Water Resources Research

RESEARCH ARTICLE
10.1029/2020WR027556

Special Section:
The Quest for Sustainability of Heavily Stressed Aquifers at Regional to Global Scales

Comparison of Groundwater Storage Changes From GRACE Satellites With Monitoring and Modeling of Major U.S. Aquifers

Ashraf Rateb[1], Bridget R. Scanlon[1], Donald R. Pool[2], Alexander Sun[1], Zizhan Zhang[3], Jianli Chen[4], Brian Clark[5], Claudia C. Faunt[6], Connor J. Haugh[7], Mary Hill[8], Christopher Hobza[9], Virginia L. McGuire[9], Meredith Reitz[10], Hannes Müller Schmied[11], Edwin H. Sutanudjaja[12], Sean Swenson[13], David Wiese[14], Youlong Xia[15], and Wesley Zell[16]

[1]Bureau of Economic Geology, University of Texas at Austin, Austin, TX, USA, [2]Don Pool Hydrogeologist, LLC, Tucson, AZ, USA, [3]State Key Laboratory of Geodesy and Earth's Dynamics, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan, China, [4]Center for Space Research, University of Texas at Austin, Austin, TX, USA, [5]Water Resources Mission Area, U.S. Geological Survey, Reston, VA, USA, [6]California Water Science Center, U.S. Geological Survey, San Diego, CA, USA, [7]Lower Mississippi-Gulf Water Science Center, U.S. Geological Survey, Nashville, TN, USA, [8]Department of Geology, University of Kansas, Lawrence, KS, USA, [9]Nebraska Water Science Center, U.S. Geological Survey, Lincoln, NE, USA, [10]Hydrologic Remote Sensing Branch, U.S. Geological Survey, Reston, VA, USA, [11]Institute of Physical Geography, Goethe University Frankfurt and Senckenberg Leibniz Biodiversity and Climate Research Centre (SBiK-F), Frankfurt am Main, Germany, [12]Department of Physical Geography, Utrecht University, Utrecht, Netherlands, [13]National Center for Atmospheric Research, University of Colorado, Boulder, CO, USA, [14]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, [15]M. Systems Group at Environmental Modeling Center, National Centers for Environmental Prediction, College Park, MD, USA, [16]Integrated Modeling and Prediction Division, U.S. Geological Survey, Reston, VA, USA

Key Points:
- GRACE satellites track GW storage in U.S. aquifers based on good agreement between GRACE and in situ GW level monitoring (23,000 wells)
- Regional models capture GRACE GW storage in most aquifers, except the Mississippi Embayment, which overestimates GRACE GWS trends by approximately four times
- Global hydrologic models overestimate GW depletion compared with GRACE by ~2.4× in heavily depleted aquifers in the SW and SC United States

Supporting Information:
- Supporting Information S1

Correspondence to:
A. Rateb,
ashraf.rateb@beg.utexas.edu

Citation:
Rateb, A., Scanlon, B. R., Pool, D. R., Sun, A., Zhang, Z., Chen, J., et al. (2020). Comparison of groundwater storage changes from GRACE satellites with monitoring and modeling of major U.S. aquifers. *Water Resources Research*, *56*, e2020WR027556. https://doi.org/10.1029/2020WR027556

Received 21 MAR 2020
Accepted 31 OCT 2020
Accepted article online 5 NOV 2020

Abstract GRACE satellite data are widely used to estimate groundwater storage (GWS) changes in aquifers globally; however, comparisons with GW monitoring and modeling data are limited. Here we compared GWS changes from GRACE over 15 yr (2002–2017) in 14 major U.S. aquifers with groundwater-level (GWL) monitoring data in ~23,000 wells and with regional and global hydrologic and land surface models. Results show declining GWS trends from GRACE data in the six southwestern and south-central U.S. aquifers, totaling −90 km^3 over 15 yr, related to long-term (5–15 yr) droughts, and exceeding Lake Mead volume by ~2.5×. GWS trends in most remaining aquifers were stable or slightly rising. GRACE-derived GWS changes agree with GWL monitoring data in most aquifers (correlation coefficients, $R = 0.52–0.95$), showing that GRACE satellites capture groundwater (GW) dynamics. Regional GW models (eight models) generally show similar or greater GWS trends than those from GRACE. Large discrepancies in the Mississippi Embayment aquifer, with modeled GWS decline approximately four times that of GRACE, may reflect uncertainties in model storage parameters, stream capture, pumpage, and/or recharge rates. Global hydrologic models (2003–2014), which include GW pumping, generally overestimate GRACE GWS depletion (total: approximately −172 to −186 km^3) in heavily exploited aquifers in southwestern and south-central U.S. by ~2.4× (GRACE: −74 km^3), underscoring needed modeling improvements relative to anthropogenic impacts. Global land surface models tend to track GRACE GWS dynamics better than global hydrologic models. Intercomparing remote sensing, monitoring, and modeling data underscores the importance of considering all data sources to constrain GWS uncertainties.

Plain Language Summary The major U.S. aquifers provide an ideal system to assess GRACE (Gravity Recovery and Climate Experiment) satellite data. We compared GRACE groundwater storage anomalies (GWSAs) with groundwater level anomalies (GWLAs) from ~23,000 wells and with groundwater storage (GWS) from regional and global models in 14 major U.S. aquifers. Results show large GWS declines from GRACE in southwestern (Central Valley and Arizona Alluvial Basins) and south-central (Central and Southern High Plains and Texas) aquifers from multiyear droughts (5–15 yr). In contrast, GWS trends in aquifers throughout the rest of the U.S. showed mostly stable or rising values. Time series of GRACE GWSAs compared favorably with GWLAs from most aquifers, suggesting that GRACE data track groundwater (GW) dynamics. Regional GW models show similar or greater declines in GWS compared with GRACE data, with the largest discrepancy of a factor of four times greater modeled depletion in the

Mississippi Embayment. Global hydrologic models show minimal storage dynamics but greatly overestimated GWS declines by ~2.4× in southwestern and south-central aquifers with intensive irrigation compared with GRACE data. In contrast, global land surface models show similar GW dynamics to GRACE data but underestimated GWS declines in heavily exploited aquifers because these land surface models do not include human intervention.

1. Introduction

There is increasing interest in groundwater storage (GWS) changes because of concerns about overexploitation and depletion in many aquifers globally (e.g., U.S. Central Valley, U.S. High Plains, northwest India, North China Plain) (Arciniega-Esparza et al., 2017; Brena-Naranjo et al., 2014; Long et al., 2016; Longuevergne et al., 2010; Scanlon et al., 2012; Wei et al., 2017). Traditional hydrologic approaches to assessing GWS changes have relied on groundwater level (GWL) monitoring and regional groundwater (GW) modeling (Konikow, 2013; McGuire, 2017). In the United States, large numbers of wells are monitored annually (e.g., ~7,500 wells in the High Plains aquifer; McGuire, 2017), and ~1,700 wells are monitored continuously in the United States (https://waterdata.usgs.gov/). Scaling up point GWL data to develop composite estimates of GWL variations over an aquifer is complicated because some wells may reflect impacts of local GW pumping and declustering data may be difficult. Converting GWLs or heads to GWS (units of length or volume/unit area) requires information on aquifer storage coefficients (S_C, unitless):

$$\Delta GWS = S_c \Delta GWLs, \tag{1}$$

where Δ is change relative to the long-term mean. However, storage coefficients can vary over several orders of magnitude from unconfined aquifers (e.g., specific yield, ~0.02–0.3) to confined aquifers (e.g., storativity, ~0.0001–0.001) (Freeze & Cherry, 1979). However, there is a continuum between unconfined to confined conditions with some aquifers predominantly semiconfined, such as the California Central Valley. In systems with vertically stacked aquifers, it is often difficult to determine whether wells are screened in unconfined or confined aquifers, or both, increasing uncertainty in GWS estimates from GWL data.

Regional GW models have been developed for many of the major U.S. aquifers by the U.S. Geological Survey (USGS) and other agencies (Deeds & Jigmond, 2015; Faunt, 2009; Peterson et al., 2016, 2020). These models are data intensive and use reported or estimated pumpage (Dieter et al., 2018). Some of these models rely primarily on GWL data for calibration, which can result in nonunique solutions for fluxes and storage (Hill & Tiedeman, 2007). Integration of additional information, such as base flow to streams and chemical tracers, better constrains model parameters and predictions. GWS information could similarly be leveraged to improve model calibration and GW predictability, but little emphasis has been placed on GWS in the past because of lack of independent GWS data at regional scales until the advent of the Gravity Recovery and Climate Experiment (GRACE) satellites in 2002.

Limited availability of ground-based monitoring and regional modeling in many areas, particularly in developing countries, has resulted in increasing interest in using global models and remote sensing to assess changes in GWS. Two basic types of global models exist: (1) global hydrologic models (GHMs) and (2) global land surface models (LSMs). GHMs were originally developed to evaluate water scarcity; they are based on water balance and most do not include energy balance. These models generally include all component storages (surface water, SW; soil moisture, SM; and GW) and human intervention (HI; water abstraction, return flow from SW or GW abstraction, and reservoir storage). GHMs have been used to identify hot spots of GW depletion in different regions (North China Plain, NW India, Tigris-Euphrates Basin, etc.) (Döll et al., 2016; Wada et al., 2010). In contrast, LSMs were developed to model the lower boundary condition for climate models and include water and energy conservation equations. Most LSMs do not include HI and generally limit storage to snow and SM.

GRACE satellites have been widely used to monitor changes in total water storage (TWS) globally since the original satellites were launched in 2002 (Tapley et al., 2004). Monitoring the distance between the two GRACE satellites provides data on spatiotemporal variability in the Earth's gravity field that is controlled

primarily by variations in water storage caused by many processes, such as floods, droughts, ice melting, and GW pumping (Rodell et al., 2018). Global scale analyses have been used to assess impacts of terrestrial TWS changes on sea level rise (Reager et al., 2016). GRACE TWS changes are represented as anomalies (TWSAs) relative to the long-term mean. TWSAs are vertically integrated from the land surface to the deep subsurface at scales ≥ ~100,000 km^2. TWSAs include the following component storages:

$$\mathrm{TWSA} = \mathrm{SnWSA} + \mathrm{SWSA} + \mathrm{SMSA} + \mathrm{GWSA}, \quad (2)$$

where the ending SA refers to storage anomalies and SnW is snow water. Therefore, to estimate GWSAs, all other component storages must be subtracted from TWSAs. SnWS and SMS are generally estimated from models (e.g., snow from Snow Data Assimilation System [SNODAS], SM from LSMs), and SW is estimated from reservoir-monitoring data. In some regions, there is considerable uncertainty in the relative contribution of different storage components to TWSAs. For example, in various studies of the Tigris-Euphrates Basin, GRACE TWS changes have been attributed to different components, including GW irrigation pumpage (Voss et al., 2013), surface reservoirs (Longuevergne et al., 2013), and natural GWS and reservoir storage (Mulder et al., 2014). Similarly, different studies in the U.S. Colorado River Basin have attributed GRACE TWS changes to GW pumping (Castle et al., 2016) or to SW storage, SM storage, and mostly natural GWS changes in response to droughts (Scanlon et al., 2015).

None of the previous studies integrated GWS estimates from combined data on GWL monitoring, regional and global modeling, and GRACE data. Such integration should help assess the reliability of different approaches for estimating GWS trends and test these various methods. The objective of this study was to attempt to fill this gap in previous studies by addressing the following questions:

1. How much do estimates of GWSAs contribute to TWSAs from GRACE?
2. How well do GRACE GWSAs compare with GWSAs from GWL monitoring data?
3. How do GRACE GWSAs compare with GWSAs from regional and global models?

The major components of the study are shown in a flow chart (Figure 1). The major U.S. aquifers provide an excellent field laboratory to address these objectives because of the intensive monitoring over thousands of wells that cover the GRACE period and regional modeling of many of the aquifers (Konikow, 2013). This study builds on previous studies that compared GRACE GWSAs with GWL data in individual aquifers, such as the Central Valley (Scanlon et al., 2012) and the High Plains (Brena-Naranjo et al., 2014; Brookfield et al., 2018; Longuevergne et al., 2010; Seyoum & Milewski, 2016; Strassberg et al., 2009) by expanding the analysis to compare with GWL data in most of the major U.S. aquifers. More detailed comparisons of GRACE GWSAs with GWSAs from regional models were also conducted, leveraging the results of previously published individual aquifer studies (Argus et al., 2017; Xiao et al., 2017). We also expanded beyond evaluation of long-term trends to include an analysis of interannual variability in GWS. Particular attention was given to estimating uncertainties in GRACE GWSAs from uncertainties in component storages. The primary foci of the work are (i) to advance interpretation of GRACE data by combining it with GWL monitoring data and (ii) to evaluate the use of GRACE data as an independent constraint on GWS estimates from regional and global modeling.

2. Materials and Methods

2.1. Major Aquifers

We selected 14 major aquifers within the United States for the analysis (Miller, 1999). These aquifers are generally intensively monitored, and regional models have been developed and are available for eight of the 14 aquifers (Figures 2 and 3). Some of these aquifers are also notable because they are hot spots of GW depletion (e.g., Central Valley, High Plains, Mississippi Embayment) (Konikow, 2013). These aquifers represent a variety of aquifer types, including varying aquifer material: unconsolidated alluvial deposits (Central Valley, High Plains, Mississippi River Valley Alluvium, Arizona Alluvial Basins), semiconsolidated sedimentary aquifers (Coastal Lowlands, Texas Gulf Coast), sandstone and carbonate aquifers (Colorado Plateau, Edwards Trinity Plateau, Pennsylvanian, and Floridan aquifers), and aquifers of basaltic and volcanic rock with interbedded alluvium (Columbia Plateau and Eastern Snake Plain) (Miller, 1999). Some aquifers are used for intensive irrigation (Central Valley, High Plains, and Mississippi Embayment).

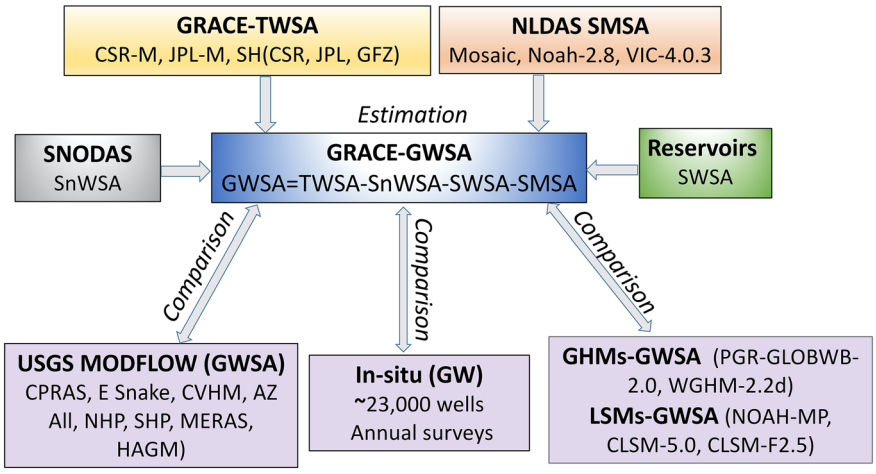

Figure 1. Flowchart describing the analysis conducted in this study. GRACE GWSA refers to GRACE-derived GWS anomaly as shown in the equation where TWSA is GRACE Total Water Storage Anomaly (GRACE solutions from University of Texas Center for Space Research, CSR; NASA Jet Propulsion Lab., JPL; M = Mascons; SH = spherical harmonics; GFZ = GeoForschungsZentrum); SnWSA is snow water storage anomaly from Snow Data Assimilation System; SWSA is surface water storage anomaly (mainly reservoirs); and SMSA is soil moisture storage anomaly from National Land Data Assimilation System Land Surface Models (NLDAS models: Mosaic, Noah-2.8, and VIC-4.03). GRACE GWSAs were calculated from GRACE TWSAs based on mean of CSR and JPL mascons. The long-term variability in GRACE-GWSAs (trend and interannual variability) were compared with GWSAs from regional models (CPRAS = Columbia Plateau Regional Aquifer System; E Snake = Eastern Snake Plain Aquifer Model; CVHM = Central Valley Hydrologic Model; AZ All = Arizona Alluvial Basins; NHP = Northern High Plains; SHP = Southern High Plains; MERAS = Mississippi Embayment Regional Aquifer System; and HAGM = Houston Area Groundwater Model. GRACE GWSAs were also compared with composite GWL hydrographs from each aquifer, totaling 23,000 wells using synoptic data from winter periods. GRACE GWSAs were compared with global hydrologic models (GHMs: PCR-GLOBWB-2.0; WGHM-2.2d) and land surface models (LSMs: NOAH-MP, CLM-5.0, CLSM-F2.5) that included a GWS component.

The 14 regional aquifers include a range of unconfined to confined conditions that are important to consider in evaluating GWL records and comparing with GRACE GWSAs. Predominantly unconfined aquifers include the Eastern Snake Plain, Edwards Trinity Plateau, and the High Plains (George et al., 2011; McGuire, 2017; Twining & Fisher, 2012). The remaining aquifers generally include both unconfined and confined portions with unconfined aquifers laterally adjacent to or underlain by confined aquifers separated by low-permeability clay layers. Aquifers in the Central Valley and Arizona Alluvial Basins are generally considered semiconfined with degree of confinement linked to clay distribution (Faunt, 2009; Pool & Anderson, 2008). Most of the pumpage (~90%) in the Mississippi Embayment Regional Aquifer System (MERAS) is in the shallow unconfined alluvial aquifer system (Mississippi River Valley Alluvium) with limited pumpage in the deeper confined Sparta aquifer (Clark & Freiwald, 2011; Clark et al., 2011). Shallow parts of the basalt and volcanic aquifers of the Columbia Plateau Regional Aquifer System (CPRAS) include alluvial unconfined regions of limited extent with the remainder of the system confined (Ely et al., 2014). The Upper Colorado River Basin includes shallow alluvial aquifers adjacent to rivers and deeper confined aquifers, with very little GW use (Scanlon et al., 2015). The Floridan aquifer system includes the unconfined Upper Floridan aquifer underlain by the confined Lower Floridan aquifer; however, these aquifers are highly connected and behave as a single system (Bellino et al., 2018). The Gulf Coast aquifers (Texas Gulf Coast and Coastal Lowlands) include unconfined and confined aquifers (Kasmarek, 2012).

2.2. Data Sources and Analysis

2.2.1. GRACE Data (TWS Anomalies)

GRACE data are based on GRACE Release 06 solutions (section S2 in the supporting information). We focused primarily on GRACE mass concentration (mascon) solutions from The University of Texas at Austin Center for Space Research (CSR, RL06, Version 1; Table S3) and from the NASA Jet Propulsion Laboratory (JPL, RL06, Version 2; Table S4). Goddard Space Flight Center (GSFC) mascons were omitted because data after July 2016 were not available. The mascon solutions were selected because they reduce

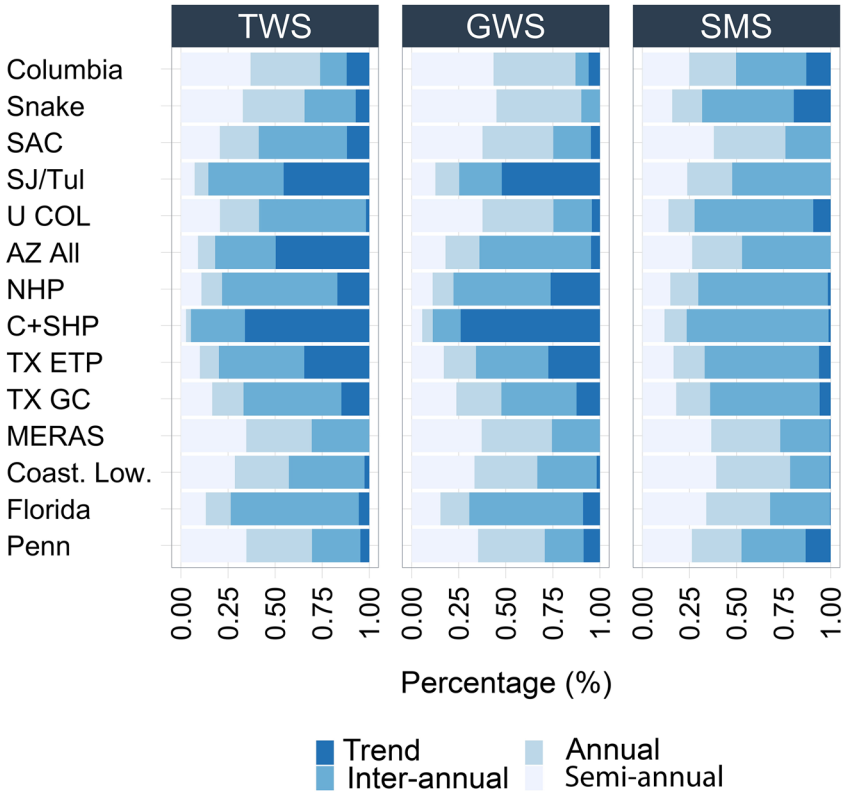

Figure 2. Temporal dynamics of water storage based on variances in temporal components of total water storage (TWS), groundwater storage (GWS), and soil moisture storage (SMS) (Table S2). The temporal components are based on STL analysis (section S6.1). The variance of each temporal component (e.g., trend and interannual) was divided by the variance in the raw time series to estimate the relative contribution of each temporal component to the total (section S6.3). Temporal components include long-term trends, inter-annual variability, and annual and semi-annual variability. GRACE TWS is the average of CSR. M and JPL.M.dsf solutions. SMS is the ensemble of the NLDAS models (MOSAIC, NOAH-2.8, and VIC-4.0.3). GWS is the residual of the water budget equation. The *y* axis refers to the 14 major aquifers: Columbia = Columbia Plateau; Snake = Snake River Basin; SAC = Sacramento; SJ/Tul = San Joaquin and Tulare; U COL = Upper Colorado; AZ All = Arizona Alluvial Basins; NHP = Northern High Plains; C + SHP = Central and Southern High Plains; TX ETP = Texas Edwards Trinity Plateau; TX GC = Texas Gulf Coast; MERAS = Mississippi Embayment Regional Aquifer System; Coast. Low = coastal lowlands; Florida = Floridan; and Penn = Pennsylvanian aquifers.

leakage issues on land and between land and ocean. GRACE data are provided at about monthly intervals as an equivalent water thickness (in centimeters). Data from April 2002 through June 2017 (~15.25 yr) were used in this analysis. Missing months were imputed using linear interpolation. The gridded GRACE data used for this study range from 0.25° for CSR (derived from the 1° native resolution) to 0.5° for JPL (derived from the 3° native resolution). The Coastline Resolution Improvement was applied to the JPL data (Wiese et al., 2016). Uncertainties in GRACE TWSA data were based on the standard deviation of five GRACE solutions (CSR and JPL mascons and CSR, JPL, and GeoForshungsZentrum [GFZ] spherical harmonic [SH] solutions) (section S2; Table S9). Seasonal Trend decomposition using Loess (STL) methodology was used to disaggregate TWSA time series and storage components into long-term variability (trends + interannual variability) as well as annual and semiannual variability (Cleveland et al., 1990) (Equation 3; Figure S1; section S6.1). Analysis in this study focused on long-term variability, here defined as time series components with frequency ≥13 months (trends + interannual variability).

Monthly GRACE TWSA observations of each aquifer system were extracted from the global data sets by masking the aquifer boundaries in the GRACE global grid and area-weighting the GRACE data within the relevant grid cells. GRACE monitors changes in water storage, which are shown as anomalies relative to the long-term mean (2002–2017) in this analysis. TWSA within the aquifers was calculated by

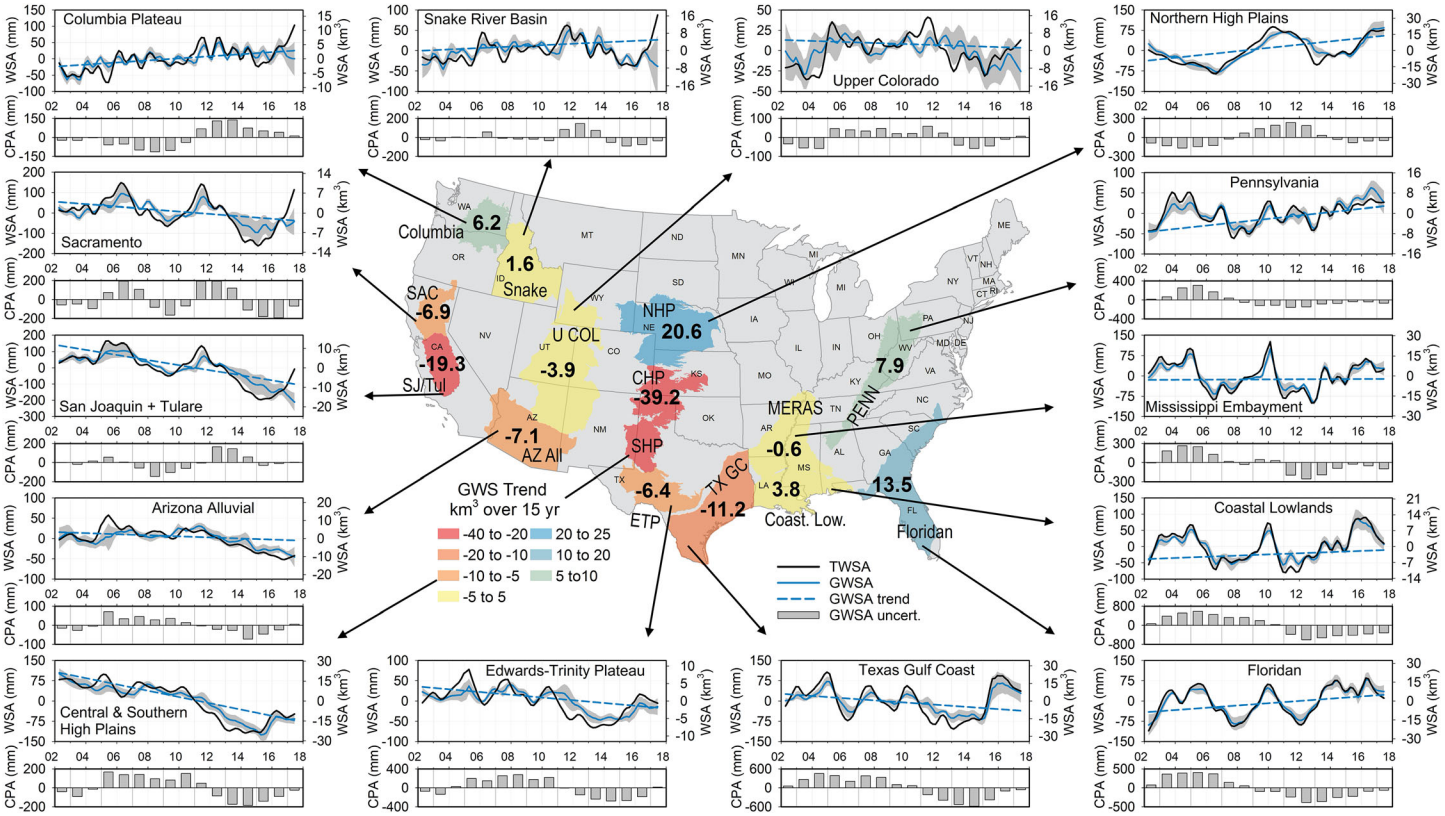

Figure 3. GWS trends from GRACE in km^3 over the 15 yr monitoring period (2002–2017) shown for the 14 major aquifers in the U.S. map (Tables 1 and S1). Time series plots: (upper panels) long-term variability (interannual variability and linear trend) in monthly water storage anomalies (WSA) including TWSA (mean CSR-M and JPL-M, black line) and GWSA from GRACE (blue line) (Tables S3–S8). The trend in GWSA is represented by blue dashed line. Uncertainty in GWSA is shown as the gray bands based on the propagated errors from TWS, snow water storage, surface water storage, and soil moisture storage (Tables S10–S20). (Lower panels) Annual cumulative precipitation anomaly (CPA) as gray bars (Table S21).

integrating the water volumes of contributing cells. GRACE polygons corresponded to the aquifer extent for the large regional aquifers, such as the High Plains. However, smaller regional aquifers were evaluated using GRACE by extending the system boundaries to the surrounding larger river basins (e.g., Sacramento and San Joaquin/Tulare Basins [155,000 km^2], which encompass the Central Valley aquifer [55,000 km^2]). A similar process was applied to the Eastern Snake Plain aquifer (40,700 km^2) and surrounding Snake River Basin adjudication (186,500 km^2) and the region comprising the ~70 alluvial basins of the Arizona Alluvial Basins aquifer system (225,400 km^2) and surrounding nonaquifer regions consisting primarily of crystalline and volcanic rocks (424,350 km^2).

2.2.2. SM, Snow, SW, and GWS

GWSAs were estimated from GRACE TWSAs by subtracting SnWSAs, SWSAs, and SMSAs (Equation 2) (section S2; Table S19). Data on SnWSAs were obtained from SNODAS (estimated uncertainty ±15%; Tables S15 and S16) (Barrett, 2003; Marton et al., 2015). SWSAs were obtained from ground-based monitoring of surface reservoirs with storage capacities exceeding ~0.5 km^3 from a variety of sources (estimated uncertainty ±15%) (Tables S17 and S18). SMSAs were derived from LSMs included within the North American Land Data Assimilation System (NLDAS Version 2: MOSAIC, NOAH-2.8, and VIC-4.0.3; Tables S10–S13) (Xia et al., 2013, 2014). Variability in SMS from other LSMs (CLSM, NOAH-MP) is similar to those from NLDAS, but NLDAS was selected because of higher spatial resolution and longer simulated period (≤2017) for North America relative to other LSMs (≤2014). Uncertainty in SMS was estimated from the standard deviation of the three NLDAS models (Table S14). Most of these models limit SMS to the upper 2 m with three to four soil layers. Uncertainty in GWSAs was calculated as the square root of the sum of the squared errors of each storage component (Equation S2.0). Deeper storage below the shallow modeled soil

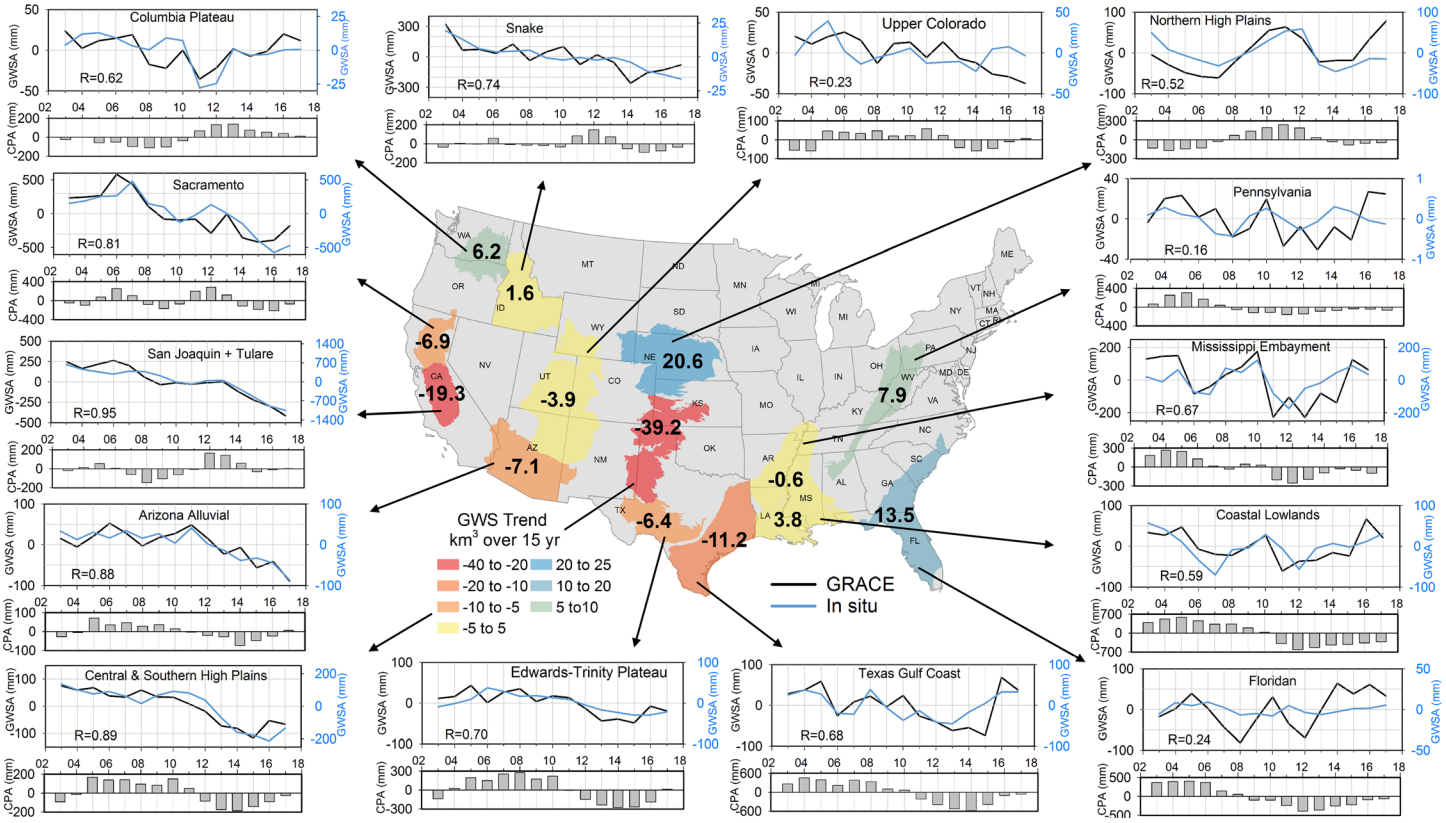

Figure 4. The time series show annual variability in GRACE GWSAs (black line) and annual variability in GWSAs from composite GW level anomalies (GWLAs) from monitoring data (blue line) based on spring and winter seasons (Table S24). In situ GWSAs were estimated using the storage coefficients from regional models (Table S26.1). In situ GWSAs were calculated from 10-km gridding resolution in the 14 major aquifers. The results for the GRACE data for the Snake, Sacramento, San Joaquin, Arizona Alluvial, and Mississippi Embayment were scaled based on the ratio of the basin area to the aquifer area under the assumption that storage change is restricted to the aquifers. The different percentages of areas are shown in Table S23.4. The gray bar graphs show annual cumulative precipitation anomalies (CPA) (Table S21). The U.S. base map is similar to that in Figure 3 showing trends in GRACE GWS in km^3 over the 15 yr monitoring period (2002–2017) (Tables 1 and S1).

zone and above the GW should not contribute substantially to GWS uncertainty because long-term storage variations in this zone are thought to be low, especially in semiarid regions (Scanlon et al., 2005, 2007). The contribution of each storage component (long-term variability) to the sum of the storage components was calculated from the mean absolute deviation (MAD) of each storage component as a percent of the sum of the MADs of all storage components (proxy for TWS; Table S2.2).

GWSAs were calculated as the residual of the water budget equation 2 using GRACE TWSAs (mean of CSR and JPL mascon solutions), SnWS from SNODAS data, SWS from in situ reservoirs, and SMS from NLDAS models (MOSAIC, NOAH-2.8, and VIC-4.0.3) (Table S19). GWSAs accumulate uncertainties in each of the water budget terms (Table S20).

2.2.3. Groundwater Level Monitoring Data

GRACE-derived GWSAs were compared with GWSAs from ground-based GWL monitoring data. GWL data were converted to GWS using the reported storage coefficient based on regional modeling (Figure 4; Table S26.1). GWLAs were also converted to GWSAs using an effective storage coefficient from the relationship between GRACE GWSAs and GWL anomalies (Figure S4; Tables S24 and S26.1).

Storage coefficients are generally high and fairly uniform in unconfined aquifers (Edwards Trinity Plateau, Eastern Snake Plain, and High Plains, aquifers). However, many of the studied aquifers include unconfined and confined systems with order of magnitude variability in storage coefficients. In most of these systems we used different approaches to isolate the shallow, predominantly unconfined portions to compare with GRACE data, as detailed later in this section. The calculated effective storage coefficients for aquifers

(Table S26.1) were also compared with storage coefficients from regional GW models. Some previous studies that compared GRACE GWSAs to ground-based data used regional estimates of storage coefficients for the aquifers (section S3).

The GWL data used in this analysis include water levels from ~22,600 wells and were obtained from a variety of sources, such as the USGS National Water Information System (NWIS) and the California Statewide Groundwater Elevation Monitoring (CASGEM) (section S3; Table S23.2). Comparison with GRACE data focused on GWLs monitored during the winter period (December through May) to lessen the impact of variations in the magnitude and timing of irrigation pumping in many aquifers. GWSAs from GRACE long-term trend analysis were also averaged for winter months for comparison with the GWL data (Table S24).

GWL data were filtered to exclude outliers (i.e., groundwater level anomalies [GWLAs] > 1.5 × interquartile range) assumed to be impacted by GW pumping. GWL data were gridded at 10, 25, 50, and 100 km resolutions to assess the sensitivity of results to gridding resolution (Table S25.1). Results from the different grid resolutions were similar, but the 10 km resolution likely best represents spatial variations in GWLs and was selected for comparison with GRACE data.

It is important to carefully analyze GWL data because small GWL variations in unconfined aquifers would result in much larger GWS changes than similar GWL variations in confined aquifers because of orders of magnitude lower storage coefficients in confined aquifers (Equation 1). GWL variations in unconfined aquifers can be analyzed as a single system and should behave similarly (Edwards Trinity Plateau, High Plains, and Eastern Snake Plain). We limited analysis in the Mississippi Embayment to the shallow unconfined Mississippi River Valley Alluvium wells and excluded GWLs from the deeper confined Sparta aquifer. Data on the status of wells in terms of unconfined and confined systems are available for a limited number of aquifers (Columbia Plateau, Coastal Lowlands, Mississippi Embayment, and Floridan) (section S3; Table S25.2). Previous studies indicate that unconfined and confined sections of the Floridan aquifer system are highly connected and behave as a single system (Bellino et al., 2018). For wells with specific unconfined and confined sections (Columbia Plateau, Texas Gulf Coast, and Coastal Lowlands), we applied an unsupervised machine learning algorithm (*k*-mean clustering) to partition wells in these aquifers into *k* clusters according to their location, depth, trend, long-term mean, and standard deviation of GWLs (Lloyd, 1982) (Table S23.1). Results from the cluster analysis were used to identify unconfined portions of the aquifers for comparison with GRACE GWSAs. There were insufficient data in the Upper Colorado to apply the *k* clustering approach. The Central Valley and Arizona Alluvial Basin aquifers behave as semiconfined systems with the degree of confinement varying with clay content (Faunt, 2009; Faunt & Sneed, 2015), and all well data were used for comparison with GRACE data.

2.2.4. Regional Groundwater Models

Regional GW models were available for eight of the 14 aquifers and were compared with the GRACE data over the common time period (Tables S27–S28). The regional models are the CPRAS, Eastern Snake Plain Aquifer Model (ESPAM), California Central Valley Hydrologic Model (CVHM, Arizona Alluvial Basins (six models), Northern High Plains (NHP), Southern High Plains, the Houston Area Groundwater Model (HAGM) including the northern part of the Texas Gulf Coast aquifers (Kasmarek, 2012), and the MERAS. Many of these models are described in Konikow (2013) with additional details in the supporting information (section S4; Table S27). The time period for most regional models did not extend over the entire GRACE period, which limited comparisons for most aquifers. All regional models were developed using the USGS finite difference GW code MODFLOW (Harbaugh et al., 2000). Simulated water-budget information for each stress period was used to calculate the total GWS change for each regional aquifer through the simulated period. For one regional aquifer, the Eastern Snake Plain, storage change was not derived directly from the model, as it includes only a brief period coincident with GRACE. Instead, storage change was estimated based on the product of specific yield values derived from the calibrated MODFLOW model and annual synoptic surveys of winter GWL observations throughout the GRACE period. Most models consist of an early predevelopment (i.e., prior to intensive pumping) steady-state model followed by a multidecadal transient model. Models are primarily based on 1 square mile (2.6 km^2) or higher resolution grids with annual or shorter stress periods during the GRACE data period. Model input data generally include initial and transient conditions of recharge, pumpage, evapotranspiration, stream/spring base flow discharge, and state variables

(hydraulic conductivity and storage coefficient). Model output for each stress period generally includes GWLs or heads, and GW budgets, including GWS, well withdrawals, and base flow to streams and springs. Only annual simulated GWS results were compared with GRACE GWSAs. These regional GW models were calibrated using GWL data and, in most cases, GW discharge to streams.

2.2.5. Global Models

Global models evaluated in this study include GHMs: PCRaster-Global Water Balance (PCR-GLOBWB) (Sutanudjaja et al., 2018) and WaterGAP Global Hydrology Model (WGHM) (Mueller-Schmied et al., 2016). LSMs were evaluated that simulate GWS (LSMs: CLSM, NOAH-MP) (section S5) (Ek et al., 2017; Lawrence et al., 2019). CLM-5.0 was also included although it does not simulate GW as a separate compartment. The top 2 m of the CLM-5.0 profile was assumed to equal SM to make direct comparisons with the other LSMs. The LSMs, except CLM-5.0, include SnWS, SMS, and GWS whereas the GHMs include all storage components (Equation 2). HI is considered in GHMs, relying on simulated or calculated GW pumpage, but LSMs do not simulate HI. None of these models were calibrated except WGHM-2.2d, which was calibrated to mean annual streamflow. Additional details are given in (section S5; Table S29).

2.2.6. Time Series Decomposition

The time series of storage components can be decomposed using the STL algorithm developed by Cleveland et al. (1990). The temporal components are as follows:

$$S_{raw} = S_{Long-term} + S_{Annual} + S_{Semi-annual+resdiual}, \tag{3}$$

where S_{raw} is the original storage data, long-term (trends + interannual) and annual, semiannual, and residual components. More details are provided in section S6.1.

2.2.7. Component Storage Contributions to TWS

The component storage contributions relative to sum of the storage components in each aquifer was calculated as the percentage of the MAD of long term storage (trend + interannual variability) relative to the MAD of all other components (Kim et al., 2009) as

$$SI_i = \frac{MAD\,(S_i)}{\sum_i^n MAD}, \; MAD = \frac{1}{N}\sum_t^N \left|S - \overline{S}\right|, \tag{4}$$

where SI_i is the storage intensity for storage component (S) (SnWS, SWS, SMS, and GWS), and n is the total number of storage components, N is the number of months, and $\overline{S}$ is the long-term mean.

3. Results

3.1. Variations in Groundwater Storage Anomalies From GRACE in Major Aquifers

GWS variation is the primary contributor to long-term (trend + interannual) variability in GRACE TWSAs in most aquifers, accounting for more than 50% in nine of the 14 aquifers, except those in the northwest (Figure S3; Table S2). The dominance of GWS variability could be partly related to underestimation of SMS variability in regions with thick soils because NLDAS SMS, used to estimate GWS, is restricted to the upper 2 m. There is little variability in SMSAs among the NLDAS models (Figure S5). The contribution of snow water storage (SnWS) to TWS was only important in the western United States, accounting for ~10–14% in the California Central Valley and Upper Colorado River Basin and ~30% in the northwest. SWS contributed 13–29% of the variability in the Southwest, in the Central Valley (including reservoirs in the Sacramento and San Joaquin/Tulare Basins), Upper Colorado (mostly Lake Powell), and Arizona Alluvial Basins (Lake Mead) but much less elsewhere. The impact of Lake Mead on water storage may be overestimated because it is located near the basin boundary, where storage signals are amplified (Longuevergne et al., 2013). SMS generally accounted for ~16–45% of TWS variability, with no systematic variation among aquifers.

Temporal variability in GWS is dominated by long-term trends in the San Joaquin/Tulare Basins (53% of total GWSA signal) and Central and Southern High Plains (65%), which dominate long-term trends in TWS in these basins (Figure 2; Table S2). Interannual variability in GWSs was highest in the Arizona Alluvial Basins, NHP, Texas Gulf Coast, and Floridan aquifers (each ~50%). Annual and semiannual GWS

contributions were lower than contributions from trends and interannual variability in a number of aquifers (e.g., Central Valley and High Plains).

Temporal variations in GRACE GWSAs are similar to those of GRACE TWSAs, with correlation coefficients (R) of 0.57 to 0.98 in most aquifers (Table S22). However, GWSA results are conditioned by accumulating uncertainties in other modeled water-budget components, such as snow, reservoirs, and soils, which are important in the Columbia and Upper Colorado systems and in humid regions (Pennsylvanian and Floridan aquifers).

Trends in GWSAs derived from GRACE were negative in the southwestern and south-central United States, totaling approximately -90 km^3 over the past 15 yr (April 2002 to June 2017; trends based on Sen slopes; Sen, 1968), ~2.5× the capacity of Lake Mead (~35 km^3), the largest surface reservoir in the United States (Figure 3; Tables 1 and S1). This depletion reflects GWS losses totaling -26 km^3 in the Central Valley, -7 km^3 in the Arizona Alluvial Basins, -39 km^3 in the Central and Southern High Plains, -6.4 km^3 in the Texas Edwards Trinity Plateau, and -11 km^3 in the Texas Gulf Coast. There was almost no trend in GRACE GWSAs in the Mississippi Embayment (-0.6 km^3/15 yr). In contrast, trends in GRACE GWSAs were slightly positive over the past 15 yr in aquifers in the northwest (Columbia Plateau and Snake River Plain: 2–6 km^3), north (~21 km^3, NHP), and east (8–13 km^3, Pennsylvanian and Floridan aquifers). Decreasing GWSAs during the 15 yr GRACE period generally reflect responses to long-term droughts as shown in the cumulative precipitation anomalies (Figure 3; Table S21) and related irrigation pumpage in California, Arizona, and Texas (Dieter et al., 2018). Both California and Texas have been subjected to 4–5 yr droughts in recent years with only partial recovery during wet periods, resulting in long-term net declines in GWSAs. GWSAs in the Arizona Alluvial Basins aquifers were maintained by two wet periods (2005 and 2010); however, there was no significant water input after 2010 resulting in GWS loss. Uncertainties in GWSAs were generally low in most aquifers and fairly uniform over time.

Although many aquifers are generally represented as a single system, for example, Central Valley and High Plains (Konikow, 2013), there is substantial variability in GWSA trends within these aquifers. For example, trends in GWSAs, when normalized by aquifer area and expressed as equivalent water height, were approximately two times greater in the southern (San Joaquin/Tulare Basins; -15 mm/yr) than in the northern (Sacramento, SAC; -6.3 mm/yr) Central Valley; however, GWSAs in the two regions were highly correlated ($R = 0.85$) (Table S22). Many studies emphasize GW depletion throughout the High Plains aquifer (Konikow, 2013); however, GWS increased in the NHP (5.4 ± 0.1 mm/yr) but decreased markedly in the combined Central and Southern High Plains (-12.6 ± 0.8 mm/yr), and GWSAs were poorly correlated between the Northern and Central + Southern High Plains ($R = -0.30$) (Table S22).

3.2. Comparison Between GRACE-Derived and In Situ Groundwater Storage Anomalies

Generally high correlations between GRACE derived GWSAs and in situ GWSAs for many aquifers suggest that GRACE satellites track the dynamics of GWS changes in these aquifers (Figure 4; Table S26.1). The reported storage coefficients or the effective storage coefficients (based on the relationship between GWLAs and GRACE GWSAs) simply scale the GWL anomalies; therefore, the correlation coefficients between GRACE GWSAs and in situ GWL anomalies or in situ GWSAs are the same. Poor correlations in some aquifers may reflect uncertainties in GRACE GWSAs or uncertainties in generating the composite hydrographs of GWLAs or both. The poorest correlations are found in regional aquifers dominated by low amplitude GWSAs and low signal-to-noise ratio, low storage coefficients (<0.05), or confined GW conditions (e.g., Columbia Plateau, Upper Colorado, Floridan, and Pennsylvanian aquifers).

Annual in situ GWSAs in unconfined and semiconfined aquifers (Edwards Trinity Plateau, High Plains, Eastern Snake Plain aquifers, Central Valley and Arizona alluvial) were highly correlated with GRACE GWSAs ($R = 0.52$–0.95) (Figure 4; Table S26.1). GWLAs from the unconfined portion of the Mississippi Embayment and GRACE GWSAs were also highly correlated ($R = 0.67$). Aquifers with data for unconfined and confined portions of the aquifers show good correspondence in the time series (Figure S8). The k cluster analysis did not help distinguish different well groups with different depths in the Central Valley; therefore, in the Central Valley, all GWL data were combined to compare with GRACE GWSAs ($R = 0.81$–0.95).

Cluster analysis identified a shallow group of wells (~900 wells) in the Columbia Plateau, likely unconfined, that resulted in a much higher correlation with GRACE GWSAs ($R = 0.62$) relative to using the entire well

Water Resources Research

10.1029/2020WR027556

Table 1
Trends of the GRACE-Derived GWSAs Over 15 yr (2002–2017) and 12 yr (2003–2014) to Correspond to the Model Periods

		GRACE-GWSA			GHMs (km^3/12 yr) (2003–2014)				LSMs (km^3/12 yr) (2003–2014)			
ID	Name	mm/yr	km^3/15 yr (2003–2017)	km^3/12 yr (2003–2014)	PCR-GLOBWB-HI	PCR-GLOBWB-NHI	WGHM-HI	WGHM-NHI	CLM-5.0	CLSM-F.2.5	NOAH-MP	Ensemble-LSMs
1	Columbia Plateau	3.6 ± 1.8	6.2 ± 3.1	6.2 ± 0.1	−1.5 ± 0.1	−0.0 ± 0.1	−0.3 ± 0.1	−0.1 ± 0.2	1.2 ± 0.4	2.8 ± 1.0	0.2 ± 0.3	1.6 ± 0.4
2	Snake	0.6 ± 0.02	1.6 ± 0.1	6.6 ± 0.8	−5.9 ± 0.8	−0.2 ± 0.7	−4.7 ± 0.5	0.2 ± 0.4	0.9 ± 0.4	3.3 ± 2.2	2.1 ± 0.6	1.9 ± 0.7
3	Sacramento	−6.3 ± 2.0	−6.9 ± 2.2	−4.4 ± 0.5	−7.0 ± 0.5	−4.2 ± 0.4	−0.9 ± 0.3	−1.8 ± 0.5	−1.9 ± 0.2	−4.2 ± 1.7	−0.9 ± 0.4	−2.4 ± 0.6
4	San Joaquin + Tulare	−15.2 ± 2.1	−19 ± 2.7	−11.3 ± 2.0	−65 ± 2.0	−3.4 ± 0.6	−29 ± 0.5	−0.7 ± 0.2	−5.0 ± 0.3	−2.2 ± 1.0	−0.6 ± 0.4	−2.6 ± 0.6
5	U. Colorado	−0.7 ± 0.01	−3.9 ± 0.1	1.3 ± 0.4	−1.7 ± 0.4	−1.5 ± 0.3	−1.5 ± 0.3	0.2 ± 0.4	−1.2 ± 1.0	−1.9 ± 3.2	−1.2 ± 1.4	−2.3 ± 1.9
6	Arizona Alluvial	−2.1 ± 0.6	−7.1 ± 2.1	0.5 ± 0.5	−15 ± 0.4	−0.9 ± 0.1	−9.4 ± 0.2	−0.2 ± 0.1	−4.6 ± 0.8	−3.0 ± 1.3	−3.9 ± 1.4	−5.1 ± 6.4
7	N-High Plains	5.4 ± 0.1	21 ± 0.4	17.8 ± 4.2	−43 ± 4.2	2.0 ± 1.1	−39 ± 5.0	2.8 ± 1.3	18 ± 1.4	19 ± 9.6	3.1 ± 1.9	9.1 ± 5.9
8	C + S-High Plains	−12.6 ± 0.8	−39 ± 2.5	−31 ± 2.2	−90 ± 2.1	−0.9 ± 0.1	−127 ± 4.7	−1.5 ± 0.4	−10 ± 1.4	−29 ± 4.4	−3.1 ± 0.8	−15 ± 2.8
9	Edwards Trinity Plateau	−3.6 ± 0.6	−6.4 ± 1.1	−6.1 ± 0.1	−2.1 ± 0.1	−0.9 ± 0.1	−1.3 ± 0.2	−1.1 ± 0.4	0.6 ± 0.5	−6.4 ± 0.9	−1.8 ± 0.5	−2.7 ± 0.4
10	TX-Gulf coast	−3.2 ± 0.5	−11 ± 1.8	−21.8 ± 0.5	−6.1 ± 0.5	−2.9 ± 0.5	−4.1 ± 0.7	−5.1 ± 1.6	−9.4 ± 1.5	−16 ± 1.0	−6.3 ± 1.0	−10 ± 1.7
11	Mississippi Embayment	−0.2 ± 0.2	−0.6 ± 0.6	−11.5 ± 0.6	−4.7 ± 0.5	−0.2 ± 0.5	−74 ± 0.8	0.4 ± 0.4	3.2 ± 3.6	−9.7 ± 4.4	−1.0 ± 1.0	−3.6 ± 4.4
12	Coastal Lowlands	1.8 ± 0.6	3.8 ± 1.2	−3.8 ± 0.5	−3.7 ± 0.5	−1.2 ± 0.5	−0.1 ± 0.7	−0.4 ± 0.1	−14 ± 3.8	−11 ± 3.0	−1.6 ± 0.8	−8.8 ± 3.5
13	Floridan	4.1 ± 1.2	13.5 ± 4	2.4 ± 2.8	−6.2 ± 2.7	−4.9 ± 2.6	−2.1 ± 1.3	−3.1 ± 2.4	1.3 ± 4.0	−13 ± 3.3	−0.5 ± 0.8	−4.0 ± 3.5
14	Pennsylvania	3.2 ± 0.2	8.0 ± 0.6	−0.6 ± 5.1	−6.7 ± 5.1	−6.7 ± 5.1	0.9 ± 1.0	1.8 ± 2.0	−3.7 ± 0.6	−12 ± 5.6	−2.3 ± 1.1	−6.6 ± 2.0

Note. GWSA trends from global hydrological models over 12 years (2003–2014) with uncertainty based on Sen slope method. HI = model simulation with human intervention; NHI = model simulation with no human intervention.

network ($R = 0.2$) (Table S26.1). The Texas Gulf Coast and Coastal Lowlands include distinct unconfined and confined well clusters with higher correlations between the shallow well clusters and GRACE GWSAs ($R = 0.59–0.68$) than those based on the entire well network ($R = 0.33–0.44$).

Effective regional aquifer storage coefficients, estimated from the slope of GRACE GWSAs and GWLAs (Equation 1; Table S26.1) are generally consistent with storage coefficients from previous studies. In the Northern and Southern High Plains, these effective regional storage coefficients (0.07–0.10) are similar to recent estimates in the Central High Plains (0.06–0.10) (Butler et al., 2016) but lower than previous modeled estimates for the Northern and Southern High Plains (0.15–0.17) (Deeds & Jigmond, 2015; Peterson et al., 2016). Lower effective storage coefficients for the Sacramento and San Joaquin/Tulare Basins (0.06–0.27) than the range in modeled values (0.09–0.40) (Faunt, 2009) may result from variations in well populations (specifically well depth and confining conditions) and degree of aquifer confinement. The median effective storage coefficient of the Arizona Alluvial Basins (0.10) is similar to the average specific yield of the uppermost model layer (0.11) in the six GW models assessed for this study. In contrast, the median effective storage coefficient for the Eastern Snake Plain Aquifer (median 0.17) exceeds the modeled value (0.06), which may be related to the low water storage changes increasing uncertainties in effective storage coefficients. In the Mississippi Embayment, the effective storage coefficient (0.23) is within the range of modeled specific yields of the shallow unconfined Mississippi River Valley Alluvium (0.10–0.30). In contrast, the effective storage coefficients are much lower in the U. Colorado, Floridan, and Pennsylvania aquifers, where the correlation with GRACE-GWSA is low. In the Snake and Coastal lowlands aquifers, GRACE-GWSA is more highly correlated with in situ GWSA (for the shallow wells and unconfined aquifers); however, the reported storage coefficients from the models are much lower than the effective value. The reported values for theses aquifers may represent the confined units where the change in storage is small and may result in higher uncertainties in storage coefficients.

3.3. Comparison Between GRACE-Derived Groundwater Storage Anomalies and Those From Regional Groundwater Models

Trends and magnitudes of GWSAs from GRACE were qualitatively compared with those from eight regional GW models (Figures 5; Table S28). The current analysis focused on four models that extend to periods after 2011: Central Valley, Eastern Snake Plain, Southern High Plains, and Mississippi Embayment models, and also the NHP that ends in 2008. There is relatively good correspondence between simulated and GRACE GWSAs in the Central Valley with comparable overall trends, including similar increases in storage during wet periods (2005–2006 and 2010–2011) (Faunt et al., 2016). The good agreement between GRACE GWSAs and the model may be related in part to use of subsidence observations to constrain storage properties more than other regional models. The main difference is the response to the 2007–2009 drought with modeled GWS decline approximately two times that from GRACE GWSAs that may be related to lack of data on SW deliveries for this model. GWSAs from the ESPAM and GRACE both show a significant declining trend approximately -15 km^3/15 yr). In the NHP, there is good agreement between simulated GWS depletion (-20 km^3) (Peterson et al., 2016) and that calculated from GRACE GWSAs (-18 km^3) during the 2003–2006 drought period. The simulated decline in GWSA in the Southern High Plains through 2012 (Deeds & Jigmond, 2015) is approximately two times the decline from GRACE GWSAs, and the regional model does not capture the increase in GWSAs during 2006–2008 related to elevated precipitation. There is a large discrepancy between simulated and GRACE-derived depletion in GWSAs in the Mississippi Embayment, with approximately four times greater modeled GWS depletion (approximately -95 km^3) over the common 2002–2014 period than that derived from GRACE (-24 km^3/12 yr). Possible causes of the discrepancy include overestimation of specific yield or pumpage or underestimation of stream-aquifer interactions or recharge rates. Simulated specific yields for the Mississippi River Valley Alluvium range from 0.10 to 0.30, but values are mostly 0.30, which is very large. Smaller values (<0.20) would improve the match with GRACE GWSAs. Stream leakage may be extremely important in this system as 90% of the GW pumpage is in the Mississippi River Valley Alluvium, which is likely well connected to the stream network. Greater simulated stream leakage could replace storage loss because simulated storage loss is ~2 orders of magnitude less than the flow rates of the large streams simulated in the model (mean annual flow >2.8 m^3/s). In addition, the dense network of smaller streams is not simulated; however, these streams may be an important water source to the aquifer during periods of increased pumping after 1980. Finally, underestimation of simulated

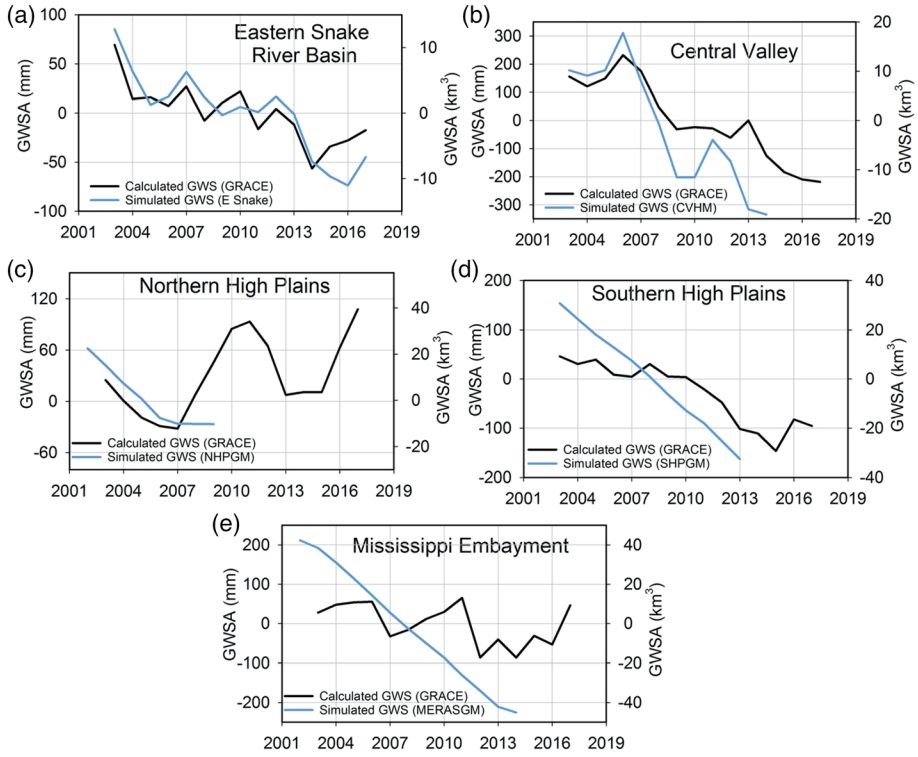

Figure 5. Annual GWSAs from GRACE (black line) and regional models (blue line) for the overlap period in (a) Eastern Snake, (b) California Central Valley Hydrologic Model (CVHM), (c) Northern High Plains (NHP), (d) Southern High Plains (SHP), and (e) Mississippi Embayment Regional Aquifer System (MERAS) groundwater models (GM). The GRACE GWS trend is referenced to the overlap period with the model. The results for the Columbia Plateau, Arizona Alluvial basins, and Texas Gulf Coast (Houston Area model) are shown in Figure S9 (Tables S27 and S28). The Central valley results are based on Sacramento and San Joaquin systems polygon scaled by the ratio of the aquifer area to the hydrological unit area. The results for the Mississippi Embayment in this figure differ from those in Figure 4 because GWS in Figures 3 and 4 is scaled by a factor of 2.6 (ratio of Mississippi Embayment to Mississippi River Valley Alluvium) under the assumption that most of the storage changes occur in the alluvium.

GW recharge, especially during wet periods, could also result in underestimation of the storage increases seen in GRACE data. Current efforts to improve the regional model focus on the shallow Mississippi River Valley Alluvial aquifer with enhanced hydrogeologic mapping based on airborne electromagnetic surveys, a more detailed stream network with aquifer connectivity data, improved estimates of GW pumpage, and revised recharge estimates, including those based on a soil water balance model (Killian et al., 2019; Miller et al., 2018; Reitz et al., 2017; Westenbroek et al., 2010).

Additional models include the Columbia Plateau, which shows little change in modeled GWS, consistent with GRACE estimates, and the Houston Area GW Model, which indicates slight GWS decline, also consistent with GRACE (Figure S9). The Arizona models do not show the increase in storage through 2010 estimated from GRACE. However, after 2011, both simulated and GRACE GWS loss rates are similar when GW use in adjacent unmodeled basins with no capturable streams was included and assumed to be derived from GWS; inclusion of the unsimulated basin withdrawals as derived from storage may slightly overestimate storage loss but is a reasonable assumption especially during the extended drought after 2010.

In summary, GRACE GWS trends are similar or less than those from regional models for many aquifers, with the largest discrepancy in the Mississippi Embayment model which shows approximately four times greater depletion than that from GRACE. The GRACE GWS variability also seems to show more dynamic response to climate variability than regional models do. The relatively good comparison for the Central Valley, especially the increases in storage during wet periods, may be related to the additional constraints on GW withdrawals that result from including subsidence measurements, SW gains/losses, and use of the

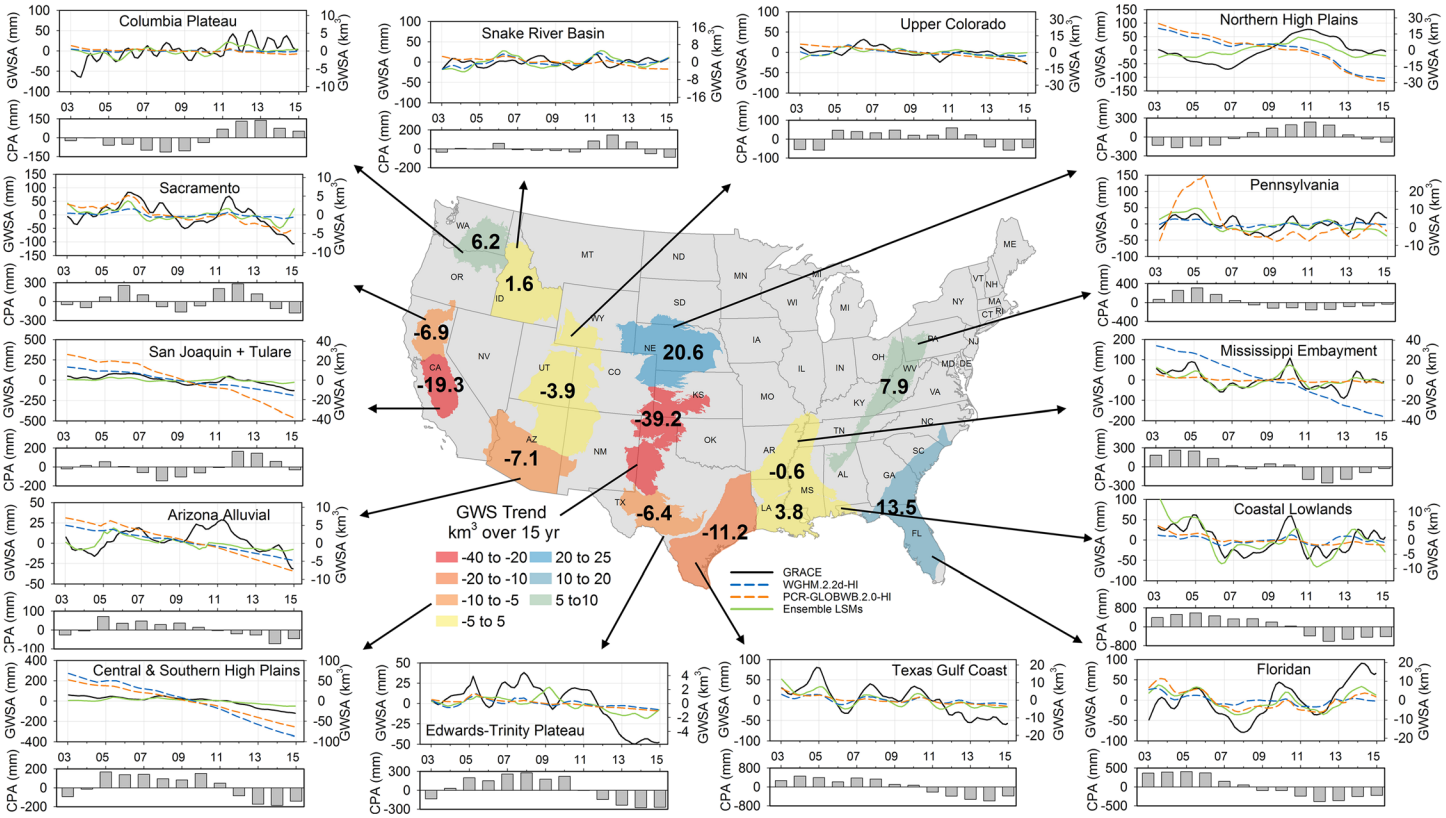

Figure 6. Time series plots: (upper panels) the long-term variability (interannual variability and linear trend) in GWSAs from GRACE (black line) and from global hydrologic models (WGHM-HI and PCR-GLOBWB-HI, HI, human intervention) (Tables S30–S33) and ensemble of land surface models (NOAH-MP, CLSM-F2.5, and CLM-5.0) (Tables S34–S36) in 14 major aquifers in the U.S. (Lower panels) Annual cumulative precipitation anomaly (CPA) as gray bars. The U.S. base map shows GRACE GWS trends in km^3 over the 15 yr monitoring period (2002–2017) (Table 1).

USGS Farm Process Package (Schmid & Hanson, 2009) to help constrain net GW withdrawals (Faunt, 2009). The analysis generally shows that regional GW models may benefit from including GRACE GWSAs as an independent observation, especially where rates of recharge, storage change, pumpage, or stream-aquifer interactions are poorly constrained.

3.4. Comparison Between GRACE-Derived Groundwater Storage Anomalies and Those From Global Models

Global models evaluated include GHMs and LSMs. Trends in GWSAs were calculated for the common modeling period for GHMs and LSMs (April 2003 to December 2014; ~12 yr). The GHMs simulated GW depletion totaling −172 to −186 km^3 in southwestern aquifers (Central Valley and Arizona Alluvial Basins), Central and Southern High Plains, and south-central aquifers (Edwards Trinity Plateau and Texas Gulf Coast), about a factor of 2.4× that from GRACE (−74 km^3; 2003–2014). (Figure 6; Tables 1 and S1). The impact of HI in GHM output was isolated by comparing GHM runs with and without HI (Figure S10). HI accounted for ~95% of the difference between GHM GWSA trends and those from GRACE. Similarly, in the NHP, the GHM simulated depletions of approximately −39 to −43 km^3 are almost entirely due to HI, whereas GRACE data suggest a GWS increase of ~18 km^3. Modeled GWS depletion for the Mississippi Embayment range from −4.7 km^3 (PCR-GLOBWB-2-HI) to −74 km^3 (WGHM-2,2d HI) compared to −12 km^3 from GRACE. The large WGHM modeled depletion is attributed almost entirely to HI. In addition to overestimation of depletion, the GHMs do not track the dynamics of GWSAs relative to GRACE GWSAs. Some challenges in modeling elements of HI can result in large variations in GWS from GHMs (Döll et al., 2016; Wada et al., 2017). Examples include (1) inapplicability of global algorithms for modeling human

management to specific aquifers; (2) uncertainty in water source (SW or GW) for different sectoral uses; (3) challenges with modeling hydrological extremes (e.g., droughts and floods); (4) incorporating irrigation return flow into GW; (5) lack of modeling land uses (e.g., vegetation) and climate interactions; and (5) uncertainty in model inputs (e.g., climate and human use) and water availability.

The ensemble of the LSMs (CLM-5.0, CLSM-F2.5, and NOAH-MP) matches GRACE GWSAs better than the GHMs, which is surprising because most LSMs do not include SWS or HI (Figure 6; Tables 1 and S1). This better match between trends in GRACE GWSAs and LSMs is not thought to be an artifact of the use of LSM SMS to estimate GRACE GWSAs because long-term trends in SMS from all LSMs and GHMs are negligible. There is substantial variability among the LSMs, with some LSMs matching GRACE GWSAs in some basins better than in others (Figure S11). For example, in the San Joaquin/Tulare Basins, the ensemble LSMs simulate -2.6 km^3 relative to -11 km^3 from GRACE. In the NHP, the ensemble LSMs simulate a median increase in GWSA of 9 km^3 relative to 18 km^3 from GRACE. In the Central and Southern High Plains, the ensemble LSM GWSA trend is -15 km^3 relative to GRACE GWSA of -31 km^3. In the Mississippi Embayment, the trend in GWSA from the LSMs is -3.6 km^3 relative to -12 km^3 from GRACE GWSAs.

4. Implications

Groundwater storage is extremely important for water resource assessments, but it has received limited consideration in regional hydrologic studies because of the general lack of independent data for evaluation. GRACE satellites can provide these independent data for large regional aquifers. Estimating GRACE GWSAs relies on subtracting SMSAs, mostly derived from LSMs. LSMs may underestimate SMSAs in areas of thick soils because NLDAS LSMs, used to estimate GWS in this study, only simulate SMS in the top 2 m of the soil profile. However, LSMs may also overestimate SMS because most do not simulate other component storages (SWS or GWS) but may implicitly include them in simulated SMS. Therefore, subtracting LSM-derived SMS from GRACE TWSA could underestimate or overestimate GWSAs. Unsimulated SMS below the simulated 1–2 m depths could also contribute to interannual and decadal variability.

This study demonstrates the utility of aggregating GWL data for comparison with GRACE data and the subsequent estimation of GWS at the aquifer scale. Issues affecting interpretation of results include clustering data, impacts of pumpage, and degree of aquifer confinement. The gridding analysis of GWLs suggests that the composite time series for the analyzed aquifers is independent of the grid resolutions tested, ranging from 10–100 km (Figure S6; Table S25.1). Impacts of pumpage and lack of recovery from irrigation was managed by focusing on winter periods and removing outliers, generally representing <10% of the well records. Previous studies have limited comparison of GRACE data with GWL data in unconfined aquifers and seasonal variations in GWLs (Li et al., 2019); however, limiting the GWL data set to wells with monthly data would reduce the data set by up to 70%. Our analysis focuses on synoptic GWL monitoring during the winter when impacts of irrigation pumpage on GWLs are expected to be low. In addition, completion of wells in unconfined, semiconfined, or confined aquifers is not reported for most aquifers.

Regional GW models are highly complex, with multiple data inputs, many parameters, and potentially high construction costs. Many of these models are designed to improve understanding of aquifer system response to pumping stress to enhance sustainable development. Assessment of these models has generally relied on comparison with GWL data and baseflow to streams (Hill & Tiedeman, 2007). However, results of regional models constrained by comparison with limited observations can be highly nonunique for many reasons, including uncertainties in aquifer parameters (e.g., recharge rates, pumping rates, independent GWS change estimates, and stream aquifer interactions) and limited observations. Also uncertain are GWS variations in highly exploited aquifers, which are commonly estimated from the product of observed GWL change and poorly defined aquifer storage coefficients. Uncertainties in each water budget component accumulate in the GWS term because it is normally unmeasured and often the most uncertain term in the water budget equation. Unfortunately, current parameter estimation methods for models often indicate that the aquifer storage coefficient is not a highly sensitive parameter, resulting in little interest in better definition of this parameter and follow-on studies that focus on other uncertain parameters. Improved definition of aquifer storage coefficients should help reduce model nonuniqueness despite the greater model sensitivity to other parameters. Butler et al. (2020) recognized the need for improved storage coefficients in the Central High Plains and developed a unique approach to the problem using detailed water budget records and stable

climate forcing and recharge. Similarly, observations of GWSAs from GRACE, ground-based gravity monitoring, weighing lysimeters, and recently developed seismic methods provide important tools to reduce uncertainty in models that have large storage changes. Flow between aquifers and streams is another water budget component that is difficult to observe and may be poorly quantified especially for systems with large streams, such as the Mississippi River where rates of stream-aquifer interactions may be far less than streamflow or even less than streamflow uncertainties. Because of these many source of model uncertainty, many regional models may oversimplify aspects of the systems, as is done, for example, with the use of constant recharge rates (Hunt et al., 2007) resulting in underestimation of GWS dynamics. While the current comparisons are generally qualitative because of limited overlap periods with many regional models, results suggest there is potential value in using GRACE data to better constrain regional model calibration. For example, large discrepancies between GRACE GWSAs and regional model results for the Mississippi Embayment are well beyond the uncertainties in GRACE GWSAs and highlight the need to conduct additional studies to better constrain the results.

Large discrepancies between GRACE-derived GWSAs and those from GHMs (PCR-GLOBWB and WGHM) reflect uncertainties in simulating HI in highly impacted aquifers using general rules related to sourcing water (SW and GW), volumes withdrawn and consumed, and irrigation (McDonald et al., 2014; Siebert et al., 2013). These GHMs also underestimate the dynamics in GWS relative to GRACE. While many of the LSMs capture GWS dynamics, some may overestimate natural GW depletion considering that HI is not included in these models. Incorporating GW abstraction and dynamic irrigation schemes in LSMs (e.g., NOAH-MP and HiGW-MAT) help to match GRACE GWS at a coarse resolution; however, it becomes challenging for finer resolution and small aquifers due to lack of representation of the subsurface hydrology in these models (Nie et al., 2019; Pokhrel et al., 2015). Some previous studies attribute discrepancies between GRACE and LSMs to lack of HI; however, the GHMs, which include HI, do not perform better than the LSMs. Hence, more emphasis is required to quantify abstractions and improve the GHMs in such intensely impacted regions.

GRACE data highlight the value of GWSAs for understanding water resources. One of the primary disadvantages of the satellite data is the low spatial resolution. However, the GRACE data have been shown here to have potential value for regional GW model calibration and could be complemented with ground-based gravity monitoring to observe spatial variations in water storage at much higher resolution than GRACE, as shown in the Arizona Alluvial Basins aquifer studies (Pool, 2008; Pool & Anderson, 2008) and other smaller-scale studies (Kennedy, 2016; Pool & Schmidt, 1997).

5. Conclusions

GRACE satellites provide an independent data set of TWSAs and estimated GWSAs in major aquifers in the United States to better understand spatiotemporal variability in water storage of individual aquifers. Generally good agreement between GRACE GWSAs and GWLAs for most aquifers indicates that both are tracking the dynamics of GWS, which increases during wet periods and decreases during drought. Both data sets show long-term declining trends in the southwestern and south-central United States (California Central Valley, Arizona Alluvial Basins, Central and Southern High Plains, Edwards Trinity Plateau, and Texas Gulf Coast), totaling approximately −90 km^3 over the 15 yr period (2002–2017). GWS in most other aquifers is stable or increasing by up to 21 km^3 in the NHP. Good correspondence between GRACE GWSAs and GWLAs is shown by correlation coefficients ranging from 0.52 to 0.95 in most systems, which increases confidence in using GRACE data to evaluate regional GWSAs. Effective storage coefficients derived from GWL and GRACE GWSA data are generally consistent with reported storage coefficients in the uppermost unconfined parts of aquifer systems.

Qualitative comparison of GRACE GWSAs and GWSAs from eight regional GW models shows generally good correspondence for the regional aquifers with the most GW use (Central Valley, High Plains, and Arizona Alluvial Basins aquifers). However, modeled GW depletion in the Mississippi Embayment is approximately four times greater than that estimated from GRACE, greatly exceeding GRACE uncertainties. This discrepancy may reflect uncertainties in modeled storage parameters, overestimated pumpage, underestimated stream capture, and/or underestimated GW recharge.

GHMs overestimate depletion in major aquifers, such as the Central Valley (GHM = −30 to −72 km^3/12 yr; GRACE = −15 km^3/12 yr), NHP (GHM = −39 to −43 km^3/12 yr; GRACE = 18 km^3/12 yr), Central and Southern High Plains (GHM = −90 to −127 km^3/12 yr; GRACE = −31 km^3/12 yr), and Mississippi Embayment (GHM = −5 to −74 km^3/12 yr; GRACE = −12 km^3/12 yr). This overestimation can mostly be accounted for by missing elements in modeling the HI. In contrast, GWSAs from global LSMs agree better with GRACE than the GHMs; however, some of the modeled depletion may be for the wrong reason as these models do not include HI. There is substantial variability between the GHMs and also among the LSMs, suggesting high levels of model uncertainties.

This study highlights the value of water storage data in assessing GW resources and demonstrates that incorporating GWS into regional and global models would help to reduce uncertainties. Each source of data has uncertainties; however, combining remote sensing, global and regional modeling, and ground-based monitoring should help to reduce conceptual and numerical model uncertainties in GWS, which should also help constrain recharge rates and stream-aquifer interactions. This type of analysis substantially advances our understanding of large-scale aquifer systems and raises new questions for further study.

Data Availability Statement

Data supporting this research are available online without restrictions for GRACE, NLDAS models, in situ reservoirs, in situ groundwater, SNODAS, and global hydrological models (supporting information). The shape file of major aquifers in the United States and results of this study are available at online (https://doi.org/10.18738/T8/JSUIJT).

Acknowledgments
This work was supported by John Wesley Powell Center for Analysis and Synthesis, U.S. Geological Survey (USGS), the National Science Foundation (NSF), and the Jackson school of Geosciences Endowment. We thank the editor and the three anonymous reviewers for their insightful comments and suggestions, which have greatly improved this manuscript.

References

Arciniega-Esparza, S., Breña-Naranjo, J. A., Hernández-Espriú, A., Pedrozo-Acuña, A., Scanlon, B. R., Nicot, J. P., et al. (2017). Baseflow recession analysis in a large shale play: Climate variability and anthropogenic alterations mask effects of hydraulic fracturing. *Journal of Hydrology*, *553*, 160–171.

Argus, D. F., Landerer, F. W., Wiese, D. N., Martens, H. R., Fu, Y., Famiglietti, J. S., et al. (2017). Sustained water loss in California's mountain ranges during severe drought from 2012 to 2015 inferred from GPS. *Journal of Geophysical Research: Solid Earth*, *122*, 10,559–10,585. https://doi.org/10.1002/2017JB014424

Barrett, A. P. (2003). National operational hydrologic remote sensing center snow data assimilation system (SNODAS) products at NSIDC, National Snow and Ice Data Center, Cooperative Institute for Research in Environmental Sciences.

Bellino, J. C., Kuniansky, E. L., O'Reilly, A. M., & Dixon, J. F. (2018). Hydrogeologic setting, conceptual groundwater flow system, and hydrologic conditions 1995–2010 in Florida and parts of Georgia, Alabama, and South Carolina, U.S. Geol. Surv. Sci. Inv. Rept 2018–5030, 103.

Brena-Naranjo, J. A., Kendall, A. D., & Hyndman, D. W. (2014). Improved methods for satellite-based groundwater storage estimates: A decade of monitoring the High Plains aquifer from space and ground observations. *Geophysical Research Letters*, *41*, 6167–6173. https://doi.org/10.1002/2014GL061213

Brookfield, A. E., Hill, M. C., Rodell, M., Loomis, B., Stotler, R. L., Porter, M. E., & Bohling, G. C. (2018). In situ and GRACE-based groundwater observations: Similarities, discrepancies, and evaluation in the High Plains aquifer in Kansas. *Water Resources Research*, *54*, 8034–8044. https://doi.org/10.1029/2018WR023836

Butler, J. J. Jr, Bohling, G. C., Whittemore, D. O., & Wilson, B. B. (2020). A roadblock on the path to aquifer sustainability: underestimating the impact of pumping reductions. *Environmental Research Letters*, *15*(1), 014003.

Butler, J. J. Jr., Whittemore, D. O., Wilson, B. B., & Bohling, G. C. (2016). A new approach for assessing the future of aquifers supporting irrigated agriculture. *Geophysical Research Letters*, *43*, 2004–2010. https://doi.org/10.1002/2016GL067879

Castle, S. L., Reager, J. T., Thomas, B. F., Purdy, A. J., Lo, M. H., Famiglietti, J. S., & Tang, Q. H. (2016). Remote detection of water management impacts on evapotranspiration in the Colorado River Basin. *Geophysical Research Letters*, *43*, 5089–5097. https://doi.org/10.1002/2016GL068675

Clark, B. R., & Freiwald, D. A. (2011). A new tool to assess groundwater resources in the Missisippi Embayment U.S. Geol. Surv. Fact Sheet 2011–3115.

Clark, B. R., Hart, R. M., & Gurdak, J. J. (2011). Groundwater availability of the Mississippi Embayment, U.S. Geol. Surv. Prof. Paper 1785, 62.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, *6*(1), 3–73.

Deeds, N. E., & Jigmond, M. (2015). Numerical model report for the High Plains Aquifer System Groundwater Availability Model Final Contract Report prepared by INTERA Inc. for Texas Water Development Board, variably paginated.

Dieter, C. A., Maupin, M. A., Caldwell, R. R., Harris, M. A., Ivahnenko, T. I., Lovelace, J. K., et al. (2018). Estimated use of water in the United States in 2015, U.S. Geological Survey Circular 1441, 65.

Döll, P., Douville, H., Guntner, A., Müller Schmied, H., & Wada, Y. (2016). Modelling freshwater resources at the global scale: Challenges and prospects. *Surveys in Geophysics*, *37*(2), 195–221.

Ek, M. B., Cai, X., Mitchell, K. E., Rodell, M., Li, B., Huang, M., et al. (2017). Comparison and assessment of three advanced land surface models in simulating terrestrial water storage components over the United States. *Journal of Hydrometeorology*, *18*(3), 625–649.

Ely, D. M., Burns, E. R., Morgan, D. S., & Vaccaro, J. J. (2014). Numerical simulation of groundwater flow in the Columbia Plateau Regional Aquifer System, Idaho, Oregon, and Washington, U.S. Geol. Surv. Scientific Inv. Rept. 2014–5127, 89.

Faunt, C. C. (2009). Groundwater availability of the Central Valley Aquifer, California; US Geological Survey Prof. Paper 1766, 173.
Faunt, C. C., & Sneed, M. (2015). Water availability and subsidence in California's Central Valley. *San Francisco Estuary and Watershed Science*, *13*(3). https://doi.org/10.15447/sfews.2015v13iss3art4
Faunt, C. C., Sneed, M., Traum, J., & Brandt, J. T. (2016). Water availability and land subsidence in the Central Valley, California, USA. *Hydrogeology Journal*, *24*(3), 675–684.
Freeze, R. A., & Cherry, J. A. (1979). *Groundwater* (Vol. 176, pp. 161–177). Englewood Cliffs, NJ: Prentice-Hall.
George, P. G., Mace, R. E., & Petrossian, R. (2011). Aquifers of Texas, Texas Water Development Board Report No. 380, 172.
Harbaugh, A. W., Banta, E. R., Hill, M. C., & McDonald, M. G. (2000). MODFLOW-2000, the U. S. Geological Survey modular ground-water model-user guide to modularization concepts and the ground-water flow process, open-file report. U. S. Geological Survey(92), 134.
Hill, M. C., & Tiedeman, C. R. (2007). *Effective groundwater model calibration: With analysis of data, sensitivities, predictions, and uncertainty* (1st Ed.). New Jersey, USA: John Wiley & Sons.
Hunt, R. J., Doherty, J., & Tonkin, M. J. (2007). Are models too simple? Arguments for increased parameterization. *Ground Water*, *45*(3), 254–262. https://doi.org/10.1111/j.1745-6584.2007.00316.x
Kasmarek, M. C. (2012). Hydrogeology and simulation of groundwater flow and land-surface subsidence in the northern part of the Gulf Coast Aquifer System, Texas, 1891–2009, U.S. Geol. Surv. Scientific Inv. Rept. 2012–5154, 55.
Kennedy, J. R. (2016). Understanding infiltration and groundwater flow at an artificial recharge facility using time-lapse gravity data.
Killian, C. D., Asquith, W. H., Barlow, J. R. B., Bent, G. C., Kress, W. H., Barlow, P. M., & Schmitz, D. W. (2019). Characterizing groundwater and surface-water interaction using hydrograph-separation techniques and groundwater-level data throughout the Mississippi Delta, USA. *Hydrogeology Journal*, *27*(6), 2167–2179.
Kim, H., Yeh, P. J. F., Oki, T., & Kanae, S. (2009). Role of rivers in the seasonal variations of terrestrial water storage over global basins. *Geophysical Research Letters*, *36*, L17402. https://doi.org/10.1029/2009GL039006
Konikow, L. F. (2013). Groundwater depletion in the United States (1900–2008). U.S. Geol. Surv. Scientific Investigation Report 2013–5079, 63.
Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., et al. (2019). The Community Land Model Version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, *11*, 4245–4287. https://doi.org/10.1029/2018MS001583
Li, B. L., Rodell, M., Kumar, S., Beaudoing, H. K., Getirana, A., Zaitchik, B. F., et al. (2019). Global GRACE data assimilation for groundwater and drought monitoring: Advances and challenges. *Water Resources Research*, *55*, 7564–7586. https://doi.org/10.1029/2018WR024618
Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137.
Long, D., Chen, X., Scanlon, B. R., Wada, Y., Hong, Y., Singh, V. P., et al. (2016). Have GRACE satellites overestimated groundwater depletion in the Northwest India aquifer? *Scientific Reports*, *6*, 24398. https://doi.org/10.1038/srep24398
Longuevergne, L., Scanlon, B. R., & Wilson, C. R. (2010). GRACE hydrological estimates for small basins: Evaluating processing approaches on the High Plains Aquifer, USA. *Water Resources Research*, *46*, W11517. https://doi.org/10.1029/2009WR008564
Longuevergne, L., Wilson, C. R., Scanlon, B. R., & Cretaux, J. F. (2013). GRACE water storage estimates for the Middle East and other regions with significant reservoir and lake storage. *Hydrology and Earth System Sciences*, *17*, 4817–4830. https://doi.org/10.5194/hess-17-4817-2013 2013
Marton, D., Stary, M., & Mensik, P. (2015). Analysis of the influence of input data uncertainties on determining the reliability of reservoir storage capacity. *Journal of Hydrology and Hydromechanics*, *63*(4), 287–294.
McDonald, R. I., Weber, K., Padowski, J., Flörke, M., Schneider, C., Green, P. A., et al. (2014). Water on an urban planet: Urbanization and the reach of urban water infrastructure. *Global Environmental Change*, *27*, 96–105. https://doi.org/10.1016/j.gloenvcha.2014.04.022
McGuire, V. L. (2017). Water-level and recoverable water in storage changes, High Plains aquifer, predevelopment to 2015 and 2013–15: U. S. Geological Survey Scientific Investigations Report 2017–5040, 14, https://doi.org/10.3133/sir20175040
Miller, B. V., Adams, R. F., Stocks, S., Wilson, J. L., David, S. C., & Kress, W. H. (2018). Waterborne resistivity surveys for streams in the Mississippi Alluvial Plain 2017, edited, U.S. Geological Survey.
Miller, J. A. (1999). Groundwater Atlas of the United States: Introduction and National Summary, Hydrologic Atlas HA730-A.
Mueller-Schmied, H., Adam, L., Eisner, S., Fink, G., Flörke, M., Kim, H., et al. (2016). Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use. *Hydrology and Earth System Sciences*, *20*(7), 2877–2898. https://doi.org/10.5194/hess-20-2877-2016
Mulder, G., Olsthoorn, T. N., Al-Manmi, D. A. M. A., Schrama, E. J. O., & Smidt, E. H. (2014). Identifying water mass depletion in Northern Iraq observed by GRACE. *Hydrology and Earth System Sciences Discussions*, *11*, 11,533–11,563.
Nie, W., Zaitchik, B. F., Rodell, M., Kumar, S. V., Arsenault, K. R., Li, B., & Getirana, A. (2019). Assimilating GRACE into a land surface model in the presence of an irrigation-induced groundwater trend. *Water Resources Research*, *55*, 11,274–11,294. https://doi.org/10.1029/2019WR025363
Peterson, S. M., Flynn, A. T., & Traylor, J. P. (2016). Groundwater-flow model of the northern High Plains aquifer in Colorado, Kansas, Nebraska, South Dakota, and Wyoming Rep. 2328–0328, US Geological Survey.
Peterson, S. M., Traylor, J. P., Guira, M. (2020). Groundwater availability of the Northern High Plains aquifer in Colorado, Kansas, Nebraska, South Dakota, and Wyoming, U.S. Geological Survey Professional Paper (Vol. 1864, p. 57). https://doi.org/10.3133/pp1864
Pokhrel, Y. N., Koirala, S., Yeh, P. J. F., Hanasaki, N., Longuevergne, L., Kanae, S., & Oki, T. (2015). Incorporation of groundwater pumping in a global land surface model with the representation of human impacts. *Water Resources Research*, *51*, 78–96. https://doi.org/10.1002/2014WR015602
Pool, D. (2008). The utility of gravity and water-level monitoring at alluvial aquifer wells in southern ArizonaGravity and water-level monitoring in Arizona. *Geophysics*, *73*(6), WA49–WA59.
Pool, D. R., & Anderson, M. T. (2008). Ground-water storage change and land subsidence in Tucson Basin and Avra Valley, Southeastern Arizona, 1998–2002 Rep. 2328–0328, Geological Survey (US).
Pool, D. R., & Schmidt, W. (1997). Measurement of ground-water storage change and specific yield using the temporal-gravity method near Rillito Creek, Tucson, Arizona, US Department of the Interior, US Geological Survey.
Reager, J. T., Gardner, A. S., Famiglietti, J. S., Wiese, D. N., Eicker, A., & Lo, M. H. (2016). A decade of sea level rise slowed by climate-driven hydrology. *Science*, *351*(6274), 699–703. https://doi.org/10.1126/science.aad8386
Reitz, M., Sanford, W. E., Senay, G. B., & Kress, W. H. (2017). Estimating water budget components of evapotranspiration, recharge, and runoff for Mississippi and the Mississippi Alluvial Plain, Abstract: Mississippi Water Resources Research Institute Conference, 46th, 2017.

Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoing, H. K., Landerer, F. W., & Lo, M. H. (2018). Emerging trends in global freshwater availability. *Nature*, *557*(7707), 651–659. https://doi.org/10.1038/s41586-018-0123-1

Scanlon, B. R., Levitt, D. G., Reedy, R. C., Keese, K. E., & Sully, M. J. (2005). Ecological controls on water-cycle response to climate variability in deserts. *Proceedings of the National Academy of Sciences*, *102*(17), 6033–6038.

Scanlon, B. R., Longuevergne, L., & Long, D. (2012). Ground referencing GRACE satellite estimates of groundwater storage changes in the California Central Valley, USA. *Water Resources Research*, *48*, W04520. https://doi.org/10.1029/2011WR011312

Scanlon, B. R., Reedy, R. C., & Tachovsky, J. A. (2007). Semiarid unsaturated zone chloride profiles: Archives of past land use change impacts on water resources in the southern High Plains, United States. *Water Resources Research*, *43*, W03437. https://doi.org/10.1029/2006WR005769

Scanlon, B. R., Zhang, Z., Reedy, R. C., Pool, D. R., Save, H., Long, D., et al. (2015). Hydrologic implications of GRACE satellite data in the Colorado River Basin. *Water Resources Research*, *51*, 9891–9903. https://doi.org/10.1002/2015WR018090

Schmid, W., & Hanson, R.T. (2009). *The farm process version 2 (FMP2) for MODFLOW-2005: modifications and upgrades to FMP1*. California Water Science Center: US Geological Survey.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's Tau. *Journal of the American Statistical Association*, *63*, 1379–1389. https://doi.org/10.2307/2285891

Seyoum, W. M., & Milewski, A. M. (2016). Monitoring and comparison of terrestrial water storage changes in the northern high plains using GRACE and in-situ based integrated hydrologic model estimates. *Advances in Water Resources*, *94*, 31–44. https://www.sciencedirect.com/science/article/pii/S0309170816301166?via%3Dihub#!

Siebert, S., Henrich, V., Frenken, K., & Burke, J. (2013). Global map of irrigation areas Version 5. *Rheinische Friedrich-Wilhelms-University Bonn, Germany/Food and Agriculture Organization of the United Nations, Rome, Italy*, *2*, 1299–1327.

Strassberg, G., Scanlon, B. R., & Chambers, D. (2009). Evaluation of groundwater storage monitoring with the GRACE satellite: Case study of the High Plains aquifer, central United States. *Water Resources Research*, *45*, W05410. https://doi.org/10.1029/2008WR006892

Sutanudjaja, E. H., van Beek, R., Wanders, N., Wada, Y., Bosmans, J. H. C., Drost, N., et al. (2018). PCR-GLOBWB 2: A 5 arcmin global hydrological and water resources model. *Geoscientific Model Development*, *11*(6), 2429–2453. https://doi.org/10.5194/gmd-11-2429-2018

Tapley, B. D., Bettadpur, S., Watkins, M., & Reigber, C. (2004). The gravity recovery and climate experiment: Mission overview and early results. *Geophysical Research Letters*, *31*, L09607. https://doi.org/10.1029/2004GL019920

Twining, B. V., & Fisher, J. C. (2012). Multilevel groundwater monitoring of hydraulic head and temperature in the Eastern Snake River Plain Aquifer, Idaho National Laboratory, Idaho, 2009–10, U.S. Geol. Surv. Sci. Inv. Rept. 2012–5259, (p. 43).

Voss, K. A., Famiglietti, J. S., Lo, M. H., de Linage, C., Rodell, M., & Swenson, S. C. (2013). Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region. *Water Resources Research*, *49*, 904–914. https://doi.org/10.1002/wrcr.20078

Wada, Y., Bierkens, M. F., De Roo, A., Dirmeyer, P. A., Famiglietti, J. S., Hanasaki, N., et al. (2017). Human–water interface in hydrological modelling: Current status and future directions. *Hydrology and Earth System Sciences*, *21*(8), 4169–4193.

Wada, Y., van Beek, L. P. H., van Kempen, C. M., Reckman, J., Vasak, S., & Bierkens, M. F. P. (2010). Global depletion of groundwater resources. *Geophysical Research Letters*, *37*, L20402. https://doi.org/10.1029/2010GL044571

Wei, F., Longuevergne, L., Kusche, J., Liang, S., Zhang, Y., Scanlon, B. R., et al. (2017). Groundwater storage variations in the North China Plain using multiple space geodetic observations G53B-04 presented at 2017 AGU Fall Meeting, New Orleans, La., 11–15 Dec.

Westenbroek, S. M., Kelson, V. A., Dripps, W. R., Hunt, R. J., & Bradbury, K. R. (2010). SWB—A modified Thornthwaite-Mather Soil Water Balance code for estimating groundwater rehcarge: U.S. Geological Survey Techniques and Methods 6-A31, (p. 60).

Wiese, D. N., Landerer, F. W., & Watkins, M. M. (2016). Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution. *Water Resources Research*, *52*, 7490–7502. https://doi.org/10.1002/2016WR019344

Xia, Y., Cosgrove, B. A., Ek, M. B., Sheffield, J., Luo, L., Wood, E. F., et al. (2013). Overview of the North American Land Data Assimilation System (NLDAS). In *Land surface observation, modeling and data assimilation* (pp. 337–377). New Jersey, USA: World Scientific.

Xia, Y., Sheffield, J., Ek, M. B., Dong, J., Chaney, N., Wei, H., et al. (2014). Evaluation of multi-model simulated soil moisture in NLDAS-2. *Journal of Hydrology*, *512*, 107–125.

Xiao, M., Koppa, A., Mekonnen, Z., Pagan, B. R., Zhan, S. A., Cao, Q. A., et al. (2017). How much groundwater did California's Central Valley lose during the 2012–2016 drought? *Geophysical Research Letters*, *44*, 4872–4879. https://doi.org/10.1002/2017GL073333