundwater storage fields from the WaterGAP covers a period of Jan 2003 till Dec 2016 with a spatial resolution of 0.5°. The WaterGAP provides fluxes and storages of water on all continents, apart from Antarctica, considering the anthropogenic influence on the natural freshwater systems due to water abstractions and impoundments (cf. Dll et al., 2014). The model is developed at the University of Kassel and the University of Frankfurt in Germany. The WaterGAP model consists of WGHM (Dll et al., 2003) and five models of water uses consisting of irrigation, livestock, households, manufacturing, and cooling of thermal power plants.

3.1.3. Earth Observation derived surface water storage

3.1.3.1. Elevation-area relationship for the major lakes of YRB.

The water levels for Dongting Lake, Poyang Lake, and Lake Tai were retrieved from the Database for Hydrological Time Series of Inland Waters (DAHITI), whose further description can be found in Schwatke et al. (2015). Outliers were removed for all water levels larger than three standard deviations from the mean value for the respective series. The water levels for Dongting Lake have records of water levels in DAHITI starting from Jan 2002 to Aug 2019, Poyang Lake Jun 2002 to Jul 2017, and Lake Tai from Jun 2002 to Jun 2016. The water level series of Lake Tai presents many missing values, especially 2011 and 2012. Indeed, Lake Tai has not fully been covered by many altimetry missions. Consequently, other EO satellite, i.e., satellite imagery, is considered to provide surface area changes necessary for volume estimations and to fill the gaps in altimetry time series (or vice-versa) over, for example, Lake Tai.

The surface area changes for the three lakes were computed based on the data delineated from the Moderate Resolution Imaging Spectroradiometer (MODIS) vegetation indices (MYD13Q1) and Terra MOD09A1 Version 6 products with resolutions of 250 and 500 m, respectively. For almost every month, one set of the 16-day product was acquired, which consisted of scenes that covered the respective lakes. However, due to the low quality of the images in some months due to cloud cover, only the best images were used, and a mask was created to separate the water pixels from land using a given threshold. The normalized difference vegetation index (NDVI) values ranged from -1 to +1 with water values mostly being classified with values less than zero. The threshold was manually adjusted for every image and then compared with the true color image to avoid misclassifications. This method is somewhat like the ones used by Wang et al. (2014) and Sheng et al. (2016). The range of values for each image was found
considering the values for the middle (water) and outside (land and or vegetation) of the lake. The mask was then applied, and the images were converted to binary images containing pixels 1 for water and 0 for land and vegetation. The binary images were further used for delineating the shape of the lakes and extracting their respective shorelines, providing monthly surface areas.

Notably, it is desired here only the volume anomalies, which might be possible to estimate assuming that the water level and surface area variations, regarding their respective mean values, might have a linear relationship for the three lakes. For each lake, an elevation-area relationship was derived in which elevation or area can be inferred, given that one of them must be accessible. Due to the correlation of the known elevations and surface areas, a linear polynomial was chosen to describe the elevation-area relationship (Figures 2a, 2c, and 2e). The Spearman’s correlation coefficient (\( \rho \)) presents a moderated monotonic relationship between elevation and area for Dongting and Tai with values of 0.44 and 0.48, respectively. For Poyang Lake there is a strong relation with \( \rho \) equals to 0.78. The moderated monotonic relationship between elevation and area for Dongting Lake and Lake Tai might be due to the poor quality of the imagery-derived surface areas using NDVI, which is not solid as can be seen from Figures 2a and 2e. The relationship would be linear if and only if the lakebed, characterized by the bathymetry, would be linear, which is absolutely not the case here. Accurate extraction of water masks would be more efficient using, for example, a method based on maximum a posteriori estimation of Markov random fields considering pixel intensity, spatial correlation between neighboring pixels, and temporal behavior of the water body (cf. Elmi et al., 2016). This will be considered in future work.

Albeit the scatter plots shown in Figures 2a, 2c, and 2e show moderate relationship between area \((a)\) and elevation \((h)\), linear regression was considered. The relationship for Dongting, Poyang, and Tai presented linear polynomials as \(h = 1.5 \times 10^{-3} \cdot a\), \(h = 2.1 \times 10^{-3} \cdot a\), and \(h = 1.0 \times 10^{-3} \cdot a\), with the associated \(R^2\) as 0.27, 0.66, and 0.22, respectively. The elevation-area relationships were then used for making provisions for the missing months of surface areas or water level data to build both time series covering the period between Dec 2002 to Dec 2016. Finally, volumes of the three lakes were estimated considering the pairs of area and elevation of the current month \((i)\) and the previous one \((i - 1)\). This was done by considering the water changes describing a volume of a trapezoid prism where the mean area was computed as \(\bar{a}_i = 0.5(a_i + a_{i-1})\), and the elevation difference as \(\Delta h_i = h_i - h_{i-1}\), which was used to estimate the volume variations as \(\Delta v_i = \bar{a}_i \cdot \Delta h_i\). The volume changes over the entire study period for the respective lakes are shown in Figures 2b, 2d, and 2f. The missing periods seen in Figure 2 were filled using a periodic signal with annual and semi-annual amplitudes.

### 3.1.3.2. Elevation-volume relationship for TGR.

Figure 2h shows the water level series of TGR due to the water impoundment of the TGD, obtained from the China Three Gorges Corporation database over the period of Jan 2003 to Dec 2016. However,
Figure 2: Panels (a), (c), and (e) show the relationship between water level and surface area considering the periods that both datasets are available for Dongting Lake, Poyang Lake, and Lake Tai, respectively. Panels (b), (d), and (f) present the volume variations for the available water level and surface areas for Dongting Lake, Poyang Lake, and Lake Tai, respectively. Panel (g) presents the water level variations (right-axis) for the TGR, the left-axis depicts the volume storage computed by using the water level and power law relationship proposed by Wang et al. (2011).

since surface area changes in the reservoir were not computed like the other three lakes (Section 3.1.3.1), an elevation-volume relationship, which is generally described by a power function, can be considered. In Wang et al. (2011), a power-law relationship between volume and water level as $v = 0.2968 \cdot 1.0284^h$, with $R^2 \approx 1.00$ has been proposed. This power-law relationship has been used in other studies (e.g., Wang et al., 2019) and an overall mean absolute difference of about 1.0-0.3 km³ has been reported (cf. Wang et al., 2011).

Here, the power function as $v = k \cdot h^a$ was considered since it allows some interpretations in the sense that the coefficients $k$ and $a$ are related to the hillside concavity and the openness of the half pyramid, which is taken as a first approximation for the reservoir (e.g., Liebe et al., 2005). Thus, given the linear relationship presented in Figure 2g, considering $\log (h)$ and $\log (v)$, the power function for TGR was estimated as $v = 1.0 \times 10^{-6} \cdot h^{3.355}$ with $R^2$ equals to 0.97. Then, the volume storage anomalies of TGR were estimated as shown in Figure 2h, even for the filling stages of the TGR. The results are similar to those presented in Wang et al. (2019), which were based on the power-law relationship presented in Wang et al. (2011).
3.1.4. Earth Observation derived precipitation

Quasi-global grids of precipitation estimates supplied by the Tropical Rainfall Measuring Mission (TRMM) Near Real-Time Precipitation (TMPA-RT) L3 1 day 0.25° x 0.25° V7 (TRMM 3B42RT v7) were retrieved from https://disc.gsfc.nasa.gov (GES-DISC, 2016). The 3B42RT v7 product (e.g., Huffman et al., 2007), which consists of daily precipitation averages, is one of the most relevant TRMM-related datasets for surface hydrology studies. TRMM precipitation (TRMM-PRC) is considered here for investigating the contribution of precipitation on the bulk of TWS mass gain within YRB.

3.2. Methodology

In this section, the overall methodology and computations are presented. Table 1 summarizes the main differences between the remotely-sensed and modeled products (cf. Section 3.1) that are necessary for the computations and investigations.

Table 1: Characteristics of the remotely-sensed and modeled products used in this study for inversion of GWS.

<table>
<thead>
<tr>
<th>Data</th>
<th>Source</th>
<th>Resolution</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stokes’s coefficients</td>
<td>CSR</td>
<td>-</td>
<td>monthly</td>
</tr>
<tr>
<td>TWC</td>
<td>NASA GES-DISC</td>
<td>0.25°</td>
<td>unitless</td>
</tr>
<tr>
<td>GWS</td>
<td>Universities of Kasel and Frankfurt</td>
<td>1.0°</td>
<td>monthly</td>
</tr>
<tr>
<td>altimetry</td>
<td>DAHITI</td>
<td>-</td>
<td>9-35 days</td>
</tr>
<tr>
<td>imagery</td>
<td>NASA/USGS-EROS</td>
<td>250-500 m</td>
<td>16 days</td>
</tr>
<tr>
<td>PRC</td>
<td>NASA GES-DISC</td>
<td>0.25°</td>
<td>kg/m²/day</td>
</tr>
</tbody>
</table>

3.2.1. Equal-area point-mass modeling

In this paper, the point-mass approach proposed by Forsberg & Reeh (2007) and further investigated by Baur & Sneeuw (2011) is used to invert the water masses causing the gravitational potential \( \delta V \) due to the gravitational attraction of those masses. This approach is somewhat like the so-called mass concentration (mascon) solution as pointed out by Baur & Sneeuw (2011). The Newton integral for the gravitational potential of infinitesimal surface elements \( dS \) with height \( \Delta r \) (i.e., the equivalent water thickness) is given as

\[
\delta V(r, \varphi, \lambda, t) = \frac{G \varrho_w}{\ell} \int \int_S \frac{\Delta r(\varphi', \lambda', t)}{\ell} dS,
\]

In Eq. (3), \( G \) is Newton’s gravitational constant \( (6.67259 \times 10^{-11} \text{ m}^3 \text{kg}^{-1} \text{s}^{-2}) \), \( \varrho_w \) is the density of water (position-dependent, which can be considered constant \( \sim 1,000 \text{ kg/m}^3 \)), and \( \ell \) is the Euclidean distance between the (attracted) computation point \( P(r, \varphi, \lambda) \) and the running integration (attracting) point \( Q(r', \varphi', \lambda') \) defined by

\[
\ell = \sqrt{r^2 + r'^2 - 2rr' \cos \psi},
\]
with the spherical distance given by the al-Kshi formula as:

$$\cos \psi = \sin \varphi \sin \varphi' + \cos \varphi \cos \varphi' \cos(\lambda - \lambda').$$  (5)

The numerical evaluation of Eq. (3) relies on a mass discretization, where the integration domain $S$ is decomposed into elementary spherical panels $s_k \in S$, assuming a constant surface density, $\mu_k = \Delta r_k \cdot \rho_w$, over each compartment. Hence, the effect of the whole mass distribution can be approximated by the sum of the impact over all individual compartments as:

$$\delta V(r, \varphi, \lambda, t) = GR^2 \sum_k \mu_k(t) \Delta \varphi_k \Delta \lambda_k \frac{\cos \varphi_k}{\ell_k}.$$  (6)

According to the requirements of the specific application, the decomposition can be performed using different types of compartments (e.g., templates, blocks). Here, the whole Earth was divided into 41,334 equal-area compartments covering its entire surface. The size of the panels is equivalent to a quadrangle of $1^\circ$-by-$1^\circ$ defined at the equator with an area of approximately 12,367.2 km$^2$. Partition of a sphere into equal-area grids can be found in, for example, Gringorten & Yepez (1992), which is the basis of the partition used here.

However, while computing the gravitational potentials using Eq (2), the functional model in Eq. (6) is not correct since there are spectral inconsistencies between the observations generated by means of Eq. (2) and the model as per Eq. (6). This means that the observations that are generated using Eq (2) are band-limited (expanded up to degree and order 96) and it is necessary to apply a low-pass filter to the model. Here, the filter was considered by using the decomposition of the reciprocal distance ($1/\ell_k$) appearing in Eq (6) as (Hotine, 1968, p. 310):

$$\frac{1}{\ell_k} = \frac{1}{r} \sum_{n=0}^{n_{\text{max}}} \left( \frac{R}{r} \right)^n P_n(\cos \psi),$$  (7)

where $P_n(\cos \psi)$ is the Legendre polynomial of degree $n$. That is, the upper limit of the summation in (7) is considered up to degree 96, just like it has been considered in Eq. (2). This is considered for both, forward and inverse modelings in order to bring the datasets to the same spectral resolutions. In this context, the modified version of Eq. (6) is used as the functional model, and the inverse problem is solved as per Eq. (10) with parameters $\mu_k$ (apparent surface densities) estimated at each panel $k$ at monthly time step. This has not been considered in previous works (e.g., Forsberg & Reeh, 2007; Baur & Sneeuw, 2011) while a different view on this problem has been addressed by Ran et al. (2018).
3.2.2. Parameter estimation

In the inverse modeling approach, the observed gravitational potential $\delta V$ over the region that the GRACE satellites overfly in their orbit can be used to compute the unknown parameters $\mu_k$ at the Earth’s surface. That is, the amount of water expressed in kg/m$^2$ at each equal-area panel. Suppose that from a linear forward model $Gm$, as right hand side (RHS) of Eq. (6), it is required to find the solution with the smallest $\|Lm\|$, which comes close enough to matching the data. This problem is formulated in terms of minimizing the misfit as a damped least squares problem (Aster et al., 2005, p. 192):

$$\min \|Gm - d\|^2_2 + \alpha^2 \|Lm\|^2_2,$$

(8)

where $d$ is the ($m \times 1$)-vector that contains the observations ($\delta V$), the ($n \times 1$)-vector $m$ takes in the unknown parameters ($\mu_k$, that is, TWS for the $k$-th panel), the ($m \times n$)-matrix $G$ is the coefficient matrix (or design matrix), and $\alpha$ is the regularization parameter.

However, the determination of the unknown parameter $\mu_k$ from the gravitational potential observed in space is an ill-posed problem ($G$ is rank-deficient and the normal matrix $G^TG$ is singular). Considering $L = I$ in Eq. (8) leads to the so-called zeroth-order Tikhonov regularization. Taking the partial derivatives w.r.t. the unknown parameters $m$ into account, the Tikhonov-regularized least squares solution is defined by:

$$(G^TG + \alpha^2I)m = G^Td.$$

(9)

The estimated parameters of the regularized solution can then be obtained by:

$$m_{\alpha} = (G^TG + \alpha^2I)^{-1}G^Td.$$

(10)

The regularization parameter, $\alpha$ (i.e., a small positive number), which is not known a priori, can be estimated based in a variety of heuristic approaches proposed to determine its optimal value. An important point is that, when plotted on a log-log scale, the curve of the optimal values of $\|m_{\alpha}\|_2$ and $\|Gm_{\alpha} - d\|_2$ often describes an L-like shape (Aster et al., 2005, p. 91). In this work, it was used the L-curve criterion proposed by Hansen (2000).

3.2.3. Leakage correction

An issue associated with estimated mass variations from GRACE data is the signal leakage. In the context of GRACE monthly solution, leakage effects are mainly due to the band-limited nature of the representation of the field in terms of Stokes’s coefficients and the post-processing procedures to reduce the stripes (Baur et al., 2009). The latter might not be an issue here since the Stokes’s
coefficients have not been filtered. Spatially, leakage manifests as signals spreading out from the source region (e.g., Greenland) to its surrounding. Consequently, the accuracy of surface mass changes derived from GRACE depends on the ability to identify, quantify, and remove leakage signals. There are many methods available to deal with leakage effects (see, e.g., Swenson & Wahr, 2002; Chen et al., 2006; Klees et al., 2008; Baur et al., 2009; Guo et al., 2010; Tang et al., 2012; Wiese et al., 2016; Vishwakarma et al., 2017). In this study, it was considered an approach that can be applied to a global set of surface densities covering the continents and that relies only on the GRACE data itself.

For this purpose, the approach proposed by Tang et al. (2012) was considered and modified in such a way that no filtering is needed. Following Wahr et al. (1998), the surface density \( \mu \), at a particular epoch \( t \), can be represented in terms of a sum of the respective Stokes’s coefficients, \( \Delta \hat{C}(t)_{nm} \) and \( \Delta \hat{S}(t)_{nm} \) given in kg/m\(^2\):

\[
\mu(\varphi, \lambda, t) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} \bar{P}_{nm}(\sin \varphi) \times \left[ \Delta \hat{C}(t)_{nm} \cos(m \lambda) + \Delta \hat{S}(t)_{nm} \sin(m \lambda) \right].
\]  

(11)

The coefficients in Eq. (11) are defined by (Wahr et al., 1998):

\[
\begin{align*}
\left\{ \Delta \hat{C}(t)_{nm} \right\} & = \frac{1}{4\pi} \int \int_{\Omega} \mu(\varphi, \lambda, t) \bar{P}_{nm}(\sin \varphi) \left\{ \cos(m \lambda) \sin(m \lambda) \right\} d\Omega, \\
\left\{ \Delta \hat{S}(t)_{nm} \right\} & = \frac{1}{4\pi} \int \int_{\Omega} \mu(\varphi, \lambda, t) \bar{P}_{nm}(\sin \varphi) \left\{ \cos(m \lambda) \sin(m \lambda) \right\} d\Omega,
\end{align*}
\]

(12)

where \( d\Omega = \cos \varphi d\varphi d\lambda \) is the surface element on the unit sphere. Considering the surface densities in Eq. (11), it is necessary to carefully distinguish between apparent surface densities \( \mu' \), as obtained from GRACE harmonic solutions and land surface densities \( \mu'_{\ell} \), which are corrected for leakage and describe the mass variations only over the continents.

The integral in Eq. (12) can be discretized considering only the spherical panels on the land and substituting it in Eq. (11). By using the trigonometric identity product-to-sum, this results in

\[
\mu'(\varphi_k, \lambda_k, t) = \frac{1}{4\pi} \sum_{i=0}^{1} \sum_{n=0}^{\infty} \sum_{m=0}^{n} \mu'_i(t) \cos \varphi_i \Delta \varphi_i \Delta \lambda_i \times \cos[m(\varphi_i - \varphi_k)] \bar{P}_{nm}(\sin \varphi_i) \bar{P}_{nm}(\sin \varphi_k).
\]  

(13)

In Eq. (13), \( \mu' \) are the apparent surface density values at each panel \( k \) of the equal-area partition. Consequently, 41,334 apparent surface density values are used as observations in the left hand side (LHS) of Eq. (13). The unknowns are the land surface densities \( \mu'_{\ell} \), which are associated with the 13,567 equal-area panels lying within the continents. However, as pointed out by Tang et al. (2012), the land surface densities can also contain some signals from offshore, for example, seismic events like the
2004 Indian Ocean earthquake, which is clearly depicted in the GRACE products (e.g., Chen et al., 2007). In such a case, a land mask defined by coastlines needs to be extended to include the correspondent mass changes as it was done here since the contributions of, for example, earthquakes, were not reduced from GRACE observations (Section 3.1.1).

3.2.4. “Tie-in” the sub-domains of TWS

Regarding the signal “tie-in” procedure of the sub-domains of the TWS-budget equation, as per Eq. (1), a statistical approach to support TWS analysis was reported by Forootan et al. (2014) and is extended here to account for groundwater. The process considers a priori spatial patterns of terrestrial and surface water storage changes to separate the TWS into its main sub-domains properly. Their study and other region-specific statistical analyses of GRACE-derived TWS (e.g., Ndehedehe et al., 2016) drew insights from the earlier work of Schmeer et al. (2012), who decomposed the mass signals within GRACE monthly gravity field models through principal component analysis (PCA). Although the water budget scheme proposed for this study is similar to previous studies (cf. Forootan et al., 2014), and the a priori spatio-temporal patterns are as suggested by Schmeer et al. (2012), that is, using PCA is however considered.

In this context, the water storage given by Eq. (1) can be decomposed as:

\[ T = c_G V_G^T + c_S V_S^T + c_T V_T^T + c_C V_C^T + c_R V_R^T, \]

where \( c \) and \( V \) are matrices with dimensions \( t \times k \) and \( n \times k \) describing the series coefficients representing the temporal evolution (i.e., the principal components PCs) and the orthonormal spatial base functions representing the spatial patterns (i.e., the empirical orthogonal functions EOFs), respectively. Since the water stored in soil, ice/snow, and canopy are those derived from GLDAS-Noah, Eq. (14) can be shortened to:

\[ T = c_G V_G^T + c_M V_M^T + c_R V_R^T, \]

where the subscript M stands for model, that is, TWC derived from GLDAS-Noah hydrologic model (cf. Section 3.1.2). Using singular value decomposition (SVD) a \( t \times n \) matrix \( Z \) (e.g., soil moisture) can be factored into:

\[ Z = U \cdot S \cdot V^T, \]

where the EOFs are given by \( V \), a \( n \times k \) orthogonal matrix, the PCs are given by \( U \cdot S \) which are respectively a \( t \times k \) orthogonal matrix and a \( k \times k \) diagonal matrix with non-negative diagonal elements called singular values.

However, the coefficients appearing in Eq. (15), that is, \( c_G \), \( c_M \), and \( c_R \), can be estimated by
least-squares given the observations $nL_t = T^T$, that is, the GRACE-derived TWS fields, and an a priori information for each variable describing the components of the TWS-budget as given by (1). Noteworthy, the first term in the RHS of Eq. (15) has not been considered by Forootan et al. (2014) since their contributions deals with the reconstruction of TWS given GLDAS-TWC and EO-SWS. The least squares problem such as

$$X = (J^T W J)^{-1} J^T W L,$$

(17)
can be used to estimate the coefficient of each water storage compartment of the TWS-budget equation given by $X$ with dimension $K \times t$. The matrix $J$, $n \times K$, is the coefficient matrix in which $K$ is the total number considered of modes to be estimated considered the individual variables in Eq. (15), that is, $K$ is the sum of $k_G$ (number of modes to compress groundwater), $k_M$ (number of modes for GLDAS-Noah water content), and $k_R$ (number of nodes of inland reservoirs) which has the following configuration:

$$J = \begin{bmatrix}
v_1^G & \cdots & v_{k_G}^G \\
v_{k_G+1}^G & \cdots & v_{2k_G}^G \\
\vdots & \ddots & \vdots \\
v_{Nk_G}^G & \cdots & v_{Nk_G}^G \\
v_1^M & \cdots & v_{k_M}^M \\
v_{k_M+1}^M & \cdots & v_{2k_M}^M \\
\vdots & \ddots & \vdots \\
v_{Nk_M}^M & \cdots & v_{Nk_M}^M \\
v_1^R & \cdots & v_{k_R}^R \\
v_{k_R+1}^R & \cdots & v_{2k_R}^R \\
\vdots & \ddots & \vdots \\
v_{Nk_R}^R & \cdots & v_{Nk_R}^R \\
\end{bmatrix},$$

(18)

The matrix $X$ with the solutions, that is, the estimated coefficients $\hat{c}_G$, $\hat{c}_M$, and $\hat{c}_R$, is given by

$$X = \begin{bmatrix}
c_1^G & \cdots & c_{k_G}^G \\
\vdots & \ddots & \vdots \\
\cdots & \ddots & \cdots \\
c_1^M & \cdots & c_{k_M}^M \\
\vdots & \ddots & \vdots \\
\cdots & \ddots & \cdots \\
c_1^R & \cdots & c_{k_R}^R \\
\vdots & \ddots & \vdots \\
\cdots & \ddots & \cdots \\
\end{bmatrix},$$

(19)

which contains the adjusted PCs for the respective water compartments shown in the RHS of Eq. (15). In Eq. (17), $W$ is the weight matrix where the errors for the GRACE-TWS (the observations) were considered. Basically, the errors were estimated by applying a 13-month moving mean to the de-seasoned series as proposed by Scanlon et al. (2016).

This approach is named here as “tie-in” and not signal separation since the aim is to adjust the subdomains appearing in the TWS-budget equation, GWS, TWC, and SWS, to TWS, which is regarded
as the observation. However, there is a problem here since neither the temporal (PCs) nor the spatial (EOFs) patterns of groundwater fields are available since they are unknown in the problem. Hence, in an attempt to have a preliminary approximation for the GWS, Eq. (1) was considered given the available dataset, and it is referred here as “observed” GWS. Another option would be to use the GWS fields derived from the WaterGAP (Section 3.1.2.2); however, it is used here only for assessment purposes.

3.2.5. Mode of water storage

3.2.5.1. Quantile function of storage (QFS).

The quantile function of storage (QFS) is a plot that shows the percentage of time that storage in a basin (or region) is likely to equal or exceed some specified value, in the water storage context, zero since the long-term mean is subtracted from monthly values. In Awange et al. (2014), the QFS has been named as TWS duration curve (TDC), which is probably inspired in its usage in hydrology known as flow duration curve (e.g., Smakhtin, 2001). The QFS can be used to express the percentage of time a basin storage can be expected to exceed a normal condition (zero) that exceeded some percent of the time (e.g., 50% of the time). Such information is useful to characterize the ability of a basin to retain storage of various magnitudes, which are due to the sub-domains of the TWS-budget equation as can be seen from Eq. (1). The probability of exceedance (probability that a given storage will be equaled or exceeded ratio of time) is computed as (Tourian et al., 2013):

\[ P = \frac{M}{n + 1} \times 100\%, \]  

(20)

which is known as the Weibull formula (cf. Helsel & Hirsch, 2002, p. 23). In Eq. (20), \( M \) is the ranked position on the listing (the data is sorted, ranked, from the largest to the smallest value), starting with 1 for the largest monthly TWS (or another field), involving \( n \) number of months (e.g., 168 months).

An interesting information of the QFS is its slope \( S_{\text{QFS}} \) (negative) calculated considering the interval between the 33rd (\( S_{33\%} \)) and 66th (\( S_{66\%} \)) storage percentiles (cf. Sawicz et al., 2011), which shows an almost linear behavior. This can be computed as (Sawicz et al., 2011):

\[ S_{\text{QFS}} = \frac{S_{33\%} - S_{66\%}}{66 - 33}, \]  

(21)

which can be interpreted as high values of \( S_{\text{QFS}} \) indicate a variable storage regime and, conversely, low values indicate a more damped response, which could be associated with the year-round rainfall and or dominance of the groundwater contribution, and or degradation of permafrost regions, etc. Furthermore, several factors impact the \( S_{\text{QFS}} \) of GWS, e.g., hydro-geological setting, aquifer flow, storage types, storage potential, recharge rates, and its mechanisms (cf. Awange et al., 2014).
3.2.5.2. Storage deficit (SD).

Using the long-term median, maximum and minimum GWS (can be TWS, or any other of its sub-domains), monthly percentage water storage deficit (or excess) can be computed as (Narasimhan & Srinivasan, 2005):

\[
SD = \frac{S_i - \text{median}(S_j)}{\text{max}(S_j) - \text{min}(S_j)} \cdot 100%.
\] (22)

In Eq. (22), \(S_i\) is the monthly water storage anomaly and \(\text{median}(S_j), \text{max}(S_j), \text{min}(S_j)\) are, respectively, the long-term median, maximum, and minimum monthly storage anomaly (i.e., with the mean removed from each monthly value of \(S\)) and with \(i = 1\) to 168 (168 months from Jan 2003 to Dec 2016) and \(j = 1\) to 12 (12 months of the calendar year).

3.2.6. Terrestrial water-budget

Considering the area-weighted time series of TWS (represented by \(T\)) for YRB, where daily precipitation are available (Section 3.1.4), the terrestrial water-budget equation (different from what has been called TWS-budget equation, cf. Eq. 1), can be used as (cf. Sneeuw et al., 2014):

\[
\frac{dT(t)}{dt} = P(t) - E(t) - R(t),
\] (23)

where \(R\) denotes the river discharge, \(P\) is precipitation, \(E\) represents the actual evapotranspiration, and \(dT/dt\) is the rate of water storage changes (flux quantity) over a basin. To directly compare GRACE water storage change with the other quantities, the water storage changes \(T(t)\) must be differentiated, or the RHS of Eq. (23) needs to be integrated. Thus, in this study, the area-weighted daily precipitation series can be integrated for comparison with GRACE-TWS and GWS where \(P\) and \(ET\) were ignored in Eq. (23) in a similar approach used, for example, by Crowley et al. (2006).

Integrating Eq. (23) yields:

\[
T(t) = \int_{t_1}^{t_N} [P(t') - E(t') - R(t')] dt',
\] (24)

where \(t_1\) and \(t_N\) are the first and the last day of the respective month (\(N\) being the number of days). Considering the daily precipitation (the same would hold for \(E\) and \(R\)), the trapezoidal method provides:

\[
\int_{t_1}^{t_N} P(t') dt' \approx \frac{t_N - t_1}{2N} \sum_{n=1}^{N} (P(t_n) + P(t_{n+1})),
\] (25)

which requires a \(N + 1\) evenly spaced number of days (i.e., from the first day of the current month until the first day of the sequent month). Equation (25) is used to integrate the daily precipitation at each respective months in order to be compared with the storage results.
4. Results and Discussion

This section presents the main results based on the datasets summarized in Table 1 and used in the steps depicted in the flowchart in Figure 3.

4.1. Inverted water storages and modes of variabilities

To properly invert the GWS fields, remotely-sensed and modeled datasets underwent the same inversion scheme as the GRACE in which the respective forwarded gravitational potentials at 500 km altitude (approximately the GRACE’s orbit) were inverted into the respective water masses considering point-mass approach (Sections 3.2.1 and 3.2.2). The point-mass was adopted to represent the surface masses at each equal-area panel, and the decomposition of the reciprocal distance was proposed to account for spectral inconsistencies between each of the different datasets and GRACE, which is band-limited (Sections 3.1.1 and 3.2.1). This offers an advantage in the sense that in the inverse modeling, the different types of observations can be considered at the same spectral resolution, which has not been considered
in previous contributions dealing with estimations of GWS (e.g., Rodell et al., 2009). Furthermore, the equal-area cells, in comparison with equal angular grids, offer many advantages since it allows the direct correlation between different regions, among others.

Specifically, Eq. (2) was used to synthesize the gravitational potentials ($\delta V$) on a grid with 3-by-3 arc-degree at 500 km altitude given the GRACE harmonic solutions provided by CSR (Section 3.1.1) up to degree and order 96. The 3 arc-degree somewhat reflects the nominal resolution of GRACE data. The gravitational potential is supposed to be generated from the residual signal due to the TWS, and hence they are the observations ($d$) used in Eq. (10) to estimate the unknown parameters ($m$, i.e., TWS).

There was a total of 7,200 observations (i.e., synthesized gravitational potential), which were inverted into 41,334 apparent surface densities (i.e., $\mu'$, that is TWS expressed in kg/m$^2$) covering the whole Earth at each monthly time step. However, due to the leakage effect, especially the TWS signal spreading out from the continent into the oceans, a correction was considered as described in Section 3.2.3. Equation (13) was used to estimate 13,567 panels over the continents ($\mu^f_k$) given the 41,334 apparent surface densities (i.e., $\mu'$). Due to the missing periods in GRACE solutions, linear interpolation was carried out to fill the gaps in the TWS series. Consequently, 168 monthly solutions covering the period of Jan 2003 to Dec 2016 were considered for further analysis in this section, to estimate groundwater storage and to investigate the storage regime in Section 4.2.

Regarding the spatial patterns of TWS within YRB, Figures 4a and 4b show the root-mean-square (RMS) and the Sen’s slope considering the period from Jan 2003 to Dec 2016 (14 years). Overall, the highest amplitudes depicted by the RMSs are located at the south of the Lower reach of YRB (Figure 4a), which is a region characterized by relatively high rainfall amplitudes (e.g., Wang et al., 2019). The southeast part of the Upper YRB, an area with its eastern portion that belongs to the Third Pole, high variability relative to the other portions of the basin are also noted. The Sen’s slope estimates for the TWS series at each mascon within YRB along with statistically significant slopes (x-marks) based on the Mann-Kendall test at 95% confidence level are indicated (Figure 4b). The spatial distribution of the Sen’s slope shows an overall water gain at the Middle and Lower reaches whereas the most significant water loss is seen at the Upper reach of YRB. There is also a considerable increase in TWS over the north-western section of the basin (Figure 4b).

The total water content (TWC, the sum of all layers of soil moisture, ice/snow, and canopy water storages) available in GLDAS-Noah, given at a resolution of 0.25-by-0.25 arc-degree (Section 3.1.2), was used to forward the gravitational potential in a 3-by-3 arc-degree at 500 km covering the whole Earth using Eq. (6). Noteworthy, results based on Eq. (6) considering the reciprocal distance given by Eq. (8), which was expanded up to degree 96, are band-limited as the results based on Eq. (2) for GRACE harmonic solutions. Then, as for the GRACE case, the gravitational potentials due to the water masses
Figure 4: Panels (a), (c), and (e) show the RMSs of the GRACE-TWS, GLDAS-TWC, and EO-SWS, respectively regarding the period of Jan 2003 to Dec 2016, in which the strength of each field can be inferred. Panels (b), (d), and (f) show the Sen’s slopes where the significant trends are identified with x-markers considering the Mann-Kendall test at 95% confidence level. The line in grey depicts the southeast boundary of the Third Pole, which is approximated by the isoline with elevation larger than 2000 masl.

as per GLDAS-Noah were used as observations in Eq. (10). Next, the inverted 41,334 mascons (apparent) were used in Eq. (13) to estimate the 13,567 mascons over the continents, which were corrected for the leakages. Figures 4c and 4d show the RMSs and Sen’s slopes, respectively, over YRB, which were based on the 168 monthly solutions (from Jan 2003 to Dec 2016).

The GLDAS-Noah TWC (hereafter GLDAS-TWC) presents almost the same spatial patterns of the RMS signals (Figure 4c) similar to those of GRACE-TWS (Figure 4a) whereas the amplitudes are much lower as expected since there are differences in the water contents in the two products. The mean of the differences in the RMSs of GRACE-TWS and GLDAS-TWC is 24.5 kg/m$^2$, and on average, GLDAS-TWC underestimates the TWS signals by approximately 35%. For example, the RMSs of the GLDAS-TWC show the same regions of high variability as those depicted by the RMSs of the GRACE-TWS, specifically, over the southern part of the Upper YRB. Furthermore, the south part of the Lower reach of YRB is also characterized by relatively high variability driven by rainfall. Regarding the Sen’s slope of GLDAS-TWC, there is a slight increase in moisture content at the south part of the Middle and Lower reaches of YRB (Figure 4d). However, most parts of YRB are characterized by a steady-state
changes of GLDAS-TWC shown by the not significant trends (at 95% confidence level) as depicted in Figure 4d. Indeed, linear patterns of soil moisture series over YRB regarding the period from Jan 2005 to Dec 2011 show values of $1.1 \pm 0.2$ kg/m$^2$/yr based on in-situ and WGHM datasets as shown by Wang et al. (2019). Huang et al. (2013) have compared the TWC with precipitation and found that the long-term changes in the former are mainly driven by the latter over YRB. The averaged value of the Sen’s slope shown in Figure 4 depicts approximately 4.3 and 0.1 kg/m$^2$/yr for TWS and TWC, respectively. (Section 4.2 presents values based on the averaged time series.)

The monthly volumes presented in Figure 2 (see Section 3.1.3 for details) for Dongting and Poyang Lakes as well as Lake Tai and TGR were divided by their respective standard areas (the standard shape for the reservoirs are those produced by the Digital Chart of the World - DCW - and were used for computing the respective areas). Thereafter, the standard boundary of each reservoir was discretized into cells with 500 m-by-500 m (nominal resolution of the imagery-based datasets, Section 3.1.3). Next, the monthly changes in the gravitational potential, due to the surface water storage considering the four lakes within YRB, were forwarded employing Eq. (7) on a grid of 3 arc-degree at 500 km covering the whole Earth. This is somewhat how the GRACE observations would be impacted by the gravitational attraction due to the four lakes considered during the period of Jan 2003 to Dec 2016 if there would be no other mass variations within the region. Optimally, all inland reservoirs (e.g., rivers, lakes) covering the whole Earth would be necessary and, however, it was considered only the ones within the study area as an example how to account for the contributions of surface water storage as observed by EO satellites (EO-SWS).

As for the cases of GRACE-TWS and GLDAS-TWC, the EO-SWS monthly fields were inverted at a global set of 41,334 equal-area panels considering the gravitational potential grids at 500 km with a resolution of 3 arc-degree. The apparent EO-SWS fields, at a global set of 41,334 panels, were then corrected for leakage effects where 13,567 land equal-area panels were recovered. Figure 4e shows the RMSs of the EO-SWS for each panel within YRB, where the highest values are seen at the TGR location (approximately at 30.8°N and 111.0°E). Notably, the RMS signals cover an area larger than the TGR due to the inversion, and the magnitudes seem more substantial than those depicted in GRACE-TWS (Figure 4a). However, the RMS signals for Dongting and Poyang Lakes, as well as Lake Tai present relative small variabilities regarding those of TGR. Sen’s slope for EOS-SWS is also presented (Figure 4f), and the overall mass gain is seen over the TGR, which also spreads out over the whole YRB with most rates being significant (Mann-Kendall test at 95% confidence level). So, why the same signal is not seen in GRACE results for RMSs and linear trends as respectively summarized at Figures 4a (or 4b) and 4c (or 4f)? Of course, GRACE senses the bulk mass changes mainly induced by the masses just below the space crafts - distant masses also contribute proportionally to the reciprocal distance appearing in Eq. (6) - and mass deficits due to changes in other sub-domains of TWS might contribute to sap the
signal. For example, groundwater at the Jianghan Plain contributes to the bulk trends of TWS shown in Figure 4b since TWC is not significant over that region. Similarly, discrepancies between GRACE-TWS and EO-SWS are seen over Europe and Asia w.r.t. the mass changes of Caspian Sea (cf. Forootan et al., 2014).

Principal component analysis (PCA) was used to separate the spatiotemporal patterns of GRACE-TWS fields (the same holds for GLDAS-TWC and EO-SWS fields) into principal components (PCs) and the associated empirical orthogonal functions (EOFs) as shown in Eq. (16). Figure 5 shows the loadings for the first four PCs for GRACE-TWS, GLDAS-TWC, and EO-SWS, as well as their respective spatial patterns given by the EOFs. For instance, Figure 5a shows the temporal evolution associated with the first PC depicting the seasonality inherent to the TWS series, which is mainly driven by the variations in rainfall at the southeast part of YRB, as shown by the accompanying EOF (Figure 5b). The PCs were scaled by their maximum values to be unitless, and the EOFs were scaled by the maximum values of their corresponding PCs to represent anomaly maps of water storage in kg/m$^2$. This EOF coincides with the spatial patterns depicted by the RMS of TWS signals (Figure 4a), and the first PC (Figure 5a) is characterized by a dominant annual peak and with a long-term change (significant at 95% confidence). Likewise, the second PC is also characterized by an evident annual peak (Figure 5c), and its associated EOF (Figure 4d) shows the predominant pole centered at the south part of the Upper YRB. This region is also shown in Figure 4a with high values of RMSs, indicating high variability of TWS within that region. However, the long-term changes (significant at 95% confidence) characteristic of this portion of YRB (belongs to the Third Pole) is seen on the third PC (Figure 5e) as depicted by the corresponding EOF (Figure 5f). Figure 5g shows the fourth PC of GRACE-TWS signals, which is characterized by a relatively low power with a two to four years period in which the corresponding EOF (Figure 5h) depicted processes concentrated at the south portion of YRB, and a significant long-term change is also evident.

The PCA was also applied to GLDAS-TWC. The first two PCs for GLDAS-TWC are characterized by annual variations where the first PC (Figure 5i) represents the variability of TWC at the southern part of the Upper YRB as shown by its EOF (Figure 5j). The second PC (Figure 5k) depicts the signal over the southeast portion of the Lower YRB, as demonstrated by its EOF (Figure 5l), which is a region characterized by high amplitudes rainfall relative to the other parts of the basin (Wang et al., 2019). The semi-annual signal is predominant in the fourth PC (Figure 5o), which also presents significant long-term changes, with its corresponding EOF (Figure 5p) showing variability at the Middle YRB. The third PC (Figure 5m) is predominant by a two-year period with a long-term change (significant at 95% level), and its EOF shows the signal pole to be placed at the northeast part of the Middle YRB.

The same was carried out for the EO-SWS fields in which the PCA correctly separated the signals
Figure 5: The top-rows show the PCA for GRACE-TWS in which the panels (a)-(h) contain the PCs and EOFs for the first four modes. The mid-rows show the PCA for GLDAS-TWC in which the panels (i)-(p) show the PCs and EOFs for the first four modes. The same holds for the bottom-rows, panels (q)-(x) present the PCs and EOFs for EO-SWS. The PCs are scaled by their respective maximum values to be unit-less, and the EOFs are scaled by the respective maximum values of their corresponding PCs to represent anomaly maps of water storage in kg/m².
due to the different sources. For instance, Figure 5q shows the first PC, in which the corresponding EOF (Figure 5r) depicted the pole centered at the TGR location. The time evolution of the first PC (Figure 5q) resembles the volume storage of TGR presented in Figure 2h where the PC characterizes the stages of water impoundment during the filling of the reservoir. (Note, the signal of the PCs is opposite, and, however, it must be considered together with the associated EOF.) In the respective order, the second, third, and fourth PCs depicted the volume storage of Dongting Lake, Poyang Lake, and Lake Tai, which can be seen from their EOFs shown in Figures 5t, 5v, and 5x. Only the PCs of TGR (Figure 4q) and Dongting Lake (Figure 5s) show significant long-term changes.

The total variance explained in the first four modes (see the title of the panels in sub-plots of Figure 5) for GRACE-TWS, GLDAS-TWC, and EO-SWS are about 83.2%, 85.8%, and 100.0%, respectively (Figure 5). For the surface water storages (EO-SWS), the contributions of Poyang Lake and Lake Tai are relatively insignificant (approximately 0.4%) compared to the water storage of TGR. These four modes of EO-SWS were considered in the inversion of groundwater storage, that is, the sub-matrix appearing in matrix $J$, see Eqs. (17) and (18), with the elements $N v^R_k$ composing a sub-matrix of dimension 142-by-4, in which $N$ is equal to 142 equal-area panels “within” YRB’s domain. Nevertheless, for the GLDAS-TWC, more modes must be considered in order to explain at least 99% of the total variance thus providing a relatively good approximation of the observed signals. Figure 6a shows the accumulative variance considering the first 53 modes for GLDAS-TWC in which 17 modes explained approximately 99.15% of the total variance. Consequently, the sub-matrix appearing in matrix $J$, see Eqs. (17) and (18), with the elements $N v^M_k$, is of dimension 142-by-17.

![Figure 6](image)

Figure 6: (a) Eigenvalues of GLDAS-TWC depicted by the stems, the curve in blue shows the accumulated variance (i.e., the sum of the eigenvalues) in which a percentage of 98.15% of the total variance is achieved by the first 17 eigenvalues indicated by the dashed red line. (b) Shows the same as (a) however for the approximated GWS considering Eq. (1) in which the first 21 eigenvalues explain 99.14% of the total variance.
PCA was applied to the resulting fields $G$, that is, the “observed” GWS as given by Eq. (1). Figure 6b shows the accumulative variance for the preliminary version of the GWS fields in which the first 21 modes explained about 99.14% of the total variance. Consequently, in the least-square problem as in Eq. (17), the coefficient matrix $J$ is of dimension 142-by-42 where the rows are the number of the geographical locations (the equal-area spherical panels) and the columns are the number of modes considering the respective PCs for “observed” GWS equals to 21, GLDAS-TWC equals to 17, and EO-SWS equals to 4. Consequently, the solution matrix, $X$, is of dimensions 42-by-168 where the number of rows is the number of the modes for the sub-domains (groundwater, TWC, and SWS) and columns the number of months in the period of Jan 2003 to Dec 2016 (168 months). Notably, considering the PCs (observed) and EOFs of the “observed” GWS in the inversion of the PCs (estimated) as per Eq. (17) implies that the EOFs for GWS do not change while the associated PCs are updated to minimize the residuals of the differences between the RHS and LHS of Eq. (15). At this point, the estimated PCs are not considered in further analysis, and their usage are limited only for reconstructing the sub-domains of TWS, including GWS.

Results from the parameter estimation, i.e., the estimated PCs given by the solution matrix $X$, are summarized in Figure 7a-c for the first modes for each sub-domain (i.e., GWS, TWC, and SWS). Overall, the temporal evolution of the “observed” ($c_G$) and estimated ($\hat{c}_G$) PCs for GWS shows relatively good agreement between both series (Figure 7a). Significant differences are seen over early 2004 to 2006 as well as during 2007 (Figure 7a). The estimated PC for the sub-domain TWC (Figure 7b) agree relatively well with the “observed” PC (Figure 5a, repeated in Figure 7b for convenience) in terms of seasonality with reduction in the amplitudes. Since the spatial patterns (EOF) of the first mode for TWC did not change (Figure 5j), the temporal evolution of the estimated PCs is associated with the southern portion of the Upper YRB. Regarding the first mode of SWS sub-domain, Figure 5r showed the central pole at the TGR region and its estimated PC (Figure 7c) agrees with the “observed” one (Figure 5q, and repeated in Figure 7c) only for the period 2009 afterwards indicating that the temporal evolution of the estimated PC did not capture the stages of the filling of TGR (Figure 7c). The seasonality inherent in the storage of TGR (dominant mode) is well depicted by GRACE and the SWS variations of the filling period seems noisy. Furthermore, there are phases change which might be associated with the fact that the lakes are located the middle and lower reaches of he basin. Such comparisons are not intended to validate the estimated parameters but instead try to understand how the adjusted coefficients need to be changed in order to fit better to the sub-domains of TWS considering the observations provided by GRACE. Furthermore, Figure 7d shows the correlation coefficients between the “observed” and estimated PCs for the respective number of the modes considered for the respective sub-domains (i.e., 21 for GWS, 17 for TWC, and 4 for SWS). Overall, only the two first modes presented good correlation for GWS and TWC whereas for SWS only the PCs of Lake Dongting (2 modes) presented reasonable value for correlation.
Groundwater fields were reconstructed considering the estimated PCs (\( \hat{c}_G \)) - example for the first mode is shown in Figure 7a, and the correlations between estimated and “observed” PCs are shown in Figure 7d - and the “observed EOFs (\( V^T_G \)) using Eq. (16), re-written here as:

\[
\hat{G} = \hat{c}_G V^T_G, \tag{26}
\]

where \( \hat{G} \) is the estimated GWS in comparison with \( G \) in Eq (1). The averaged time-series for YRB was computed by only taking the mean of the 142 equal-area panels, and it is displayed in Figure 8a. Furthermore, groundwater in-situ data from 136 monitoring wells (Figure 1) within YRB covering the period of Jan 2005 to Dec 2010 was used for assessment purposes. The in-situ groundwater dataset was digitally reconstructed from the time-series available in Wang et al. (2019) using the DigitizeIt digitizer software (Bormann, 2016). The overall error of the extracted groundwater time-series due to the digitization process is approximately ±0.67 mm (at 95% confidence level) based on closed-loop simulation. The extracted time-series was de-trended and a linear trend of 5.4 mm/yr added back to match the results provided by Wang et al. (2019). The “in-situ” groundwater series is shown in Figure 8a.

Overall, the results of the reconstructed GWS series for YRB (Figure 8a) performed reasonably well relative to the results based on the mere application of Eq. (1). However, there are some differences between the reconstructed and the “observed” - the one based on Eq. (1) - especially over the period of Jul 2013 to Jul 2014. The differences between the two series mainly reflect the behaviors of the first
modes of “observed” and estimated PCs, as shown in Figure 7a. Figure 8a also depicts the time series of groundwater estimations based on WaterGAP (Section 3.1.2.2), which good agreement is seen w.r.t. the in-situ and reconstructed series. Comparisons based on standard deviation of the differences between in-situ and the “observed and reconstructed times series show 13.97 and 10.86 kg/m², respectively. For WaterGAP estimated groundwater, the comparison shows a value of about 14.84 kg/m². That is, a slight improvement of approximately 22% is seen while considering the reconstructed time series in terms of standard deviation. Correlation coefficient (CC) between the respective series show values of 0.55, 0.60 and 0.76 for WaterGAP, “observed, and reconstructed series, which shows an improvement of approximately 27% indicating a better performance while using the reconstructed groundwater series. The NashSutcliffe efficiency (NSE) presents a degree of statistical similarity between in-situ measurements and the evaluated series where “observed and reconstructed show values of 0.15, 0.25, and 0.55, respectively. The reconstructed GWS series show an improvement of about 120% in terms of NSE is seen regarding to that based on Eq. (1).

4.2. Storage regime of YRB

Regarding the storage regime of YRB, it is essential to look at the integrated precipitation time series (TRMM-PRC), which is shown in Figure 8b since it is the primary source of water within YRB based on the terrestrial water-budget (Eq. 24). As expected, integrated precipitation (anomalies) generally bounds TWS (cf. Crowley et al., 2006) over intra-annual time scales, and there is a slight phase shift in some years (e.g., summer 2004) and the largest spike between the series occurs at a lag of one month with precipitation leading. On the other hand, TWS and TWC (GLDAS) present a proper agreement in
terms of phases and amplitudes; the only exception is seen in summer 2004 (Figure 8b). Generally, the maximum and minimum peaks of TWS are slightly higher than those of TWC, in which groundwater variations might play a role in the observed differences. Figure 8b also depicts the averaged surface water storages (EO-SWS) in which seasonal variations are seen after the filling of TGR 2009 afterward.

Overall, the four series present Sen’s slope of 0.92 (monthly precipitation), 4.59 (TWS), 0.20 (TWC), and 1.09 kg/m²/yr (SWS) indicating that approximately 3.30 kg/m²/yr might be due to groundwater storage (Table 2). Long-term changes based on monthly precipitation would explain only 20% in the best case, i.e., without considering the contributions of the sink terms (Eq. 24), for the overall increase in YRB’s TWS. From the study of Sun et al. (2017), it is possible to infer that the linear trend of precipitation is about 17% of that of the TWS series of YRB, albeit for the period of 2003-2014. Nevertheless, periods with unusually high/low rainfalls can produce positive/negative TWS anomalies (as for soil moisture), as shown in Figure 8b, and such effects generally persist through the months. Since quantifying TWS persistence (“memory”) would require multidecadal records of TWS, annual accumulated precipitation (Figure 8b) helps to understand that rainfall is the main driven of the wetness of YRB. For example, Sens slope of annual precipitation shows an increase of 19.05 kg/m²/yr (Table 2) and, disregarding dissipation through evaporation and other processes, rainfall plays a major role in TWS increase of 4.59 kg/m²/yr.

Table 2: Sen’s slopes for the different products regarding to the period from Jan 2003 to Dec 2016. The values between the brackets refer to Sen’s slope for the period from Jan 2005 to Dec 2011.

<table>
<thead>
<tr>
<th>Product</th>
<th>Observed Sen’s slope (kg/m²/yr)</th>
<th>Reconstructed</th>
<th>WaterGAP</th>
<th>in-situ</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRC</td>
<td>0.92a/19.05sb</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TWS</td>
<td>4.59</td>
<td>4.59</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TWC</td>
<td>0.20</td>
<td>2.69</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SWS</td>
<td>1.09</td>
<td>0.14</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GWS</td>
<td>3.30 (1.64)</td>
<td>1.76 (1.79)</td>
<td>0.48 (0.96)</td>
<td>- (5.42)</td>
</tr>
</tbody>
</table>

a Sen’s slope of monthly PRC.
b Sen’s slope of annual PRC.

The averaged time-series for reconstructed GWS (Figure 8a) shows an overall increase with a Sen’s slope of 1.76 kg/m²/yr from Jan 2013 to Dec 2016 (Table 2). Considering the “observed groundwater time series using Eq. (1), the Sen’s slope for the same period is 3.30 kg/m²/yr (Table 2). Notably, there is an overall agreement between the “observed and reconstructed groundwater time series (Figure 8a), whereas discrepancies are seen from 2011 afterward. Based on “in-situ data available throughout Jan 2005 to Dec 2010 (Figure 8a), Sen’s slope depicted an increase of approximately 5.42 kg/m²/yr for the same period. As expected, the linear trend agrees with the value estimated by Wang et al. (2019), who reported an overall increase of GWS within YRB of approximately 5.4 ±4.3 mm/yr (or kg/m²/yr, given the density of water as 1,000 kg/m³). For comparison purposes, the slopes are 1.64 and 1.79 kg/m²/yr for
the “observed, based on Eq. (1), and the reconstructed time series of groundwater storages, respectively and 0.96 kg/m²/yr for WaterGAP considering the same period (i.e., Jan 2003 to Dec 2010). The differences between the slope of “in-situ GWS series, regarding Jan 2003 to Dec 2010, and those derived from remotely-sensed and modeled datasets mainly reflect the distribution of the groundwater sites since they are located mainly at the Middle and Lower reaches of YRB (Figure 1) at which large rates of TWS increasing is seen (Figure 4b).

Considering the TWS time series (Figure 8b), exceedances were calculated based on Eq. (20). Figure 9a shows QFS for TWS, which the exceedance at the zero net shows that YRB gained water 45% of the period and lost it during 55% of the time. The same is observed for the reconstructed sub-domains of TWS (i.e., GWS, TWC, and SWS). The probability of change in the volume of TWS can be extracted by using the percentiles, for example, for less than 1% probability of occurrence, the maximum water gain is about 122.2 kg/m² for YRB. Conversely, the maximum water loss of -70.8 kg/m² is expected with a probability of occurrence of 99%. The 1% probability of exceedance 46.8 kg/m² while -27.1 kg/m² is seen for the groundwater loss at 99%. Furthermore, the slope of the QFS ($S_{\text{QFS}}$) can be computed from Eq. (21), which shows slopes of approximately -1.3 and -0.5 for TWS and GWS, respectively. These figures show that TWS has a fast response and small storage potential in comparison with GWS. This might be due to the soil moisture changes, which governs the storage potential of TWS. For example, at the seasonal time scale, TWS and TWC almost coincide (Figure 8b). Interestingly, the gradual slope of the QFS for SWS shows that this resource has high potential storage, which might be due to the successfully management of the surface water within YRB where its water retention capacity has increased.

![Figure 9](image-url) (a) Quantile function of storage (QFS) for monthly TWS and GWS over the period Jan 2003 and Dec 2017. (b) Cumulative SD for TWS and GWS.
Figure 9b shows the cumulative SDs (Yirdaw et al., 2008) for GWS computed from Eq. (22). Storage deficits are useful to compare deficit (surplus) values across seasons since the seasonalities inherent in GWS (or another compartment of TWS) were removed (Narasimhan & Srinivasan, 2005) as seen in Eq. (22). The cumulative SD represents the historical dryness (negative values) and wetness (positive values) in the changes in water storage in which a declining trend denotes a lasting SD, whereas a growing trend represents a surplus (positive SD). Figure 9b shows that the cumulative SD for GWS decreased from Jan 2003 until Jul 2008 and, despite some recovery of the dryness in early 2009 and 2011, GWS started to recover after only in latter 2011 which reached surplus in early 2016. Figure 9b also shows the response of precipitation to the changes reflected by the cumulative SD curve, which follows the same temporal evolution as GWS from 2008 afterward. However, while the cumulative SD curve for precipitation presents some oscillation from 2003 to latter 2007, the cumulative SD curve for GWS shows a steep decrease. This might be due to the fact that GWS in the current month depends on the GWS available from previous month, and the portion of groundwater recharge due to rainfall and interaction between surface and groundwaters in YRB.

The linear trends of GWS within YRB showed an overall increase during the GRACE-era (Table 2) and has also been shown in previous studies (see, e.g., Jiang et al., 2016; Xiang et al., 2016). Even though GWS showed an increase, the basin has experienced several drought periods (e.g., Zhang et al., 2016), which impacted the groundwater resources (Figure 9). For instance, the respective QFSs (storage duration curves) for TWS and GWS (Figure 9a) showed that 55% of the times, YRB lost the water stored as TWS and GWS. Furthermore, the slope of the QFS ($S_{QFS}$) showed that TWS a fast response and small storage potential relative to GWS, which might be because firstly, the water stored at the surface and soils reflect the dryness and secondly, GWS is more resilient than TWS. However, the droughts that impacted the TWS also translated to the reduction of GWS as shown by the cumulative SD (storage deficit) in Figure 9b, which depicted the temporal patterns in the historical dryness and wetness for YRB. The cumulative SD shows that the drought in YRB impacted the GWS until latter 2011 depicted by the rising limb of the curve (Figure 9b). This was a consequence of drought in YRB that started in early 2003 and persisted till 2012, with the worst severity registered during 2003-2005 and 2006-2009 (Zhang et al., 2016). Indeed, this has also be confirmed by Sun et al. (2018), who reported these two periods (Jan 2003 to Feb 2005 and Jun 2006 to Aug 2008) as the most extensive drought periods, lasting 26 and 27 months, respectively. Drought events over YRB have also been reported by Zhang et al. (2015), which the impacts over the Upper and Lower reaches of YRB have been found in Sep 2006 and Jun 2011 and have impacted all tributaries of Yangtze River (cf. Zhang et al., 2019a).

Droughts in the sub-basins of YRB are identified in Figure 5, for example, the three drought events shown by Zhang et al. (2019a) in the Jinsha River basin (source region of Yangtze River) is somewhat seen in Figure 5c, which shows the PC of the second mode of variability of TWS. There, it is possible to
see low peaks in 2004-2006, 2008-2010, and 2011, with the focal spatial pattern located in the southern part of the Jinsha River basin (Figure 5d). However, while droughts have impacted the Dongting Lake basin and Poyang Lake basin from summer 2003 to springer 2005, and Lake Tai basin from early 2003 to latter 2004 and from summer 2005 to summer 2006 (cf. Zhang et al., 2019a), the volumes of the respective lakes were not impacted by such events (Figure 2). Conversely, while drought seems to have not impacted the water level of the lakes, floods are depicted in Figure 2. For example, Poyang Lake basin in May-Aug 2010 experienced a flood peak as characterized by Sun et al. (2017), which is shown in Figures 2d and 5s.

5. Conclusions

The Yangtze River, which takes its first drop at the Earths Third Pole, plays an essential role in the economic and social developments of China. However, the Yangtze River is settled over complex physiographic provinces at which a vast section is located in mountainous and hilly areas implying that runoff limits the storage potential of the upper reaches of its basin (YRB). Yet, the abundant distribution of the water resources in YRB is increasing overall at a pace of 4.59 kg/m$^2$/yr, which reflects the storage potential of the middle and lower reaches of YRB. Thus, the main goal of this study was to understand the storage regime of the sub-domains of the TWS. Under such a scenario, it was investigated the wetness trend of YRB considering the maximum use of the available remotely-sensed (GRACE and EO satellites) and modeled (GLDAS) datasets considering a “tie-in signal approach, which optimally matches, in the least-squares sense, the different datasets to the GRACE observed TWS. And the results are summarized as follows:

1. It was found that TWC (soil moisture, ice and snow) increased at a rate of 2.69 kg/m$^2$/yr, SWS at a rate of 0.14 kg/m$^2$/yr, and GWS at a rate of 1.76 kg/m$^2$/yr. That is, considering the reconstructed sub-domains of TWS (GWS, TWC, and SWS), it was possible to close the TWS-budget equation at the inter-annual and seasonal time scales. Consequently, the sub-surface water storages are the dominant factors in the variations of TWS within YRB, which accounted for approximately 97% from Jan 2003 to Dec 2016. There is a significant increase in the annual precipitation in the basin at a rate of 19.05 kg/m$^2$/yr, and, therefore, increasing in the sub-surface storages might be due to this source rather than groundwater leakage (inter-aquifer flow). Furthermore, given the rates of TWS and precipitation (disregarding its dissipation terms like evaporation), it is recommended to increase the water storage facilities over the upper reaches of YRB to improve its overall water retention capacity, which is limited by the physiography.

2. The quantile function of storage showed that YRB lost groundwater 55% of the time during the study period and that TWS presented a fast response and small storage potential regarding GWS.
This might be associated with the fact that TWC was more prone to the drought events (mainly driven by soil moisture) that have reached YRB during the study period. Indeed, droughts impacted the GWS resources of YRB until yearly 2010, and, in the following years, the basin presented a recovery of GWS reaching water surplus in 2016 and afterward. Albeit the reconstructed GWS series was validated against “in-situ” data, which mainly reflected the differences at the seasonal time scales, a significant difference at the inter-annual scales was found. This might be due to the distribution of the stations mainly located in the regions with high groundwater potential (cf. Figure 1). Furthermore, in-situ data of soil moisture would be necessary to focus on increasing soil moisture since the reconstructed values presented a much more significant increase in comparison with the GLDAS-Noah data. Future work will thoroughly investigate the individual sub-domains of TWS, and to characterize their respective memories (persistence) based on in-situ and remotely-sensed datasets.

Despite the limitations in the present study, the results and information presented here could be of importance in the management and plan of the use of surface and sub-surface water resources in YRB of China.

Acknowledgments

V.G. Ferreira acknowledges the support from the National Key R&D Program of China (Grant No. 2018YFA0605402), the National Natural Science Foundation of China (Grant No. 41574001), and the Fundamental Research Funds for the Central Universities (Grant No. 2015B21014). The authors are grateful to the American and German taxpayers for affording GRACE mission and the University of Texas Center for Space Research (CSR) for the Level 2 products. We are also grateful to the data providers NASA (TRMM, GLDAS-Noah, and MODIS), DAHITI (altimetry products), and Dr. Hannes Miller Schmied for providing the groundwater storage from the WaterGAP model. We also would like to express our gratitude to the three Reviewers and the Editor Dr. Christian Herrera for the valuable comments and suggestions.

References


Bormann, I. (2016). Digitizeit: Digitizer software - digitize a scanned graph or chart into (x,y)-data. URL: https://www.digitizeit.de/.


CGS (2012). Groundwater serial maps of Asia - hydrogeological map, groundwater resources map, geothermal map.


Graphical abstract
Highlights

- Terrestrial water storage of Yangtze Basin grows overall at a pace of 4.59 mm/year.
- Reconstructed groundwater adds approx. 38% in the terrestrial water storage gain.
- Drought events impacted groundwater from 2003 to 2010, which recovered in 2016.
- Terrestrial water storage has a fast and small storage potential w.r.t. groundwater.
- Surface water has high potential storage due to its increased retention capacity.
Figure 4

(a) RMS of GRACE-TWS
(b) GRACE-TWS rates
(c) RMS of GLDAS-TWC
(d) GLDAS-TWC rates
(e) RMS of EO-SWS
(f) EO-SWS rates
Figure 5
(a) First PC for GWS

(b) First PC for TWC

(c) First PC for SWS

(d) Corr. coeff. between original and estimated PCs

Figure 7
Figure 8

(a) Groundwater storages of Yangtze Basin

(b) Precipitation and water storages (TWS, TWC, and SWS) of Yangtze Basin
Figure 9

(a) Quantile function of storage for TWS and its sub-domains in YRB

(b) Cumulative storage deficits for GWS and PCP in YRB