California State University, Monterey Bay

# Digital Commons @ CSUMB

AES Faculty Publications and Presentations          Department of Applied Environmental Science

2-10-2022

# Assessing the utility of remote sensing data to accurately estimate changes in groundwater storage

Aakash Ahamed
*Stanford University*

Rosemary Knight
*Stanford University*

Sarfaraz Alam
*Stanford University*

Rich Pauloo
*University of California, Davis*

Forrest Melton
*California State University, Monterey Bay*

Follow this and additional works at: https://digitalcommons.csumb.edu/aes_fac

# Assessing the utility of remote sensing data to accurately estimate changes in groundwater storage

Aakash Ahamed [a,*], Rosemary Knight [a], Sarfaraz Alam [a], Rich Pauloo [b], Forrest Melton [c,d]

[a] Department of Geophysics, Stanford University, 397 Panama Mall, Stanford, CA 94305, United States of America
[b] Hydrologic Sciences, University of California, Davis, One Shields Avenue, Davis, CA 95616, United States of America
[c] Department of Applied Environmental Sciences, California State University, Monterey Bay, 100 Campus Center, Seaside, CA 93955, United States of America
[d] Biospheric Sciences Branch, NASA Ames Research Center, Mail Stop 245, Moffett Field, CA 94035, United States of America

## HIGHLIGHTS

- Groundwater storage changes estimated using ensemble remote sensing water balance.
- Remote sensing-based method can provide reliable estimates of changes in storage.
- Long-term changes in groundwater storage during droughts were well captured.
- Uncertainty is highest in the water balance component evapotranspiration.
- Remote sensing data allow groundwater storage estimates at sub-annual timescales.

## GRAPHICAL ABSTRACT
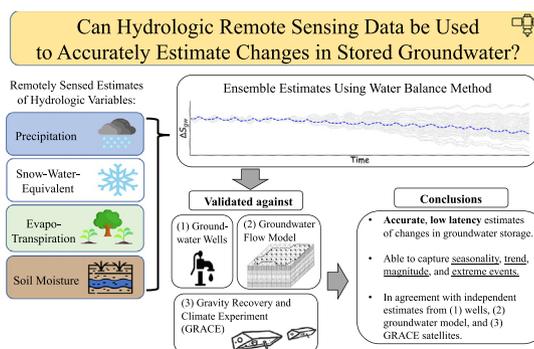


## ARTICLE INFO

## ABSTRACT

Accurate and timely estimates of groundwater storage changes are critical to the sustainable management of aquifers worldwide, but are hindered by the lack of in-situ groundwater measurements in most regions. Hydrologic remote sensing measurements provide a potential pathway to quantify groundwater storage changes by closing the water balance, but the degree to which remote sensing data can accurately estimate groundwater storage changes is unclear. In this study, we quantified groundwater storage changes in California's Central Valley at two spatial scales for the period 2002 through 2020 using remote sensing data and an ensemble water balance method. To evaluate performance, we compared estimates of groundwater storage changes to three independent estimates: GRACE satellite data, groundwater wells and a groundwater flow model. Results suggest evapotranspiration has the highest uncertainty among water balance components, while precipitation has the lowest. We found that remote sensing-based groundwater storage estimates correlated well with independent estimates; annual trends during droughts fall within 15% of trends calculated using wells and groundwater models within the Central Valley. Remote sensing-based estimates also reliably estimated the long-term trend, seasonality, and rate of groundwater depletion during major drought events. Additionally, our study suggests that the proposed method estimate changes in groundwater at sub-annual latencies, which is not currently possible using other methods. The findings have implications for improving the understanding of aquifer dynamics and can inform regional water managers about the status of groundwater systems during droughts.

* Corresponding author.
E-mail address: aahamed@stanford.edu (A. Ahamed).

# 1. Introduction

Groundwater resources are critical for worldwide food and water security. Globally, groundwater constitutes 35% of all water use, 33% of irrigation water (Siebert et al., 2010), and supports the drinking water needs of at least 2 billion people (Alley et al., 2002). Presently, groundwater resources face unprecedented stress due to climate change and population growth (Wada et al., 2010; Famiglietti, 2014; Bierkens and Wada, 2019). Arid and semiarid areas rely on groundwater to meet demand during droughts, but increasingly severe drought events due to climate change (e.g. Diffenbaugh et al., 2015; Cook et al., 2015; Swain et al., 2018) coupled with anthropogenic groundwater overdraft (e.g. Wada et al., 2010), jeopardize the sustainability and longevity of groundwater aquifers. These aquifers currently support thriving population centers and economies around the world. As such, there is a growing and critical need for reliable and timely estimates of changes in groundwater storage to establish groundwater sustainability and increase resiliency to drought events.

Central to groundwater management is the ability to reliably estimate and monitor the change in stored groundwater ($\Delta S_{gw}$) over time and at the spatial scale(s) of interest. Accurate estimates of $\Delta S_{gw}$ over a long time record and with frequent temporal sampling provide critical information on seasonal variability, long term trend, and the general health of a groundwater system. A number of approaches have been explored as a means to estimate $\Delta S_{gw}$ at various scales. Five commonly used approaches or sources of data are: (1) satellite, airborne and ground-based gravimetric measurements, (2) groundwater monitoring wells, (3) surface deformation data derived from GPS and interferometric synthetic aperture radar (InSAR), (4) numerical models, and (5) the water balance method. Although these methods have been applied in numerous previous studies, obtaining reliable and timely estimates of $\Delta S_{gw}$ is still a major global challenge (Bierkens and Wada, 2019), especially in regions without monitoring wells.

Measurements of changes in gravity can be used to directly estimate changes in the mass of subsurface water. At the continental and regional scale (>150,000 km$^2$), observations of the change in gravity from the Gravity Recovery and Climate Experiment (GRACE) satellites, corrected for mass variations in surface water, soil moisture, and snow, have been used to determine $\Delta S_{gw}$ (e.g. Rodell et al., 2009; Famiglietti et al., 2011; Scanlon et al., 2012). Though it provides a direct measure, the limitation with GRACE is the footprint of the measurement, currently >150,000 km$^2$. Thus, although GRACE is an excellent tool for regional-scale estimates of groundwater storage changes, it cannot provide monitoring capabilities at scales practical for groundwater management. Airborne gravimetric surveys (e.g. Pool and Eychaner, 1995) can resolve site-specific storage at higher resolution, but require corrections for soil moisture and vegetation water to derive estimates of storage. Measurements by ground-based gravimeters have been used to monitor $\Delta S_{gw}$ at specific locations between the times of measurement, but cannot realistically provide the consistent spatial and temporal coverage needed to support groundwater management.

Water level measurements made in wells are a common site-specific approach to estimate $\Delta S_{gw}$, and can be readily upscaled when well coverage is extensive throughout a region (e.g. Scanlon et al., 2012; Alam et al., 2021). With densely sampled measurements of the changes in the water levels and knowledge of the appropriate storage coefficients, a high level of accuracy can be presumed. The highest quality measurements are typically acquired using monitoring wells, but these are expensive to install and often suffer from inadequate spatial and temporal sampling and coverage. The accuracy of water levels measured in public supply, irrigation or domestic wells is prone to human error and the possible influence of groundwater pumping (in the well itself or in nearby wells) on ambient groundwater levels during the measurement. Wells also typically lack information on the storage coefficient of the underlying aquifer needed to accurately estimate $\Delta S_{gw}$ from water levels.

GPS and Interferometric Synthetic Aperture Radar (InSAR) data can provide indirect measurements of $\Delta S_{gw}$ at the regional and sub-regional scale (e.g. Borsa et al., 2014; Argus et al., 2017; Smith et al., 2017). These data can be used to model the link between the observed surface deformation and $\Delta S_{gw}$ in the underlying system (e.g. Argus et al., 2017; Smith et al., 2017). While invaluable as a means of monitoring changes over a large area, accurate estimates at the spatial scale required for groundwater management would require significant investment to determine the link between surface deformation and $\Delta S_{gw}$ for the specific groundwater system of interest.

Groundwater models that solve groundwater flow equations using finite-difference or finite-element grids can also be used to estimate $\Delta S_{gw}$ at local to sub-regional scales (10 km$^2$–5000 km$^2$) (e.g. Faunt et al., 2009; Brush et al., 2013). Groundwater models, however, require accurate subsurface and water level data to parameterize and calibrate, are computationally-intensive, and typically lag years behind in terms of incorporating new data to provide ongoing estimates of $\Delta S_{gw}$. For instance, regional groundwater models in the Central Valley of California (Brush et al., 2013) run until 2015. These models are very useful for considering the impact on $\Delta S_{gw}$ of changing conditions, such as climate, land use, and management practices, but are not suitable for weekly/monthly monitoring. Moreover, developing physically-based groundwater models in most parts of the world is not a realistic option due to a lack of data required for model parameterization.

An alternate approach to estimate $\Delta S_{gw}$ is through application of a water balance equation. Such an approach has the advantage of scale independence, rapid computation, fewer (and less expensive) on-the-ground data requirements compared to other methods, and an increasing number of data sources available to estimate the required variables (Sahoo et al., 2011; Lakshmi, 2016; Pan et al., 2012). The change in groundwater storage ($\Delta S_{gw}$) for an area can be approximated using the equation:

$$\Delta S_{gw} = P + Q_{in} - Q_{out} - Q_{gw\;net} - \Delta S_{sm} - \Delta S_{SWE} - ET - \Delta S_R, \qquad (1)$$

where $P$ is precipitation; $Q_{in}$ and $Q_{out}$ out are surface water flow in and out of the area; $Q_{gw\;net}$ is net groundwater flow in and out of the area; $\Delta S_{SM}$ is the change in soil moisture, defined as moisture contained within the soil and the top few meters of the vadose zone; $\Delta S_{SWE}$ is the change in snow-water-equivalent; $ET$ is evapotranspiration; and $\Delta S_R$ is the change in surface water storage. Groundwater abstraction is not included in the Eq. (1) because there is typically no transfer of pumped groundwater in/out of the study domain.

Accurate data describing each hydrologic component in Eq. (1) are needed to obtain reliable estimates of $\Delta S_{gw}$. There are typically three sources of data that can be used to solve the water balance equation: (1) ground-based measurements which include any measurement that can be made using instruments deployed from the ground surface (e.g., rain gauges, stream gauges, flux towers, snow pillows, reservoir staff gauges); (2) remote sensing measurements which include any measurements made by satellite-deployed instruments; and (3) land surface models that solve water and energy balance equations on a fixed-resolution grid to predict hydrologic variables. Among the three methods, remote sensing data is uniquely capable of providing spatially and temporally continuous measurement of water balance components at high resolution and global coverage.

Reliable remotely sensed estimates of $\Delta S_{gw}$ could eventually have applications worldwide for groundwater science and management where the only available data might be satellite data. Ground-based measurements can be highly accurate, but are always limited in terms of the spatial extent of the acquired data. Various methods are used for upscaling or interpolation of ground-based measurements, but such methods inevitably introduce a high level of uncertainty into the derived estimates. Land-surface models are extremely useful for scenario analysis, but do not typically account for anthropogenic interventions like reservoirs and irrigation, and use a highly simplified

representation of the groundwater system, often as a single layer with fixed depth (Niu et al., 2011; Barlage et al., 2015); limiting their effectiveness for ongoing monitoring.

The rapid advancement of remotely sensed data as a tool to acquire information about global hydrologic systems has presented an opportunity, as highlighted by Becker (2006), and Lakshmi (2016), to use satellite data to determine some of the variables required to calculate $\Delta S_{gw}$. The adoption of remote sensing data, wherever possible, in estimating $\Delta S_{gw}$ using the water balance method is appealing because of global coverage, free and near real-time data access, and frequent revisit time (hours to weeks), which could make possible continuous monitoring of $\Delta S_{gw}$ at regional and subregional scales practical for groundwater management and decision making. Moreover, a new generation of satellites designed to measure unknown hydrologic variables (e.g. streamflow), and improve the resolution of existing Earth observing satellites, will potentially increase the value and utility of methods that rely primarily on remote sensing data. However, one major challenge inherent to the use of remote sensing data is the uncertainty arising from multiple sources and the variable accuracy in different regions (e.g. Tian and Peters-Lidard, 2010; Senay et al., 2020). Despite widespread data availability, the strengths and weaknesses of remote sensing data to estimate $\Delta S_{gw}$ still remain unclear.

The objective of this study was to assess the viability of remote sensing data to accurately estimate groundwater storage changes through the water balance method. To address this, we (1) compared the magnitude, seasonality and uncertainty of multiple data sources (i.e., remote sensing, ground-based measurements, land surface models) for each variable in the water balance equation, (2) estimated $\Delta S_{gw}$ using different combinations of input datasets, and (3) assessed the performance and uncertainty of $\Delta S_{gw}$ estimates arising from various combinations of datasets by validating against independent estimates. Due to the availability of multiple independent $\Delta S_{gw}$ estimates from various sources and spatial scales, the Central Valley of California and its watershed were selected as suitable study areas to evaluate the performance of remote sensing based $\Delta S_{gw}$ estimates. Noting that careful attention should be given to defining boundaries without significant net groundwater flux when performing the analysis, results from this study can provide insights about how remote sensing data can help quantify $\Delta S_{gw}$ in semiarid and irrigated regions with alluvial aquifers. Moreover, there is a critical need for ongoing and up-to-date estimates of groundwater storage changes. The completed research shows how remote sensing data can help produce estimates of $\Delta S_{gw}$ on an ongoing and up-to-date basis that cannot be achieved with other approaches.

## 2. Study area

The Central Valley (CV; ~55,000 km$^2$), and Central Valley Watershed (CVWS; ~150,000 km$^2$) within California were selected as target locations to test remote sensing data as a means to estimate groundwater storage changes using the water balance method. The CV is a flat, agricultural subregion within the CVWS. Shown in Fig. 1, both share a common natural watershed outlet, the confluence of the San Joaquin and Sacramento Rivers, which drains to San Francisco Bay. The CV (Fig. 1, green outline), located approximately parallel to the California coastline, receives surface water from the Sacramento River Basin, San Joaquin River Basin, and Tulare Basin, which collectively make up the CVWS (Fig. 1, red outline). The climate of this region is Mediterranean in the north and semiarid in the south, with most precipitation occurring in the winter (November through March), out of phase with evaporative demand which is high in the summer (July through September; Cooper et al., 2018). While the CV is within the CVWS, the two regions have a distinct hydrography. The snow-dominated Sierra Nevada mountains within the CVWS are the primary source of the water that eventually travels to the CV through rivers, runoff, and groundwater. The CV, as one of the most important agricultural regions in the world,

is heavily irrigated by a vast network of reservoirs, canals, and aqueducts. The northern portion of the CV is characterized by a shallower semiconfined aquifer, while the southern portion of the CV contains a two-layer system with an unconfined aquifer overlying a deeper aquifer, confined by the Corcoran Clay lithologic unit (Faunt et al., 2009).

## 3. Methods and data

In this study, we applied the water balance method in the Central Valley (CV) and Central Valley Watershed (CVWS) to test the ability of remote sensing data to estimate $\Delta S_{gw}$ at monthly time steps from 2002 to 2020. Our methodology included three major steps. First, we compared the magnitude, seasonality and uncertainty of each hydrologic variable derived from multiple sources (i.e., ground-based measurements, remotely sensed data, and land surface models). Next, we estimated $\Delta S_{gw}$ using different combinations of hydrologic variables obtained from multiple sources, and assessed the net change, trends, and seasonality in $\Delta S_{gw}$ arising from different combinations. Lastly, we evaluated the performance of $\Delta S_{gw}$ estimates through comparison with three independent estimates from wells, a groundwater model, and GRACE. In the following sections, we describe the water balance method, data sources, and the methods used in each of the three major steps listed above.

### 3.1. Water balance method for groundwater storage changes ($\Delta S_{gw}$)

We applied the hydrologic water balance method by solving the following equation for $\Delta S_{gw}$ in the CV:

$$\Delta S_{gw}^{CV} = P - ET - R + Q_s - \Delta S_R - \Delta S_{sm}, \qquad (2)$$

where $R$ is net surface runoff, $Q_s$ is net surface water flux, including imported/exported water and the other variables are as defined in Eq. (1). We did not include changes in the volume of water stored in lakes and rivers because they constitute a relatively small part of the overall mass of water in this region. Net groundwater flow ($Q_{gw\ net}$) into the CV is relatively small (Brush et al., 2013) hence it is assumed to be negligible.

The above equation was modified to estimate $\Delta S_{gw}$ for the CVWS. Removing the runoff ($R$) term, because the CVWS is a closed watershed, and adding the change in Snow-Water-Equivalent ($\Delta S_{SWE}$) term, due to the large amounts of water stored as snow in the Sierra Nevada, yields the following:

$$\Delta S_{gw}^{CVWS} = P - ET - Q_s - \Delta S_R - \Delta S_{sm} - \Delta S_{SWE} \qquad (3)$$

We evaluated the $\Delta S_{gw}$ estimates at two spatial extents (CV and CVWS) for three main reasons: (1) to better understand the behavior of $\Delta S_{gw}$ at varying spatial scales (CV vs CVWS), (2) to evaluate the performance of the calculated $\Delta S_{gw}$ in regions with and without snow cover, and (3) the availability of validation datasets for both areas. Estimates of $\Delta S_{gw}$ from a numerical groundwater model (Brush et al., 2013) and well-based water level measurements (Alam et al., 2021) were used to evaluate the performance of water balance estimates for the CV, while GRACE data were used to evaluate performance of water balance estimates for the CVWS.

### 3.2. Data

Table 1 shows the data sources used in this study which can be grouped into the following categories: (1) ground-based measurements which include those from groundwater wells, reservoirs, stream gauges, and interpolated measurements made with rain gauges; (2) remote sensing measurements; and (3) land surface models (LSMs). The categories are denoted in the second column of Table 1. Remote sensing measurements and land surface model outputs describing the terms
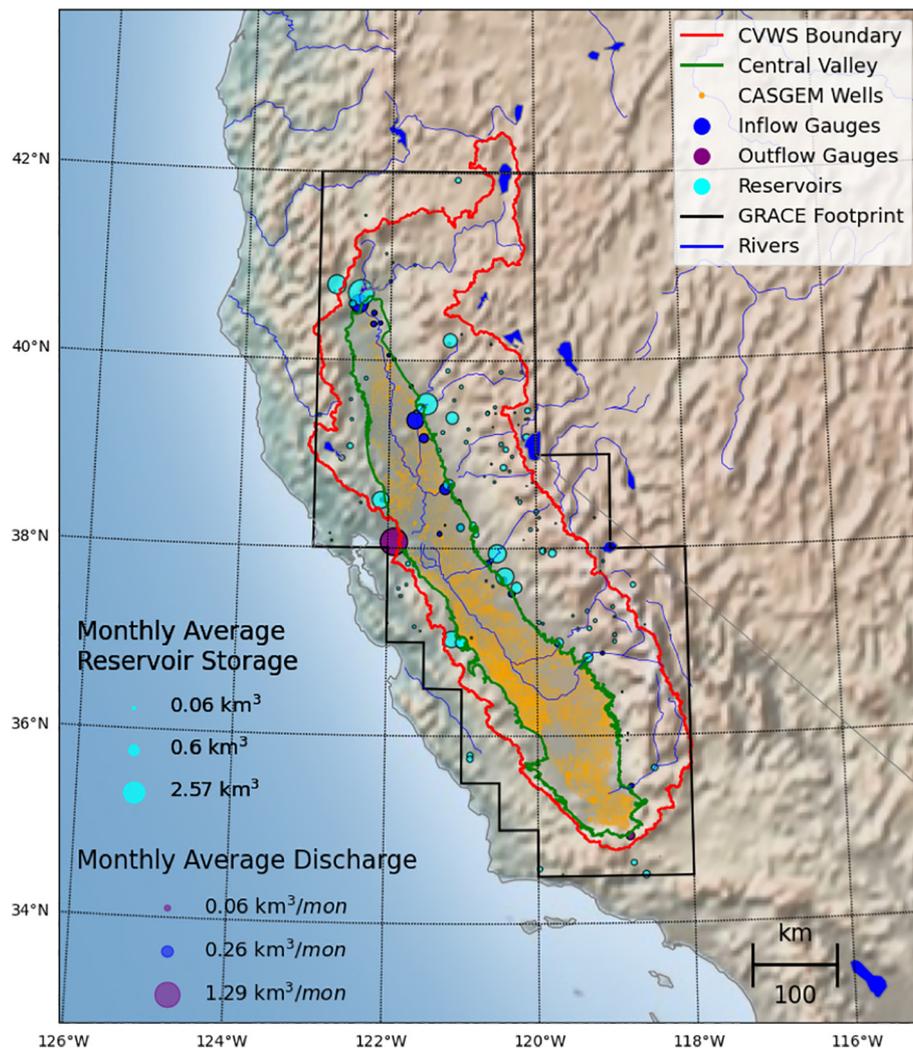
**Fig. 1.** Study Area. CVWS (red polygon), CV (green polygon), DWR CASGEM well locations (orange circles), USGS gauge stations measuring inflows ($Q_{in}$; blue circles) and outflows ($Q_{out}$; purple circles) sized by monthly average ($Q_s$), California Data Exchange Center (CDEC) reservoir locations (cyan circles) sized by mean monthly storage ($S_R$), and the GRACE satellite footprint from Argus et al., 2017 (black polygon).

in Eqs. (2) and (3) were accessed primarily through the Google Earth Engine cloud computing platform (Gorelick et al., 2017), or other public repositories (references given in Table 1), and aggregated to monthly time steps. Detailed descriptions of the source for each variable are provided in the following sections.

### 3.2.1. Precipitation (P)

Precipitation data were obtained from products utilizing remote sensing and ground-based measurements made with rain gauges. Products based on interpolation of rain gauge measurements included the Precipitation Regression on Independent Slopes Model (PRISM) (Daly et al., 2015) and the Daymet meteorological model (Thornton et al., 2018). Products based on the remote sensing measurements included the Global Precipitation Measurement Mission (GPM) (Huffman et al., 2019). We also used the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) (Ashouri et al., 2015) model, and the Climate Hazards Group InfraRed Precipitation with Station Data (CHIRPS) (Funk et al., 2015), which combines both ground-based and remote sensing measurements.

### 3.2.2. Actual Evapotranspiration (ET)

Evapotranspiration is challenging to reliably estimate (Allen et al., 2011) through ground-based measurements (Foken, 2008), remote sensing data (Liou and Kar, 2014), and land surface models (Kumar

et al., 2018). While Potential Evapotranspiration (PET) is the maximum possible evaporation given current meteorological conditions at the land-atmosphere interface, Actual Evapotranspiration (AET) is the amount of water which physically moves from the land surface to the atmosphere - the appropriate ET flux for the water balance method. It can be directly measured with eddy-covariance flux towers that provide ground-based measurements of AET, but spatial coverage is very limited. As a result, both land surface models and remote sensing measurements have emerged as attractive upscaling approaches. Due to these complicating factors, we estimated ET using LSMs as well as remote sensing data.

Land surface models estimate AET by simulating mass and energy transfer at the land surface using weather and radiation data as driving variables. They are typically calibrated through data assimilation, a procedure in which ground-based measurements are used to update parameters of a model. In this study we used GLDAS, FLDAS, and Terraclimate, which have global coverage and long data records, to estimate AET for both the CV and the CVWS, noting that these products are not attuned to accurately simulate irrigated or engineered regions.

In terms of remote sensing measurements, NASA's MOD16 AET product (Mu et al., 2013), and the Operational Simplified Surface Energy Balance (SSEBop) AET product (Senay et al., 2013) have shown good agreement when validated against flux tower measurements and have global coverage at 1 km resolution. We used MOD16 and SSEBop as

**Table 1**
Datasets considered in this study, including ground-based measurements (GBM), remote sensing products (RS), and land surface models (LSM).

**Precipitation (P)**

| Source | Data type | Spatial resolution | Availability | Spatial coverage | Reference |
|---|---|---|---|---|---|
| PRISM AN81 | GBM | 800 m or 4 km | 1895–present | United States | Daly et al., 2008 |
| DAYMET v4 | GBM | 1 km | 1980–present | North America | Thornton et al., 2018 |
| GPM v6* | RS | 10 km | 2001–present | Global | Huffman et al., 2019 |
| CHIRPS | RS and GBM | 5 km | 1981–present | Quasi-Global (50° N–50° S) | Funk et al., 2015 |
| PERSIANN-CDR | RS and GBM | 25 km | 1983–2018 | Quasi-Global (60° N–60° S) | Ashouri et al., 2015 |

**Actual Evapotranspiration (AET)**

| Source | Data type | Spatial resolution | Availability | Spatial coverage | Reference |
|---|---|---|---|---|---|
| MODIS (MOD16)* | RS | 1 km | 2001–present | Global | Mu et al., 2011 |
| SSEBop (MODIS)* | RS | 1 km | 2002–2020 | United States | Senay et al., 2013 |
| Terraclimate | LSM | 4 km | 1958–present | Global | Abatzoglou et al., 2018 |
| FLDAS v1 | LSM | 10 km | 1980–present | Global | McNally et al., 2017 |
| GLDAS v2.1 | LSM | 25 km | 1980–present | Global | Rodell et al., 2004 |

**Potential Evapotranspiration (PET)**

| Source | Data type | Spatial resolution | Availability | Spatial coverage | Reference |
|---|---|---|---|---|---|
| MODIS (MOD16)* | RS | 1 km | 2001–present | Global | Mu et al., 2011 |

**Runoff (R)**

| Source | Data type | Spatial resolution | Availability | Spatial coverage | Reference |
|---|---|---|---|---|---|
| Terraclimate* | LSM | 4 km | 1958–present | Global | Abatzoglou et al., 2018 |
| GLDAS v2.1* | LSM | 25 km | 1980–present | Global | Rodell et al., 2004 |
| FLDAS v1* | LSM | 10 km | 1980–present | Global | McNally et al., 2017 |

**Discharge (Q)**

| Source | Data type | Spatial/temporal resolution | Availability | Spatial coverage | Reference |
|---|---|---|---|---|---|
| DWR Dayflow* | GBM | Daily | 1997–2020 | California | DWR, 2021 |
| USGS Stream Gauges* | GBM | Varies, typically daily or 15 min | Varies | United States | U.S. Geological Survey (USGS), 2021 |

**Snow-Water-Equivalent (SWE)**

| Source | Data type | Spatial resolution | Availability | Spatial coverage | Reference |
|---|---|---|---|---|---|
| DAYMET | GBM | 1 km | 1980–present | North America | Thornton et al., 2018 |
| SNODAS* | RS | 1 km | 2003–present | United States | NOHRC SNODAS, 2004 |
| UCB LRM* | RS | 500 m | 2000–2019 | United States | Schneider and Molotch, 2016 |
| GLDAS v2.1 | LSM | 25 km | 1980–present | Global | Rodell et al., 2004 |
| FLDAS v1 | LSM | 10 km | 1980–present | Global | McNally et al., 2017 |
| Terraclimate | LSM | 4 km | 1958–2020 | Global | Abatzoglou et al., 2018 |

**Soil moisture (SM)**

| Source | Data type | Spatial resolution/number of layers/maximum depth | Availability | Spatial coverage | Reference |
|---|---|---|---|---|---|
| SMAP* | RS | 25 km/–/15–30 cm | 2015–present | Global | Entekhabi et al., 2010 |
| SMOS* | RS | 25 km/–/21 cm | 2010–present | Global | Bolten et al., 2010 |
| GLDAS v2.1 | LSM | 25 km/4/2 m | 1980–present | Global | Rodell et al., 2004 |
| Terraclimate* | LSM | 4 km/1/6 m | 1958–present | Global | Abatzoglou et al., 2018 |

**Reservoir storage ($S_R$)**

| Source | Data type | Temporal resolution | Availability | Spatial coverage | Reference |
|---|---|---|---|---|---|
| California Data Exchange Center (CDEC) Monthly Reservoir Storage* | GBM | Monthly | Varies | California | California Data Exchange Center (CDEC), 2021 |

**Landcover and crop cover**

| Source | Data type | Temporal resolution | Availability | Spatial coverage | Reference |
|---|---|---|---|---|---|
| USGS National Landcover Database (NLCD)* | RS | Varies; typically once per 3 years | 1992–present | United States | Yang et al., 2018 |
| USDA Cropland Data Layer (CDL)* | RS | Annual | 2008–present | United States | USDA, 2021 |

Those chosen for the Preferred Remote Sensing (PRS) scheme, described in Section 3.5, are denoted with a *. For more information on product acronyms, description of remote sensing retrieval methods, model outputs, forcing data, resolution, availability, see the reference for each dataset.

two of the data sources to obtain *ET* estimates for the CV and CVWS. However, both MOD16 and SSEBop tend to underestimate *ET* in irrigated croplands (Velpuri et al., 2013; Mu et al., 2013).

Much of the Central Valley Watershed is heavily irrigated agriculture, for which *ET* estimation is challenging for LSMs as well as traditional remote sensing approaches applicable for natural vegetation.

We therefore implemented a third approach to estimate *ET* which uses the product of potential evapotranspiration (*PET*) and crop coefficients (Allen et al., 1998) to estimate *ET* in irrigated regions. Crop coefficients are dimensionless scaling factors intended to approximate the growth stage of a plant. Crop coefficient data are available through the US Geological Survey (Schmid et al., 2004; Faunt et al., 2009). Crop

evapotranspiration calculated using this method is suitable for regions characterized by irrigated agriculture (Allen et al., 1998), like the CV. To apply the method, we identified agricultural and non-agricultural areas in CV using the National Land Cover Database (NLCD; Yang et al., 2018), a Landsat-based 30 m resolution remotely sensed product produced every 2–3 years for the United States Department of Agriculture. We considered the "Cropland", "Pasture", and "Open Water" classes within the Central Valley as irrigated, and all other areas as non-irrigated agriculture or natural vegetation. To estimate *AET* in non-irrigated regions, we simply used the MOD16 *AET* product. To estimate *AET* in irrigated regions, we obtained the *PET* from the MOD16 *PET* product, scaled by a dimensionless crop coefficient (*kc*) determined to be appropriate for the specific crop or land cover type (Allen et al., 1998). Crop types and land cover for each year were obtained from the Cropland Data Layer (CDL; USDA, 2021), and were assumed static between 2001 and 2008, for which there are no available data. Crop and land cover categories from the CDL were mapped to classes defined in Schmid et al. (2004) and used in Faunt et al. (2009), to obtain monthly *kc* coefficients used to scale MOD16 *PET* values at each irrigated pixel. The total monthly *ET* (MODkc) was subsequently obtained by summing *ET* in irrigated and non-irrigated regions.

We note that in the MODkc approach, we selected the MOD16 *PET* product over regional *PET* datasets, such as the Spatial California Irrigation Management Information System dataset (Hart et al., 2009), because it is globally available at a spatial resolution of 1 km. In addition, it uses the Penman-Monteith approach (Monteith, 1965), consistent with other *PET* datasets, and is calculated from meteorological fields from the Modern-Era Retrospective analysis for Research and Applications (MERRA) reanalysis dataset (Gelaro et al., 2017).

### 3.2.3. Discharge (Q) and runoff (R)

Discharge, the inflows and outflows to/from the study areas from rivers and streams, can only be estimated, at present, using ground-based measurements. Discharge was quantified using streamflow data from the USGS (2021) and DWR (2021) (gauge stations shown in Fig. 1). In total, we identified 41 gauges that measure streams flowing into the CV with data covering the study period; there are no inflows for the CVWS, by definition of the term watershed. For both the Central Valley Watershed and Central Valley, the sole natural surface water outflow is from the Sacramento-San Joaquin Delta, formed by the confluence of the Sacramento and San Joaquin Rivers, which drains into the San Francisco Bay estuary. Gauge stations record the discharge from the delta to the bay; this is available as daily outflow estimates, released annually through the Dayflow program (DWR, 2021). Water conveyance structures (e.g., the California Aqueduct, part of the State Water Project; Fig. 1 purple circle) transport water from the Central Valley Watershed to Southern California. This particular source of delivered water is gauged and included with USGS streamflow estimates.

Discharge alone, however, does not capture all the surface water entering the CV because some smaller watersheds cross the Central Valley boundary (Faunt et al., 2009) and thus contribute diffuse surface water runoff as opposed to streamflow at a single point. To account for this, we considered runoff estimates available for the CV from GLDAS, FLDAS, and Terraclimate LSMs. Previous studies (Nady and Larragueta, 1983; Faunt et al., 2009) found that runoff accounts for a small percentage of inflow relative to *Q*.

### 3.2.4. Changes in reservoir storage ($\Delta S_R$)

Fluctuations in reservoir levels constitute a large portion of the changes in surface water stored in the Central Valley Watershed. Typically, water depth in a reservoir is recorded on a staff gauge or data logger and converted to volumes using a depth-volume curve which depends on reservoir-specific bathymetry. Monthly reservoir storage data were downloaded from the California Data Exchange Center (CDEC, 2021) for all reservoirs in California. While availability varies by reservoir, many data records span decades and are available through the present,

especially for large capacity reservoirs, which have a greater influence on the overall water balance. CDEC (2021) contains data for 10 reservoirs within the CV and 93 within the CVWS from 2002–present. The total volumetric storage for all reservoirs within the CV or CVWS was summed to provide monthly estimates for both the total volumes across the CV and CVWS. The absolute volumes were differenced relative to the starting value at the beginning of the study period (October 1st 2002) to determine storage changes which are inputs to the water balance method. Variations in water stored in rivers, lakes, aqueducts, and streams constitute a negligible portion of the overall surface storage at the scale of the Central Valley Watershed and Central Valley (Xiao et al., 2017); hence, they were assumed to be negligible in this study.

### 3.2.5. Changes in soil moisture storage ($\Delta S_{SM}$)

Limited ground-based measurements are available for soil moisture; the Soil Climate Analysis Network, which maintains and operates a spatially distributed in-situ soil moisture sensor network, has very sparse coverage in California. Thus, we relied on remote sensing and LSM-based datasets to obtain estimates of changes in soil moisture. Passive microwave satellite data are available for 2010 to present from the Soil Moisture Ocean Salinity satellite (SMOS; Bolten et al., 2010) and for 2015 to present from the Soil Moisture Active-Passive satellite (SMAP; Bolten et al., 2010; Entekhabi et al., 2010), while land surface models including GLDAS (Rodell et al., 2004) and Terraclimate (Abatzoglou et al., 2018) are available to estimate $\Delta S_{sm}$ for the entirety of our study period between 2002 and 2020.

Soil moisture is a state variable within land surface models, and the total soil moisture at any given pixel is the sum of water stored within each depth interval. GLDAS contains 4 depth intervals: 0–10 cm, 10–40 cm, 40–100 cm, and 100–200 cm; we summed the soil moisture across the layers to determine the total soil moisture for a given pixel at a given time. In contrast, Terraclimate uses spatially distributed single-layer depth intervals, as described in Wang-Erlandsson et al. (2016), which, within the study region, range from 1 m to 6 m in depth.

### 3.2.6. Changes in water stored as snow ($\Delta S_{SWE}$)

Snow pillows and snow courses provide ground-based measurements that can be used to estimate Snow-Water-Equivalent (*SWE*), the amount of water contained within a snowpack. While there are hundreds of snow pillows and courses distributed throughout the Western US, remote mountainous locations complicate sensor dispatchment and maintenance. This results in a lack of spatial coverage and discontinuous data records, which makes it challenging to use ground-based *SWE* measurements for long term water balance studies. Spatial interpolation of ground-based meteorological measurements (e.g. Daymet) can inherit these problems. *SWE* can also be estimated from land surface models (e.g. GLDAS, FLDAS, Terraclimate) but this approach is recognized to be highly uncertain and to generally underestimate *SWE*, especially in the Western US (Wrzesien et al., 2017; Broxton et al., 2016). This underestimation is due to the forcing data and the way in which snow ablation is accounted for within land surface models (Wrzesien et al., 2017). Noting these limitations, we used *SWE* data from Daymet, GLDAS, FLDAS, and Terraclimate in this study.

Remote sensing approaches to measure and monitor *SWE* are inexpensive, have broad spatial coverage, and have validated well against measurements made in snow pillows and snow courses. We incorporated remote sensing-based estimates of *SWE* from (1) the Snow Data Assimilation System (SNODAS; NOHRC, 2004), and (2) the Linear Regression Model from University of Colorado – Boulder (UCB LRM; Schneider and Molotch, 2016). SNODAS is an assimilation-based remote sensing dataset which uses satellite measurements of snowcover fraction, radar precipitation, and measurements made in snow pillows and courses in the assimilation algorithm (Wrzesien et al., 2017). The UCB LRM model uses satellite data from the Moderate Resolution Imaging Spectroradiometer (MODIS) and streamflow measurements with a calorimetry approach to estimate *SWE* (Schneider and Molotch, 2016).

### 3.3. Estimate of correlation in water balance components

We used the Pearson Correlation coefficient (Pearson, 1895) to assess correlations between datasets describing each variable in the water balance (Eq. (1), Table 1). The Pearson Correlation coefficient measures the linear correlation between two variables. It is equivalent to the covariance divided by the product of the standard deviations. The Pearson Correlation ($P$) between two variables x and y is given by:

$$P_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{4}$$

where $x_i$ and $y_i$ are the elements of each variable, and $\bar{x}$ and $\bar{y}$ are the mean of each variable.

### 3.4. Estimate of uncertainty in water balance components

Analyzing uncertainty in input hydrologic variables can help identify sources of variability in estimates of $\Delta S_{gw}$ and aid in the selection of remote sensing products that capture realistic seasonal patterns. To estimate uncertainty in the input hydrologic variables, we used Triple Collocation (Stoffelen, 1998), a common error estimation technique applicable when the true value of a variable is not known, but multiple estimates are available. Triple collocation provides statistical estimates of the random error variances ($\sigma^2$) associated with a dataset, assuming additive and multiplicative errors among at least three independent and collocated datasets describing the same variable. Several studies have demonstrated Triple Collocation as an effective technique to estimate errors associated with hydrologic and meteorological measurements (e.g. Scipal et al., 2008; Chakraborty et al., 2013; Alemohammad et al., 2015; Gruber et al., 2016). For three area-averaged collocated estimates of the same variable, the error variances ($\sigma^2$) are given by:

$$\begin{aligned} \sigma_x^2 &= \langle (x - y_s)(x - z_s) \rangle \\ \sigma_y^2 &= \langle (y - x_s)(y - z_s) \rangle \\ \sigma_z^2 &= \langle (z - x_s)(z - y_s) \rangle \end{aligned} \tag{5}$$

where brackets denote the temporal mean, $x$, $y$, and $z$ are the three datasets, and the subscript $s$ denotes scaling of the $x$, $y$, or $z$ dataset to the mean and standard deviation of the reference dataset (Scipal et al., 2008). For variables with more than three datasets, we constructed each unique combination of triplets, estimated $\sigma^2$ as in Eq. (5), and reported the mean $\sigma$ for all triplets containing each specific dataset. For instance, there were five precipitation datasets, so the number of unique triplets is $5!/(5-3)! = 60$. Of the 60 total triplets, each individual dataset appears $(3/5) * 5!/(5-3)! = 36$ times. The random error variance ($\sigma$) for each dataset was then computed from the 36 triplets where it appears. We report the mean and standard deviation of $\sigma$ (in mm) as the uncertainty calculated across the 36 triplets where a dataset appears. A low mean and standard deviation of $\sigma$ for a specific dataset denotes a smaller relative error, while a high mean and standard deviation of $\sigma$ suggests a larger relative error.

### 3.5. Groundwater storage estimation scenarios

Estimates of $\Delta S_{gw}$ can vary depending on the data source used in Eqs. (2) and (3), but it is not well known to what degree input datasets can affect the resulting magnitudes, seasonality, and trends of $\Delta S_{gw}$, and how water balance-based estimates compare with other methods. In order to compare estimates of $\Delta S_{gw}$ derived from different sources of data and provide an insight of the strengths and weaknesses of remote sensing products in estimating $\Delta S_{gw}$, we constructed 4 water balance scenarios for the CV and CVWS which utilize data from ground-based measurements, LSMs and Remote Sensing

datasets to estimate $\Delta S_{gw}$. The description of each scenario is provided below:

(1) All data scenario (ADS): An ensemble of estimates of $\Delta S_{gw}$ was computed using all combinations of inputs shown in Table 1. The number of input datasets was five for $P$, six for $ET$, six for $SWE$, four for $SM$, and three for $R$ which resulted in 360 distinct combinations of $\Delta S_{gw}$ for the CV ($N_P \times N_{ET} \times N_{SM} \times N_R \times N_Q = 5 \times 6 \times 4 \times 3 \times 1 = 360$) and 720 for the CVWS ($N_P \times N_{ET} \times N_{SM} \times N_{SWE} \times N_Q = 5 \times 6 \times 4 \times 6 \times 1 = 720$).

(2) Ground-Based and Land Surface Model scenario (GBLSM): An ensemble of estimates for $\Delta S_{gw}$ was computed using inputs from LSMs for $ET$, $SM$, $SWE$, and $R$ while ground-based estimates are used for $P$, resulting in 36 different combinations for the CV ($N_P \times N_{ET} \times N_{SM} \times N_R \times N_Q = 2 \times 3 \times 2 \times 3 \times 1 = 36$) and 48 for the CVWS ($N_P \times N_{ET} \times N_{SM} \times N_{SWE} \times N_Q = 2 \times 3 \times 2 \times 4 \times 1 = 48$).

(3) All Remote Sensing scenario (ARS): An ensemble of estimates for $\Delta S_{gw}$ was computed using inputs of remotely sensed data for all hydrologic variables when available, those being three for $P$, three for $ET$, two for $SWE$, and three for $SM$. Because remote sensing-based estimates of $SM$ are available only from 2010 onwards, we used Terraclimate from 2002 to 2020, SMOS from 2010 to 2020, and SMAP from 2015 to 2020 for the ARS scenario ensembles. Lastly, $R$ was derived from three LSMs, resulting in 81 different combinations for the CV ($N_P \times N_{ET} \times N_{SM} \times N_r \times N_Q = 3 \times 3 \times 3 \times 3 \times 1 = 81$) and 54 for the CVWS ($N_P \times N_{ET} \times N_{SM} \times N_{SWE} \times N_Q = 3 \times 3 \times 3 \times 2 \times 1 = 54$).

(4) Preferred Remote Sensing (PRS) scenario: In this scenario, we estimated $\Delta S_{gw}$ using a subset of all remote sensing datasets used in the ARS scenario. Since the accuracy of remote sensing data can vary between sources and regions, it remains unclear how the use of different products will affect the accuracy of $\Delta S_{gw}$ estimates. We therefore selected remote sensing data for different hydrologic components that are generally found to exhibit lower error than others in the selected study regions. For precipitation, we used the state-of-the-art and newly released GPM Integrated Multisatellite Retrievals (IMERG) as our remote sensing estimate because of higher accuracies relative to legacy products (Wang et al., 2021). For $AET$, we used SSEBop and MODkc because these remote sensing approaches can more effectively capture the $ET$ patterns in irrigated regions (Senay et al., 2013), which is a key feature of the CV. In contrast, MOD16 is known to significantly underestimate $ET$ in croplands (Velpuri et al., 2013). For $SM$ and $SWE$, we find that remotely sensed data from different sources do not cover the whole study period. Selection of one product would limit our analysis to a short timespan, therefore, we used the temporal average of sources in the ARS scenario; these being Terraclimate, SMOS, and SMAP for $SM$, and SNODAS and UCB LRM for $SWE$; in order to create a continuous time series. For $R$, we used the temporal mean across the three LSMs, noting that the magnitude of $R$ was relatively small compared to other more important fluxes ($Q$, $ET$, $P$). Datasets used in the PRS scenario are denoted * in Table 1.

For the ADS, GBLSM, and ARS scenarios described above, we generated ensemble $\Delta S_{gw}$ estimates using combinations of input datasets. Because we were interested in both the trend and seasonality of $\Delta S_{gw}$ estimates produced from the water balance method, we differenced each time series estimate of $\Delta S_{gw}$ from the mean of the first five years to decrease the variance among estimates while still capturing the prevailing long-term trends. This 5-year mean differencing procedure is consistent with the 5-year calibration period used in GRACE processing (Wiese, 2015). In the ADS, GBLSM and RS Scenarios, we

used mean $\Delta S_{gw}$ across all ensembles as the final estimate to compare with validation datasets.

### 3.6. Evaluation of RS-based $\Delta S_{gw}$ estimates

An effective method of estimating $\Delta S_{gw}$ should be able to capture long-term net changes, accelerated depletion during droughts, recovery during wet years and seasonal fluctuations. To evaluate the performance of $\Delta S_{gw}$ estimates produced from the water balance method, we compared the ADS, GBLSM, ARS, and PRS scenarios with independent estimates of $\Delta S_{gw}$ from GRACE, wells, and the C2VSim groundwater model, detailed below. We evaluated the performance of $\Delta S_{gw}$ estimates for the CV region by comparing with two estimates based on: (1) groundwater levels measured in wells, available from 2002 to 2019 (Alam et al., 2021), and (2) the C2VSim groundwater model (Brush et al., 2013), available from 2002 to 2019. For the CVWS region, we evaluated the performance of $\Delta S_{gw}$ estimates through comparison with $\Delta S_{gw}$ estimates from NASA's Gravity Recovery and Climate Experiment (GRACE; Swenson, 2012), available from 2002 to 2017, and corrected for fluctuations in $\Delta S_{SWE}$, $\Delta S_{SM}$ and $\Delta S_R$. In the following sections, we first discuss the use of GRACE, groundwater well data, and a groundwater model to estimate $\Delta S_{gw}$, and then describe the performance evaluation metrics used in this study.

### 3.6.1. GRACE

GRACE satellites (Swenson, 2012) measure changes in Earth's gravity field and have been previously used to estimate changes in groundwater storage for the CV (e.g., Famiglietti et al., 2011; Scanlon et al., 2012; Argus et al., 2017; Xiao et al., 2017). GRACE data are available from 2002 to 2017, all over the globe, with a spatial resolution of 150,000 km$^2$ (National Aeronautics and Space Administration (NASA), 2021).

In this study, we used the GRACE Mass Concentration Blocks (mascons) which, relative to the traditional spherical harmonic approach, allows for more robust a priori geophysical constraints, thereby reducing noise and leakage errors (Watkins et al., 2015). We applied the Coastline Resolution Improvement (CRI) filter (Wiese et al., 2016), which better separates terrestrial and oceanic contributions to mass changes. Spatially distributed gain factors described in Wiese et al. (2016) were also applied to the gridded GRACE data to further increase the signal to noise ratio.

The GRACE domain, shown in Fig. 1 as a black outline, follows 0.5° gridlines and is consistent with the domain defined in Famiglietti et al. (2011), Wiese et al. (2016), and Argus et al. (2017) as the area over which a valid estimate of $\Delta S_{gw}$ can be obtained. As can be seen, this approximately covers the extent of the CVWS. A second-degree polynomial fit was used to interpolate missing values which occur from 2011 onwards, when battery saving measures caused data gaps approximately every 5–6 months. The use of different fitting functions exhibited very minor changes in the resulting timeseries. Similar to the procedure outlined in Famiglietti et al. (2011), Scanlon et al. (2012), and Argus et al. (2017), GRACE-based changes in terrestrial water storage were corrected for changes in soil moisture and snow-water-equivalent, and changes in reservoir storage for the GRACE study domain. To adequately capture uncertainty, we applied an ensemble correction using all *SM* and *SWE* datasets shown in Table 1 and described in Sections 3.2.5 and 3.2.6, and used the ensemble mean as our best estimate.

### 3.6.2. Groundwater storage changes from wells

We used groundwater level time series obtained from wells to calculate $\Delta S_{gw}$ for the CV for 2002–2019 (Alam et al., 2021). Previous studies (e.g. Scanlon et al., 2012) have also used water level data from groundwater level measurements to estimate $\Delta S_{gw}$ for the CV. Groundwater level data were compiled from the California Department of Water Resources (DWR) California Statewide Groundwater Elevation Monitoring (CASGEM) database (DWR CASGEM, 2021). The database contains measurements from 43,987 wells which are used for monitoring, irrigation, domestic uses, and other purposes. Of these, 23,014 groundwater wells are located within CV (Fig. 1).

Estimates of $\Delta S_{gw}$ were computed by multiplying groundwater level changes with the area of the CV and a spatially dependent storage coefficient, which describes the volume of water released from an aquifer per unit decrease in water level. Noting that both groundwater level changes and the storage coefficient are uncertain and sparsely observed in space and time, we followed three steps to estimate $\Delta S_{gw}$ from well data: First, groundwater levels measured in wells were spatially interpolated using Inverse Distance Weighting at a 10 km cell size. Next, month-to-month changes were calculated for each grid cell from water level data. Lastly, month-to-month water levels were multiplied by storage coefficient values determined from statistically weighting Specific Storage and Specific Yield values from the Central Valley Hydrologic Model (CVHM; Faunt et al., 2009) and C2VSim, based on the magnitude of water level fluctuation at a given grid cell. The full statistical weighting procedure and complete details of the method are described in Alam et al. (2021).

### 3.6.3. California Central Valley groundwater-surface water simulation model (C2VSim)

The California Central Valley Groundwater-Surface Water Simulation Model (C2VSim) (Brush et al., 2013) is a finite element numerical groundwater flow model of the CV developed by the California DWR that simulates water flux through the land surface and root zone, and into the groundwater system. The model calculates land surface, root zone, and groundwater hydrologic budgets at monthly time steps from 1921 to 2009. Grouped into 21 subregions, the C2VSim fine grid version (C2VSimFG) contains over 35,000 elements and covers the 1975–2015 hydrologic years with a spatial resolution that varies, but is typically ~1.5 km. The model is available at https://data.cnra.ca.gov/dataset/c2vsimfg-version-1-0 (last accessed: March 2021). We used Version 1.0, the current version of the C2VSim fine grid model (Integrated Water Flow Model-2015 version) in this study, herein simply called C2VSim (see Alam et al., 2021; S3 and S4 for more detail). Estimates of $\Delta S_{gw}$ in the CV from this model are similar to those obtained from the CVHM (Faunt et al., 2009; Faunt and Sneed, 2015) over the 1975–2003 period where both models overlap.

The C2VSim simulation was extended from 2015 until 2019 by running the model with three inputs for the extended period: precipitation, surface water inflow to the CV and surface water diversions (Alam et al., 2021). Monthly precipitation was estimated from PRISM (Daly et al., 2015), the same sources as for the base model (DWR, 2021). Inflow time series for the extended period were created using observations and simulated outputs from the VIC model (Alam et al., 2021). The third input, surface water deliveries, included agricultural, industrial, and residential deliveries, and was computed following the method of Hanson et al. (2012), who estimate a percentage of delivery for a given year depending if it is dry or wet, based on precipitation anomaly. The full procedure used to generate inputs required to extend the C2VSim time series is described in greater detail in Alam et al. (2021).

### 3.6.4. Metric for evaluating the performance of RS-based $\Delta S_{gw}$

While the appropriate model or method to estimate changes in groundwater storage depends largely on the area of investigation, degree of spatiotemporal granularity required, and current and historic instrumentation within the area, a reasonable estimate of $\Delta S_{gw}$ must capture the seasonality, long term change, and be able to reproduce the dynamics during extreme events (e.g., drought). As such, intercomparison and benchmarking between independent estimates derived from a variety of datasets and models can facilitate an understanding of which method(s) can best capture the dynamics of hydrologic systems, especially across spatial scales. However, it is also important to note that each of the validation datasets has limitations (described in

Sections 3.6.1–3.6.3). Here, we used multiple $\Delta S_{gw}$ estimates from different sources to enhance the confidence in our evaluation. We found that earlier studies have successfully used similar datasets to understand $\Delta S_{gw}$ behavior in CV and CVWS (Faunt et al., 2009; Famiglietti et al., 2011; Scanlon et al., 2012; Brush et al., 2013; Xiao et al., 2017; Argus et al., 2017; Hanak et al., 2019; Yin et al., 2021).

In this study, we considered three metrics to evaluate the performance of $\Delta S_{gw}$ estimates computed from the water balance method. These were (1) the net change in $\Delta S_{gw}$ for three time windows (droughts: October 2006–September 2009. October 2011–September 2015; long term: 2002–2020) during the 2002–2020 study period for which independent estimates are available, (2) the rate of change in $\Delta S_{gw}$ (km$^3$/year) during recent droughts (2006–2009 and 2011–2015) determined using least squares, and (3) the seasonal amplitude of $\Delta S_{gw}$ estimates derived from the ADS, GBLSM, ARS, and PRS scenarios evaluated against validation datasets (GRACE, well data, C2VSim).

## 4. Results and discussion

Estimates of $\Delta S_{gw}$ were produced for CV and CVWS and the quality of remote sensing data was investigated through analysis of multiple input scenarios. First, correlations and uncertainty among input datasets (Table 1) were determined for each hydrologic variable. Next, ensemble

estimates of $\Delta S_{gw}$ were calculated for the following scenarios: ADS, LSM, ARS, and PRS. Then, the performance of $\Delta S_{gw}$ estimates was evaluated through comparison with three independent estimates from wells, groundwater models, and GRACE, as well as through comparison with previous studies.

### 4.1. Uncertainty in input hydrologic variables

Estimates of (1) precipitation, (2) evapotranspiration, (3) streamflow (stream discharge and runoff), (4) soil moisture, and (5) snow-water-equivalent were compared for a variety of data sources encompassing ground-based measurements, land surface models, and remotely sensed data. In the following section, we discuss the seasonality, variability, magnitude, and uncertainty among datasets available to estimate each variable.

#### 4.1.1. Precipitation (P)

All estimates of precipitation, when aggregated to monthly sums, produced similar estimates of total precipitation for both the CV (Figs. 2 and 3) and CVWS (Figs. 4 and 5) for the length of the study period. The mean and variability in precipitation are consistent among sources, and the lowest error was found in ground-based products PRISM and Daymet. Remote sensing-based estimates of $P$ derived
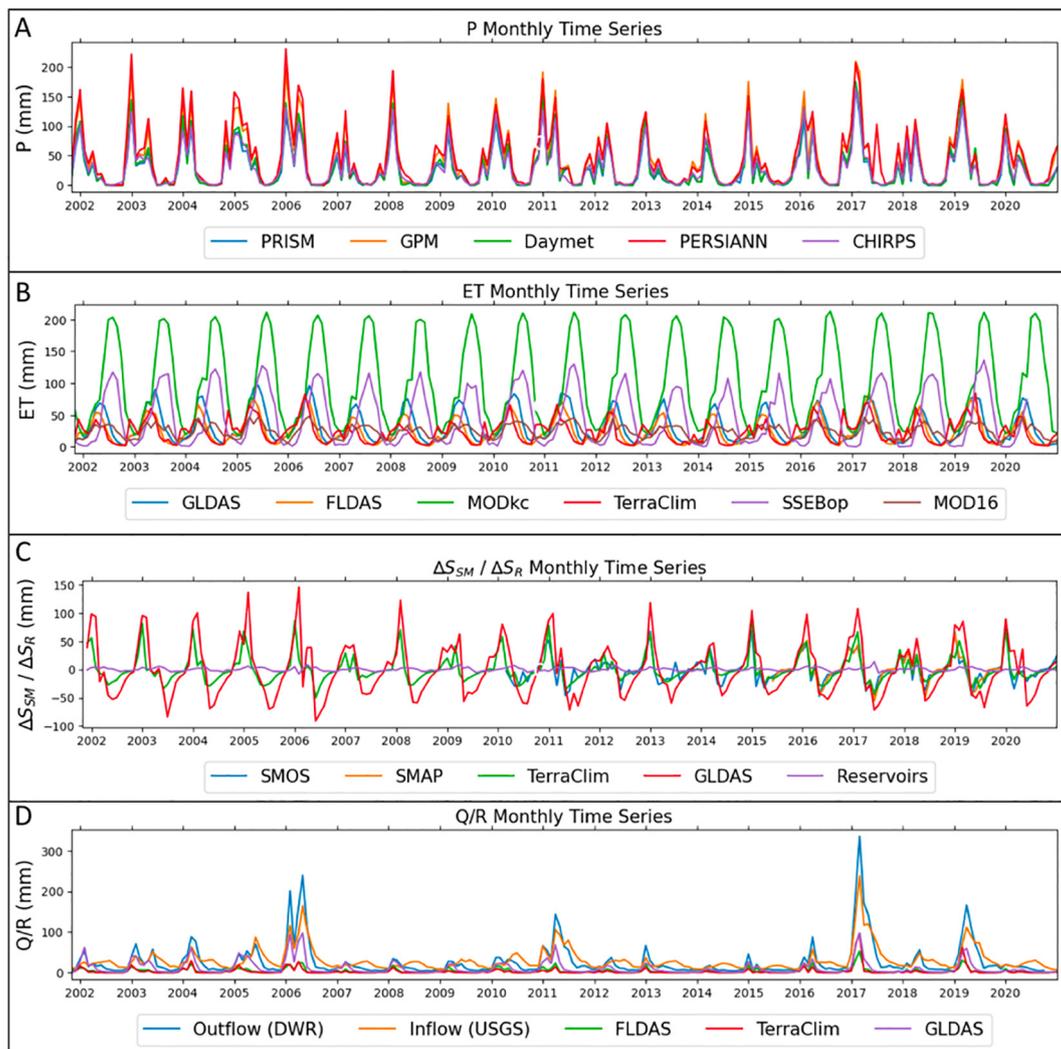


**Fig. 2.** Time series plots for datasets used in the Central Valley water balance model. The rows depict the monthly time series in millimeters from 2002 to 2020 for (A) $P$, (B) $ET$, (C) $\Delta S_{SM}$ and $\Delta S_R$, and (D) $Q$ and $R$.
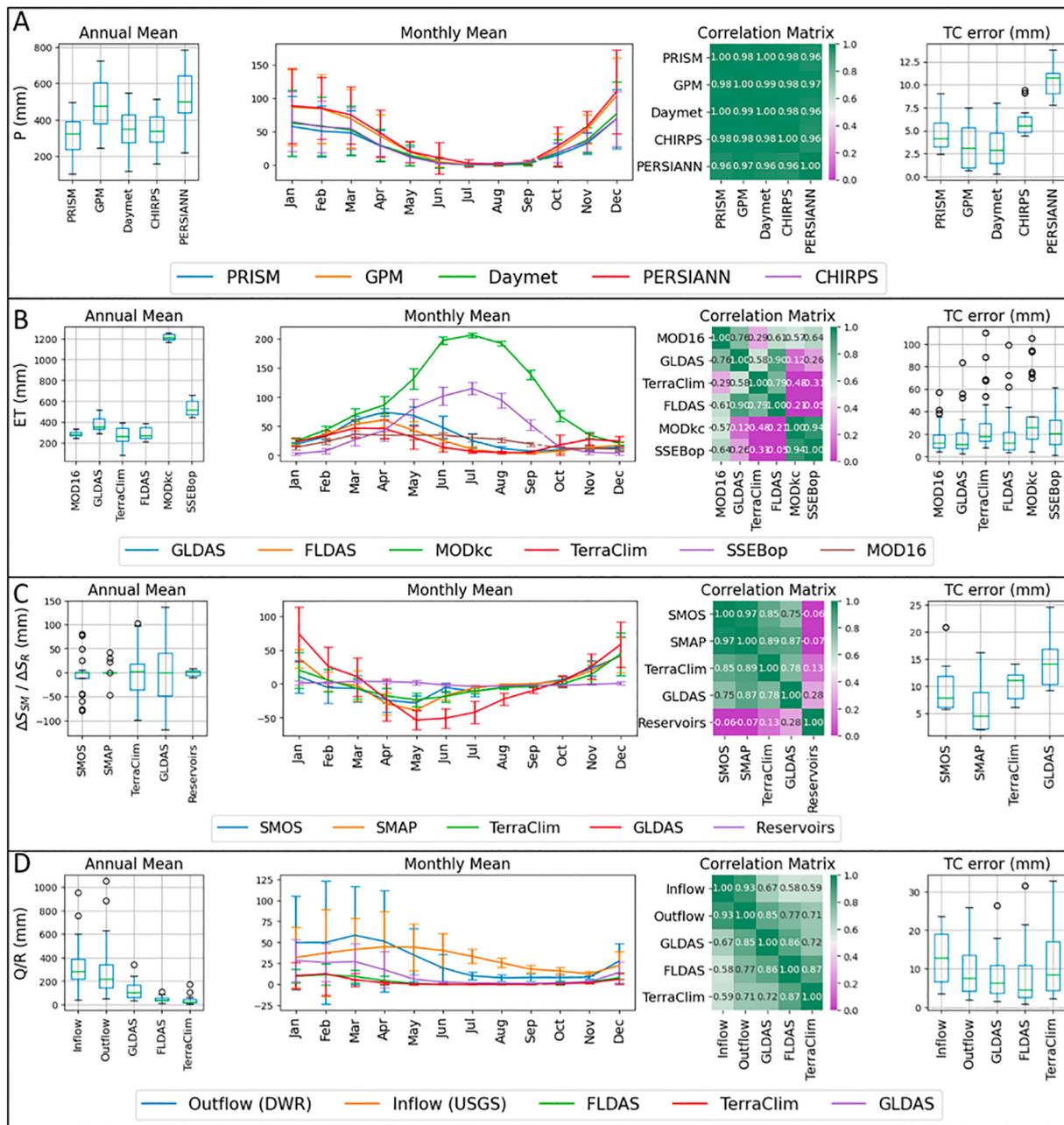
**Fig. 3.** Plots summarizing datasets describing variables used in the Central Valley water balance model. From top to bottom, the panels depict (A) *P*, (B) *ET*, (3) $\Delta S_{SM}/\Delta S_R$, and (4) *Q/R*. The columns, from left to right, depict (1) box and whisker plots showing annual means, standard deviations, and outliers, (2) monthly means, with error bars corresponding to standard deviations, (3) Pearson Correlation matrix describing correlations between datasets, (4) box and whisker plots showing Triple Collocation errors (mm).

from GPM had similar mean triple collocation errors but larger standard deviations compared to those for the ground-based estimates, shown in Figs. 3 and 5. Ground-based estimates of maximum *P* during the very wet 2017 winter were greater than remotely sensed estimates during this time. One anomalous result that contrasted with others is the high precipitation estimate for the PERSIANN dataset during summer 2017, which resulted in a larger triple collocation error (10 mm) relative to other estimates; we interpret this outlier as a regional error associated with the PERSIANN data. The overall agreement among datasets and low triple collocation errors for remotely sensed estimates gave us confidence that the adoption of the remote sensing measurements of *P* in the water balance would not sacrifice accuracy. The strong agreement among precipitation datasets found in this regional study is in contrast with global water balance studies performed on a number of

watersheds (Sahoo et al., 2011; Pan et al., 2012) who found *P* to be the most uncertain flux.

### 4.1.2. Evapotranspiration (ET)

A comparison of *ET* estimates for Land Surface models and remote sensing data is shown in Figs. 2 and 3 (CV) and Figs. 4 and 5 (CVWS). As expected, the MODkc approach we used in this study shows considerably higher monthly and annual *ET* than other estimates for the CV. Estimates from MOD16 and land surface models for the CV show *ET* peaking in March, whereas SSEBop and the MODkc capture the mid-summer peak resulting from irrigated agriculture. Terraclimate consistently shows a slight increase in *ET* for October and November, which is likely due to the inability of LSM-based *ET* approaches to adequately capture irrigated conditions. The magnitude of the increases is still
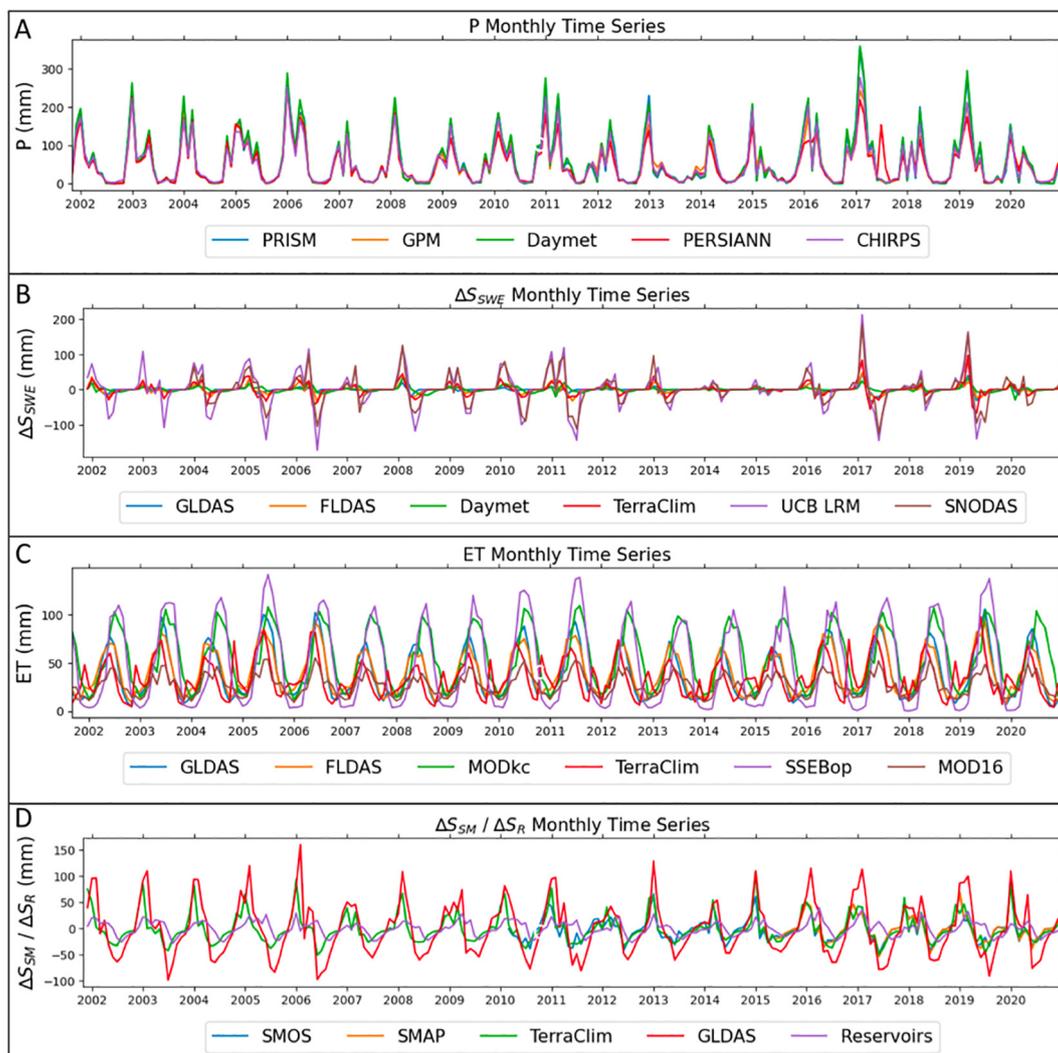
**Fig. 4.** Time series plots for datasets used in the Central Valley Watershed water balance model. The rows depict the monthly time series in millimeters from 2002 to 2020 for (A) $P$, (B) $\Delta S_{SWE}$, (C) $ET$, and (D) $\Delta S_{SM}/\Delta S_R$.

significantly less than the midsummer peak for SSEBop and MODkc. Triple collocation analysis for $ET$ in the CV suggests comparable mean errors between all sources of data - approximately 15–30 mm - with the highest mean error being attributed to the MODkc method. This may be due to the overall larger magnitude of MODkc estimates, the timing and magnitude of which contrast significantly with LSM-based estimates and remote sensing estimates from MOD16.

The results suggest that evapotranspiration is the most uncertain flux term in the balance because it had both the lowest overall correlation and the highest triple collocation error among datasets (Figs. 3, 5). This finding contrasts with water balance studies of other regions (e.g. Sahoo et al., 2011; Pan et al., 2012), which cited precipitation as the most uncertain term. The uncertainty and disagreement among $ET$ datasets likely arises because the CV is one of the most irrigated regions in the world, and the overwhelming influence of irrigation in the CV confounds conventional methods of estimating $ET$ (e.g. LSMs, MOD16), which are known to generally underestimate $ET$ for irrigated regions (Velpuri et al., 2013; Mu et al., 2013).

Many opportunities exist to improve $ET$ estimates, particularly in irrigated regions. The ability to accurately identify irrigated areas (e.g. Deines et al., 2017) at sub-annual timescales would improve $ET$ derived from crop coefficient-based approaches like the MODkc estimate produced here. Satellite-based estimates of crop coefficients (e.g. Mendiguren et al., 2017) can also be used to improve $ET$ estimates. Future research could evaluate the accuracy of satellite-based $kc$ estimates

in the CV, and incorporate these data to improve $ET$ estimates used in the $\Delta S_{gw}$ calculation. In-situ measurements obtained at flux towers are critical in order to assess the quality of remotely sensed $ET$ estimates and improve the estimation of regional accuracies and biases. Finally, the development of new tools, such as the OpenET platform (http:// etdata.org), can facilitate the distribution and comparison of numerous high-resolution $ET$ datasets over large areas. Available models within OpenET include the Mapping Evapotranspiration with Internalized Calibration (METRIC) model (Allen et al., 2007), Satellite Irrigation Management Support (SIMS) model (Melton et al., 2012; Pereira et al., 2020), Atmosphere Land Exchange Inverse (ALEXI) model (Anderson et al., 1997; Anderson et al., 2007a,b), and Priestly-Taylor JPL model (Fisher et al., 2008). These data do not require crop maps, are able to better estimate $ET$ in irrigated regions, and improve upon the 1 km resolution $ET$ estimates from MOD16 and SSEBop used in this study, as well as ET estimates derived from LSMs.

### 4.1.3. Changes in soil moisture ($\Delta S_{SM}$)

Estimates of $\Delta S_{SM}$ from remote sensing measurements and land surface models are shown for the CV in Figs. 2 and 3 and for the CVWS in Figs. 4 and 5. Variability among estimates of $\Delta S_{SM}$ is higher than for $P$, but lower than $ET$. Seasonality is consistent among datasets, but the magnitude of $\Delta S_{SM}$ tends to be greater in LSMs than microwave-based estimates. This may be due to the lack of inclusion of an explicit groundwater component to LSMs – the high magnitude soil moisture
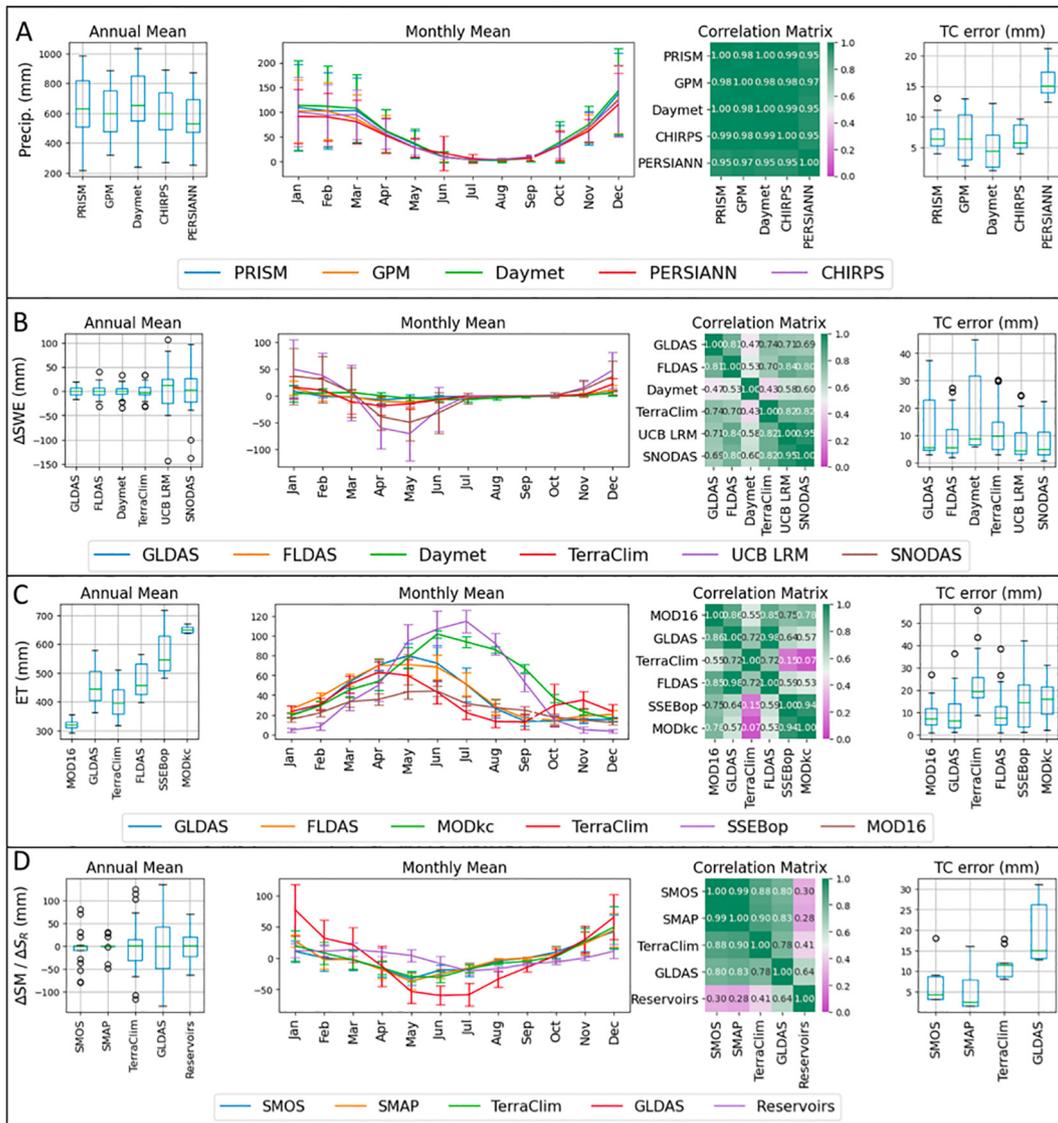
**Fig. 5.** Plots summarizing datasets describing variables used in the Central Valley Watershed water balance model. From top to bottom, the panels depict (A) $P$, (B) $\Delta S_{SWE}$, (3) $ET$, and (4) $\Delta S_{SM}/\Delta S_R$. The columns, from left to right, depict (1) box and whisker plots showing annual means, standard deviations, and outliers, (2) monthly means, with error bars corresponding to standard deviations, (3) Pearson Correlation matrix describing correlations between datasets, (4) box and whisker plots showing Triple Collocation errors (mm).

fluctuations may be capturing partial fluctuations in the groundwater system. While GLDAS and FLDAS consistently estimate larger magnitude fluctuations than passive microwave-based products derived from SMAP and SMOS, the Terraclimate Land Surface Model exhibits a high correlation (0.85–0.9) with both SMOS and SMAP, as shown for the CV in Fig. 3 and the CVWS in Fig. 5. This agreement may be due to the spatially distributed single-layer depth intervals, as described in Wang-Erlandsson et al. (2016). The high level of agreement of Terraclimate with reliable passive microwave-based products gave us confidence of the temporal extrapolation performed in the PRS Scenario, where we used Terraclimate to estimate $\Delta S_{SM}$ for the time period 2002 to 2010, which predates availability of passive microwave remote sensing products.

*4.1.4. Discharge (Q) and runoff (R)*

Discharge and runoff data, shown in Figs. 2 and 3, agree in terms of seasonality, and have relatively consistent magnitudes across data sources. The outflow to the San Francisco Bay, the outlet of both the CV and CVWS, peaks around March, while inflow to the CV peaks around May. Discharge and runoff estimates show a relatively high agreement between datasets as suggested by Pearson Correlation coefficients between 0.6 and 0.9; higher than correlations for other variables. Triple collocation performed on discharge and runoff datasets resulted in errors of 10–20 mm, with Terraclimate having the highest uncertainty among LSM-based approaches.

*4.1.5. Changes in Snow-Water-Equivalent ($\Delta S_{SWE}$)*

In Figs. 4 and 5, we compare estimates of $\Delta S_{SWE}$ for the CVWS based on ground-based measurements (Daymet), remote sensing measurements (SNODAS, UCB LRM), and land surface models (GLDAS, Terraclimate, FLDAS). Similar to $\Delta S_{SM}$, there is a high degree of variability among datasets. The remote sensing-based approaches generally produced larger magnitude estimates of $\Delta S_{SWE}$ than LSM and ground-based approaches, consistent with findings of previous studies which suggest that LSMs significantly underestimate *SWE*, especially in the Western US (Wrzesien et al., 2017; Broxton et al., 2016). Remote sensing products also demonstrated the lowest triple collocation errors, while ground-based observations had the highest. The high correlation between remote sensing-based *SWE* products (>0.9, shown in Fig. 5) (UCB LRM, available 2001–2019 and SNODAS; available 2002–2020) suggested that it was reasonable to extend the time period of $\Delta S_{SWE}$ for the PRS scenario by using UCB LRM from 2002 to 2003, using the

average of SNODAS and UCB LRM between 2003 and 2019, and using SNODAS from 2019 to 2020 (described in Section 3.5).

## 4.2. Comparison of water balance estimates of $\Delta S_{gw}$

We compared the $\Delta S_{gw}$ estimated for four scenarios (Section 3.5) at monthly time steps from 2002 to 2020 (Fig. 6). The ensemble mean calculated from the ADS, GBLSM, ARS and PRS scenarios are shown in black, purple, red and blue lines, respectively. Each scenario suggests a net loss of stored groundwater during the study period in both the CV and CVWS. The results indicate the most rapid depletion occurred during periods of severe drought between 2006–2009 and 2011–2015. There is a high degree of variance among individual ensembles, shown in Fig. 6 in light gray, which suggests that certain combinations of datasets can produce very high/low $\Delta S_{gw}$ estimates.

The net change in $\Delta S_{gw}$ derived from each water balance scenario is shown for the CV and CVWS in Table 2. $\Delta S_{gw}$ estimates are provided for the study period (2002–2020), the first drought (2006–2009), and the second drought (2011–2015). For the duration of the study period in both the CV and CVWS, the ARS scenario estimates the greatest groundwater depletion, followed by the PRS scenario, the ADS scenario, and the GBLSM scenario. Each scenario shows an accelerated loss of groundwater during the 2011–2015 drought relative to the 2006–2009 drought, with the exception of the PRS scenario in the CVWS, which shows a

**Table 2**
Comparison of $\Delta S_{gw}$ (km³) for each scenario within the CV and CVWS for the period 2002–2020, the 2006–2009 drought, and the 2011–2015 drought.

| Scenario | Net $\Delta S_{gw}$ (km³) 2002–2020 | Drought 1 Net $\Delta S_{gw}$ (km³) 2006–2009 | Drought 2 Net $\Delta S_{gw}$ (km³) 2011–2015 |
|---|---|---|---|
| CV | | | |
| ADS | −39 (−2/yr) | −19 (−6/yr) | −44 (−11/yr) |
| GBLSM | −8 (0/yr) | −8 (−3/yr) | −22 (−6/yr) |
| ARS | −75 (−4/yr) | −29 (−10/yr) | −67 (−17/yr) |
| PRS | −67 (−4/yr) | −31 (−10/yr) | −61 (−15/yr) |
| CVWS | | | |
| ADS | −85 (−5/yr) | −25 (−8/yr) | −70 (−18/yr) |
| GBLSM | −39 (−2/yr) | −18 (−6/yr) | −78 (−20/yr) |
| ARS | −246 (−14/yr) | −56 (−19/yr) | −111 (−28/yr) |
| PRS | −165 (−9/yr) | −50 (−17/yr) | −54 (−14/yr) |

Values inside parenthesis indicate changes per year.

higher rate of depletion during 2006–2009 (Fig. 6B). Further, for the 2011–2015 drought, the rate of depletion generated from the PRS scenario for the CV (−15.2 km³/year) is greater than the rate of estimated depletion from the PRS scenario in the CVWS (−13.5 km³/year). While physically possible, this implies net recharge in the CVWS, which is unrealistic, and a limitation associated with the simple modeling approach we adopted.
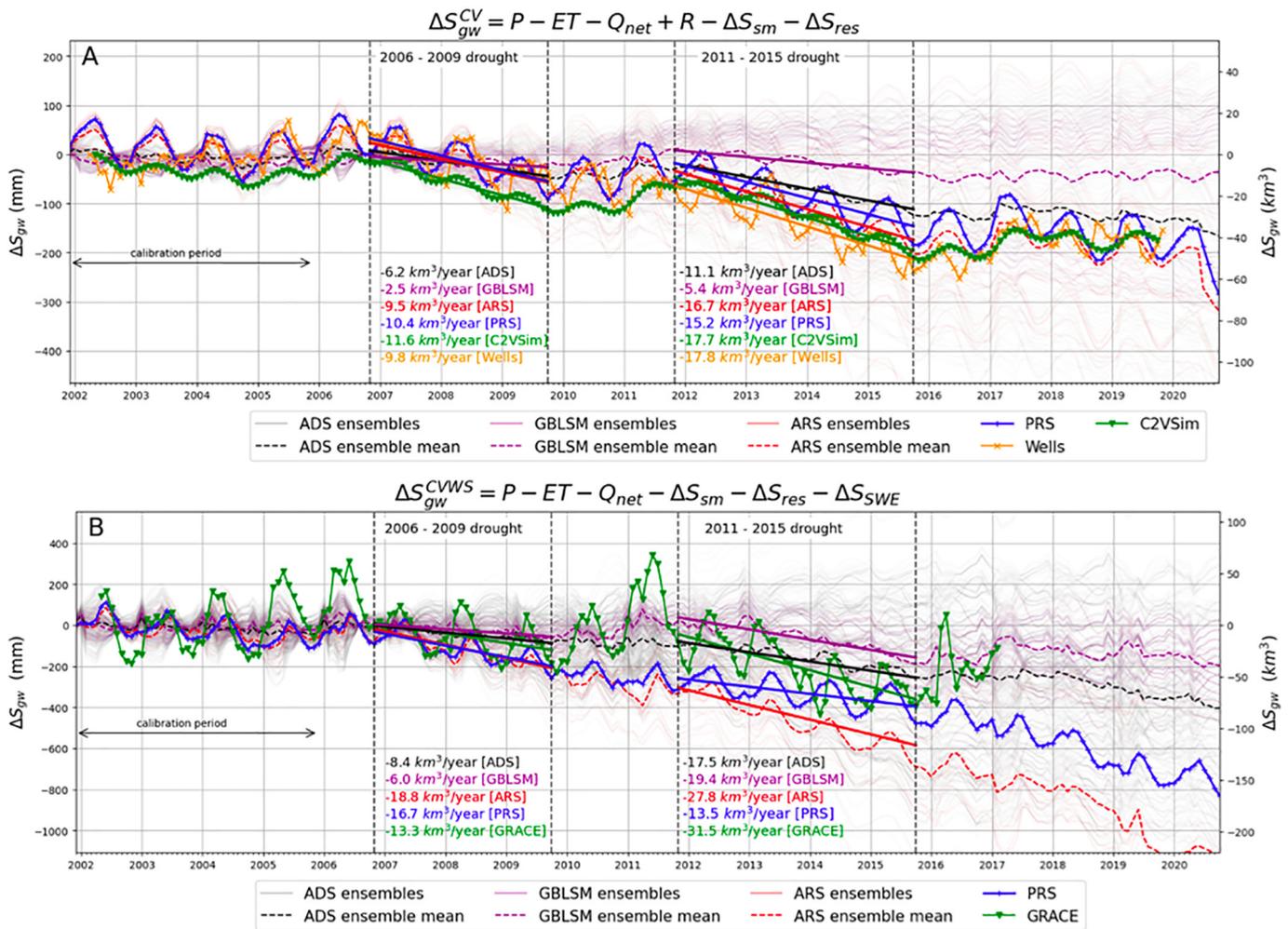


**Fig. 6.** (A) Changes in groundwater storage for the CV calculated from: (1) ADS Scenario (gray), and ADS-based ensemble mean (black dashed), (2) GBLSM scenario and (light purple) and GBLSM-based ensemble mean (purple dashed), (3) ARS Scenario (light red) and ARS-based ensemble mean (red dashed), (4) Preferred Remote Sensing (PRS) Estimate (blue), (5) C2VSim (green), and (6) wells (orange). (B) Changes in groundwater storage calculated from: (1) ADS Scenario (gray), and ADS-based ensemble mean (black dashed), (2) GBLSM scenario and (light purple) and GBLSM-based ensemble mean (purple dashed), (3) ARS Scenario (light red) and ARS-based ensemble mean (red dashed), (4) Preferred Remote Sensing (PRS) Estimate (blue), (5) GRACE (green).

### 4.3. $\Delta S_{gw}$ estimates from GRACE, wells, and C2VSim

We computed $\Delta S_{gw}$ estimates derived from GRACE, wells, and C2VSim. We note that well-based and C2VSim estimates of $\Delta S_{gw}$ cover the CV, while the footprint of GRACE-based estimates covers the entire CVWS (Fig. 1). Estimates of $\Delta S_{gw}$ are available from 2002 to 2019 for wells and C2VSim; and from 2002 to 2017 for GRACE. Table 3 shows the net change in $\Delta S_{gw}$ derived from each validation dataset in the left column, net change during the 2006–2009 drought in the middle column, and net change during over the 2011–2015 drought in the right column. Each validation dataset suggests a net loss of stored groundwater during the period of data availability, and shows a larger net decrease in stored groundwater for the 2011–2015 drought relative to the 2006–2009 drought. Among the three validation datasets, GRACE shows the highest rate of decline during droughts, but also considerable groundwater recovery toward the end of the data record (Fig. 6B).

A number of previous studies also estimated $\Delta S_{gw}$ in the CV or CVWS using a variety of methods, including groundwater wells (Scanlon et al., 2012), CVHM (Faunt et al., 2009), Land Surface Model outputs generated using the North American Land Data Assimilation System (NLDAS; Xiao et al., 2017), vertical displacement data from GPS (Argus et al., 2017), and GRACE (Famiglietti et al., 2011; Scanlon et al., 2012). Because these studies were published at different times, and span different periods of investigation, we qualitatively compare, when available, estimates of $\Delta S_{gw}$ during the 2006–2009 drought and 2011–2015 drought to estimates reported in this study. Table 3 shows the net changes in $\Delta S_{gw}$ for previous studies for the two drought periods.

Although the magnitude and trend in $\Delta S_{gw}$ varies between different methods, there are a few important patterns highlighted by all independent methods. These includes (1) the long-term groundwater depletion, (2) accelerated depletion during droughts, (3) recovery during wet years, and (4) seasonal fluctuations. An effective method for $\Delta S_{gw}$ estimation should be able to capture all these patterns. To evaluate the performance of $\Delta S_{gw}$ estimates produced from the water balance method, we compared patterns (1)–(4) generated for each scenario with the independent estimates (i.e., GRACE, well, C2VSim groundwater model). The results are detailed in Section 4.4.

### 4.4. Evaluation of remote sensing-based $\Delta S_{gw}$ estimates

We compared $\Delta S_{gw}$ estimates derived from four water balance scenarios (i.e., ADS, GBLSM, ARS, PRS) with validation datasets (i.e., GRACE, wells, C2VSim) for the CV and CVWS. In the following sections, we discuss the net change, seasonality, and the trends in $\Delta S_{gw}$ during drought and non-drought periods in the CV and CVWS as estimated by the water balance method and other approaches.

#### 4.4.1. Evaluation of $\Delta S_{gw}$ within CV region

In general, we observe that ensemble remote sensing data driven $\Delta S_{gw}$ estimates can provide upper and lower bounds to the net change (2002–2020), and effectively capture seasonality and depletion during drought events (2006–2009 and 2011–2015 drought) in the Central

Valley region. $\Delta S_{gw}$ was $-39$ km$^3$, $-8$ km$^3$, $-75$ km$^3$, and $-67$ km$^3$ in CV during 2002–2020 under ADS, GBLSM, ARS, and PRS scenarios, respectively (Table 2). Well-based estimates and the C2VSim groundwater model, available from 2002 to 2019, suggest net $\Delta S_{gw}$ of $-36$ km$^3$ and $-41$ km$^3$, respectively. Among the four scenarios (ADS, GBLSM, ARS, and PRS), remote sensing-based scenarios (ARS and PRS) showed a net change that is much closer to the C2VSim and well-based $\Delta S_{gw}$ estimates (Fig. 6A). We found a high Pearson correlation in $\Delta S_{gw}$ between the PRS and well-based (0.87) and C2VSim-based (0.94) estimates during 2002–2019 period. Both remote sensing-based estimates (ARS and PRS) showed similar seasonality, while seasonal amplitudes derived from the ARS estimate (mean) were slightly lower than the PRS amplitudes. The relatively higher net depletion found in the ARS and PRS scenarios are attributable to the higher ET magnitudes from remote sensing data compared to LSMs (Figs. 2 and 3).

The results show that the remote sensing-based $\Delta S_{gw}$ estimates (ARS and PRS) effectively capture the groundwater depletion during the two major droughts in the past two decades (2006–2009 and 2011–2015). $\Delta S_{gw}$ changed by $-31$ km$^3$ and $-29$ km$^3$ during October 2006 through October 2009 under the PRS and ARS scenarios. These estimates compare favorably to estimates of $-32$ km$^3$ from CVHM (Faunt et al., 2009); $-41$ km$^3$ from C2VSim (Brush et al., 2013); $-29$ km$^3$ from the analysis of groundwater wells (Alam et al., 2021); $-24 \pm 6$ km$^3$ from GRACE reported in Famiglietti et al. (2011) for the time period April 2006 to March 2010; and $-27$ km$^3$ and $-28$ km$^3$ calculated from wells and GRACE, respectively reported in Scanlon et al. (2012) for the time period April 2006–September 2009. For the 2011–2015 drought, estimates of $\Delta S_{gw}$ within the CV determined from the ARS and PRS scenarios agree well with independent estimates of storage losses. Both C2VSim and the well data suggest $-71$ km$^3$ of groundwater storage decline, while the PRS water balance scenario suggests $-61$ km$^3$ and the ARS water balance scenario suggest $-67$ km$^3$. These estimates are considerably higher than the ADS and GBLSM scenarios, which suggest $-44$ km$^3$ and $-21$ km$^3$, respectively; and are also higher than the $-30$ km$^3$ reported in Xiao et al. (2017). Remote sensing-based estimates of the annual rate of $\Delta S_{gw}$ CV during droughts fall within 15% of groundwater models and wells. Remote sensing estimates also suggest higher depletion during 2011–2015 drought than 2006–2009, consistent with other methods, and capture the groundwater recovery during wet years (e.g., 2006, 2011, 2017). The general agreement between the PRS and ARS scenario and independent validation datasets suggests that a remote sensing-based water balance methodology can produce reliable and timely estimates capable of capturing trends and net changes of $\Delta S_{gw}$ during droughts in a heavily irrigated region.

Seasonal fluctuations in $\Delta S_{gw}$ within the CV are reproduced in the both the ARS and PRS scenarios, shown in Fig. 6A. The ARS and PRS scenarios suggest that most recharge occurs from October to May, and most depletion occurs between May and September. Fig. 6A also shows that ARS and PRS scenarios produce seasonal fluctuations similar in magnitude to well-based estimates, but larger in magnitude than estimates from C2VSim. The timing of seasonal fluctuations from remotely sensed estimates of $\Delta S_{gw}$ is synchronous with C2VSim and slightly

**Table 3**
Comparison of net changes in $\Delta S_{gw}$ (km$^3$) derived from validation datasets within the CV and CVWS during the 2006–2009 drought (second from right column), and the 2011–2015 drought (right column).

| Method and source | Study area | Drought 1 Net $\Delta S_{gw}$ (km$^3$) 2006–2009 | Drought 2 Net $\Delta S_{gw}$ (km$^3$) 2011–2015 |
|---|---|---|---|
| Wells (this study) | CV | $-29$ | $-71$ |
| Wells (Scanlon et al., 2012) | CV | $-27 \pm 3$ | NA |
| C2VSim (this study) | CV | $-35$ | $-71$ |
| CVHM (Faunt et al., 2009) | CV | $-32$ | NA |
| NLDAS LSM (Xiao et al., 2017) | CV | $-16$ | $-30$ |
| GPS (Argus et al., 2017) | CVWS | $-59 \pm 28$ | $-104 \pm 30$ |
| GRACE (this study) | CVWS | $-40$ | $-95$ |
| GRACE (Famiglietti et al., 2011) | CVWS | $-24 \pm 6$ | NA |
| GRACE (Scanlon et al., 2012) | CVWS | $-28 \pm 5$ | NA |

precedes those in the well-based estimates. A clear seasonal cycle is not reproduced by the ADS (black dashed) and GBLSM-based scenarios (purple dashed). The ability of the remote sensing-based water balances to reproduce seasonal fluctuations suggests the possibility of monitoring ongoing seasonal or monthly $\Delta S_{gw}$ changes, which is very difficult using other methods if not impossible.

### 4.4.2. Evaluation of $\Delta S_{gw}$ within CVWS region

In the CVWS, there are fewer studies and datasets which can be used to validate estimates of $\Delta S_{gw}$. GRACE-based estimates of $\Delta S_{gw}$ for the CVWS region (red outline in Fig. 1) have previously been compared to estimates describing the CV (e.g. Scanlon et al., 2012; Xiao et al., 2017), but have also been compared to estimates for the entire CVWS (e.g. Famiglietti et al., 2011; Argus et al., 2017). Within the CVWS, Xiao et al. (2017; SI) estimated approximately −30 km³ net $\Delta S_{gw}$ between 2002 and 2017; of this, −10 km³ was attributable to the non-CV region. Argus et al. (2017) used GPS and the a priori hydrology model given in Xiao et al. (2017) to estimate approximately −100 km³ net $\Delta S_{gw}$ between 2005 and 2017. Within the CVWS for the time period 2002–2020, the ADS, GBLSM, ARS, and PRS scenarios suggest net $\Delta S_{gw}$ changes of −85 km³, −39 km³, −246 km³, and −165 km³, respectively (Table 2). The net change in $\Delta S_{gw}$ for the duration of the study period is more negative when calculated using remotely sensed data (i.e., ARS and PRS scenarios), than ground-based data and LSMs (i.e. ADS and GBLSM scenarios). This may be due to the larger-magnitude estimates of *SWE* derived from remotely sensed datasets, a key variable for large, snow-dominated regions like the CVWS, as well as inability of remotely sensed *P* to capture the highest magnitude events over mountainous terrain (Wang et al., 2021).

During the 2006–2009 drought, the PRS and ARS scenarios suggest −56 km³ and −50 km³ of $\Delta S_{gw}$, respectively. During the same time period, Argus et al. (2017) estimate −59 ± 28 km³ net change in $\Delta S_{gw}$, while Xiao et al. (2017) estimated −16 km³. Our GRACE-based estimate suggests net change in $\Delta S_{gw}$ of −40 km³, while previous GRACE-based estimates given in Famiglietti et al. (2011) and Scanlon et al. (2012), which used the older GRACE releases and spherical harmonic solution instead of the MASCON solution, suggest 24 ± 6 km³, and −28 ± 5 km³, respectively. During the 2011–2015 drought, the ARS and PRS scenarios produce estimates of −111 km³ and −54 km³, respectively. The PRS scenario underestimates $\Delta S_{gw}$ significantly relative to estimates of $\Delta S_{gw}$ determined from GRACE, which suggests −95 km³, and from GPS in Argus et al. (2017), who estimated −104 ± 30 km³. However, the PRS estimate is still 20% larger in magnitude than the estimate of −42 km³ given in Xiao et al. (2017). Argus et al. (2017) suggested that this large depletion was driven primarily by a loss of deep soil moisture or a large loss of groundwater in river alluvium and in crystalline basement in the Sierra Nevada.

While the ARS and PRS scenarios reproduce trends during droughts similar to what is found by GRACE and GPS-based approaches, the ARS and PRS water balance scenarios applied to the CVWS struggle to capture significant recharge events, which are clearly captured by GRACE in 2011 and 2017. The ARS and PRS estimates instead show a steady decline in $\Delta S_{gw}$. This may be due to an underestimation of extreme precipitation in remotely sensed products (Wang et al., 2021). For example, the 2017 winter was one of the wettest on record in California; Fig. 4 shows that the peak precipitation derived from ground-based observations (Daymet, PRISM) exceeds the peaks of previous years, but GPM-based estimates have a magnitude comparable to that in previous wet years occurring in 2006 and 2011. In contrast with estimates of $\Delta S_{gw}$ derived from remotely sensed precipitation, estimates of $\Delta S_{gw}$ derived using in situ data from PRISM and Daymet (i.e. ADS and GBLSM scenarios) do exhibit post drought recovery following 2015 as expected.

Estimates of $\Delta S_{gw}$ generated in the CVWS from both the ARS scenario and PRS Scenario reproduce seasonal fluctuations (Fig. 6B), showing most recharge occurring from October to May, and most depletion occurring between May and September. However, the magnitude of seasonal fluctuations is significantly lower than that suggested by GRACE (shown in green in Fig. 6B). The ADS (black dashed) and GBLSM-based scenarios (purple dashed) are not able to reproduce seasonal fluctuations within the CVWS.

### 4.4.3. Potential evaluation of $\Delta S_{gw}$ in other regions

In this paper, we outlined a procedure to use ensemble remote sensing data in a water balance approach to estimate changes in groundwater storage for the CV and CVWS. While the method may be extended to other areas, there are a few key steps which should be undertaken to ensure reasonable results. Firstly, a literature review should be conducted to determine and assess availability, known biases, and other limitations specific to the performance of certain datasets in the region of interest. Next, comparison and error analysis of available datasets describing each variable should be performed in order to select the optimal datasets. Lastly, ensemble calculations of $\Delta S_{gw}$ should be compared to any available existing independent estimates derived from e.g. GRACE or well data in order to assess accuracy.

### 4.5. Limitations and future outlook

Despite many advantages, a remote sensing-based solution using the water balance method is currently limited in terms of general practical applicability by a number of key factors.

Firstly, the remote sensing datasets utilized here may perform differently in regions with different climates or when applied at finer scales. Random error and biases in remote sensing products (e.g., due to elevation, temperature) can be minimized by averaging over large areas, but can persist at finer scales (Almazroui, 2011). All forms of on-the-ground data collected using, for example, snow pillows, rain gauges, stream gauges, flux towers, and groundwater monitoring wells with high sampling frequency, are therefore invaluable for assessing the regional accuracies of remotely sensed products (i.e., Table 1), and can provide an opportunity to bias-correct existing data and models to be more reflective of regional conditions. The lack of on-the-ground data in many parts of the world, to conduct the level of assessment we have completed in this study, will make it challenging to adopt, with a high level of confidence, a remote sensing-based approach for estimating $\Delta S_{gw}$.

Secondly, it remains challenging to apply the water balance method to estimate $\Delta S_{gw}$ in agriculturally intensive regions that are heavily dependent on surface water diversions. As an example, had we elected to work at a finer scale in the Central Valley, we would have required accurate quantification of water delivered through surface water infrastructure (canals, aqueducts) to apply the water balance approach. This issue is also challenging to solve using LSM-based approaches, which do not typically account for surface water infrastructure. Recent research, which uses historical climate data to estimate surface water deliveries (Goodrich et al., 2020) can help alleviate these data requirements.

Thirdly, the remote sensing-based water balance approach proposed here is suitable for regions where net groundwater inflow/outflow is relatively small compared to other water balance components. Regions with significant groundwater inflow/outflow would require additional measurements and modeling approaches in order to accurately estimate changes in groundwater storage.

The method presented in this study focused on a semiarid and highly irrigated region so it required in-situ measurements of streamflow ($Q_{in}$ and $Q_{out}$) and reservoir storage changes ($\Delta S_R$), as well as LSM-based estimates of runoff ($R$). These measurements are unavailable in many regions globally where the water in rivers and reservoirs are important components of the water balance, thus limiting the application of the proposed method. However, the data requirements of discharge and reservoir storage changes may be alleviated in the near future by the Surface Water Ocean Topography (SWOT) satellite mission (est. launch

date: November 2022, https://swot.jpl.nasa.gov/mission.htm) that will use altimetry to measure reservoir heights and water levels in rivers. Measurements of river stage can be combined with hydrologic rating curves to estimate discharge, thereby providing the necessary measurements to apply the water balance method in geographies lacking mature in-situ data collection networks. The current study is an important step toward developing a framework that can be applied to estimate $\Delta S_{gw}$ globally in all regions relying exclusively on remote sensing data.

## 5. Conclusions

Advancements in groundwater science and groundwater management require an improved ability to estimate and monitor $\Delta S_{gw}$. Water balance methods have the ability to investigate scales finer than the GRACE footprint, but without the well data requirements and computational complexity of groundwater flow models. Through validation against independent estimates, this study demonstrated the fidelity of remote sensing data to estimate $\Delta S_{gw}$ in a semiarid and agriculturally intensive region at two spatial scales using the water balance method. An ensemble procedure to estimate $\Delta S_{gw}$ using multiple remotely sensed data sources was outlined, and methods for assessing uncertainties and errors among water balance inputs were presented.

A remote sensing-based method is a promising approach for obtaining timely estimates of $\Delta S_{gw}$. Our developed approach can provide estimates of $\Delta S_{gw}$ as soon as the required data become available, which is approximately 1 to 6 months for the data used here. This could be extremely valuable in drought-prone regions where the current practice is to use streamflow or precipitation data to determine drought severity without direct awareness of the state of the groundwater system in terms of $\Delta S_{gw}$. Adopted for monitoring throughout a drought, this method could provide the critical, near-real-time information about $\Delta S_{gw}$ needed to avoid failed wells, subsidence of the ground surface, negative impacts on groundwater dependent ecosystems, and reduction of water quality. In this study, we were able to provide estimates of $\Delta S_{gw}$ until the end of 2020, which is not currently practical for other methods.

The combined use of remote sensing data and the water balance method is a promising approach that, with further testing and development, could be applied in a wide range of locations to obtain estimates of $\Delta S_{gw}$. The rapid advancement in the types, quality, and quantity of remote sensing data make it inevitable that these data will, in the not-too-distant future, become central to groundwater science and management.

## CRediT authorship contribution statement

**Aakash Ahamed:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Rosemary Knight:** Conceptualization, Resources, Writing – review & editing, Supervision, Funding acquisition. **Sarfaraz Alam:** Conceptualization, Data curation, Validation, Writing – review & editing. **Rich Pauloo:** Conceptualization, Data curation, Writing – review & editing. **Forrest Melton:** Writing – review & editing.

## Declaration of competing interest

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

## Acknowledgments

The authors declare no real or perceived financial conflicts of interest.

## References

Abatzoglou, J.T., Dobrowski, S.Z., Parks, S.A., Hegewisch, K.C., 2018. Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015. Scientific Data. https://doi.org/10.1038/sdata.2017.191.

Alam, S., Gebremichael, M., Ban, Z., Scanlon, B.R., Lettenmaier, D.P., 2021. Post-drought Groundwater Storage Recovery in California's Central Valley (Version 1). Figshare.

Alemohammad, S.H., McColl, K.A., Konings, A.G., Entekhabi, D., Stoffelen, A., 2015. Characterization of Precipitation Product Errors Across the United States Using Multiplicative Triple Collocation.

Alley, W.M., Healy, R.W., LaBaugh, J.W., Reilly, T.E., 2002. Flow and storage in groundwater systems. Science 296 (5575), 1985–1990.

Allen, R.G., Pereira, L.S., Howell, T.A., Jensen, M.E., 2011. Evapotranspiration information reporting: I. Factors governing measurement accuracy. Agric. Water Manag. 98 (6), 899–920.

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop Evapotranspiration-Guidelines for Computing Crop Water Requirements-FAO Irrigation and Drainage Paper 56. 300. FAO, Rome, p. 9 D05109.

Allen, R.G., Tasumi, M., Trezza, R., 2007. Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC)—Model. J.Irrig.Drain.Eng. 133 (4), 380–394.

Almazroui, M., 2011. Calibration of TRMM rainfall climatology over Saudi Arabia during 1998–2009. Atmos.Res. 99 (3–4), 400–414.

Anderson, M.C., Norman, J.M., Diak, G.R., Kustas, W.P., Mecikalski, J.R., 1997. A two-source time-integrated model for estimating surface fluxes using thermal infrared remote sensing. Remote Sens.Environ. 60 (2), 195–216.

Anderson, M.C., Norman, J.M., Mecikalski, J.R., Otkin, J.A., Kustas, W.P., 2007a. A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation. J. Geophys. Res. Atmos. 112 (D10).

Anderson, M.C., Norman, J.M., Mecikalski, J.R., Otkin, J.A., Kustas, W.P., 2007b. A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 2. Surface moisture climatology. J. Geophys. Res. Atmos. 112 (D11).

Argus, D.F., Landerer, F.W., Wiese, D.N., Martens, H.R., Fu, Y., Famiglietti, J.S., Thomas, B.F., Farr, T.G., Moore, A.W., Watkins, M.M., 2017. Sustained water loss in California's mountain ranges during severe drought from 2012 to 2015 inferred from gps. J. Geophys. Res. Solid Earth 122 (12).

Ashouri, H., Hsu, K.L., Sorooshian, S., Braithwaite, D.K., Knapp, K.R., Cecil, L.D., Nelson, B.R., Prat, O.P., 2015. PERSIANN-CDR: daily precipitation climate data record from multisatellite observations for hydrological and climate studies. Bull.Am.Meteorol. Soc. 96 (1), 69–83.

Barlage, M., Tewari, M., Chen, F., Miguez-Macho, G., Yang, Z.L., Niu, G.Y., 2015. The effect of groundwater interaction in northamerican regional climate simulations with WRF/Noah-MP. Clim.Chang. 129 (3), 485–498.

Becker, M.W., 2006. Potential for satellite remote sensing of ground water. Groundwater 44 (2), 306–318 Vancouver.

Bierkens, M.F., Wada, Y., 2019. Non-renewable groundwater use and groundwater depletion: a review. Environ.Res.Lett. 14 (6), 063002.

Brush, C.F., Dogrul, E.C., Kadir, T.N., 2013. Development and Calibration of the California Central Valley Groundwater-Surface Water Simulation Model (C2VSim), Version 3.02-CG. Bay-Delta Office, California Department of Water Resources.

Broxton, P.D., Zeng, X., Dawson, N., 2016. Why do global reanalyses and land data assimilation products underestimate snow water equivalent? J.Hydrometeorol. 17 (11), 2743–2761.

Bolten, J., Crow, W.T., Zhan, X., Jackson, T.J., Reynolds, C.A., 2010. Evaluating the utility of remotelysensedsoilmoistureretrievals for operationalagriculturaldroughtmonitoring. IEEE Trans.Geosci. Remote Sens. 3 (1), 57–66. https://doi.org/10.1109/JSTARS.2009.2037163.

Borsa, A.A., Agnew, D.C., Cayan, D.R., 2014. Ongoing drought-induced uplift in the western United States. Science 345 (6204), 1587–1590.

California Data Exchange Center (CDEC), 2021. . Accessed July, 2021, at URL: https://cdec.water.ca.gov/reportapp/javareports?name=MonthlyRes.

Chakraborty, A., Kumar, R., Stoffelen, A., 2013. Validation of ocean surface winds from the OCEANSAT-2 scatterometer using triple collocation. Remote Sens. Lett. 4 (1), 84–93.

Cook, B.I., Ault, T.R., Smerdon, J.E., 2015. Unprecedented 21st century drought risk in the americansouthwest and Central Plains. Sci.Adv. 1 (1), e1400082.

Cooper, M.G., Schaperow, J.R., Cooley, S.W., Alam, S., Smith, L.C., Lettenmaier, D.P., 2018. Climate elasticity of low flows in the maritime western US mountains. Water Resour.Res. 54 (8), 5602–5619. https://doi.org/10.1029/2018WR022816.

Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Pasteris, P.P., 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int. J. Climatol. 28 (15), 2031–2064.

Daly, C., Smith, J.I., Olson, K.V., 2015. Mapping atmospheric moisture climatologies across the conterminous United States. PloS One 10 (10), e0141140.

Deines, J.M., Kendall, A.D., Hyndman, D.W., 2017. Annual irrigation dynamics in the US northern High Plains derived from landsat satellite data. Geophys.Res.Lett. 44 (18), 9350–9360.

Department of Water Resources (DWR), 2021. Dayflow. [Data file]. Retrieved from: https://water.ca.gov/Programs/Environmental-Services/Compliance-Monitoring-And-Assessment/Dayflow-Data Accessed June 2021.

Department of Water Resources California Statewide Groundwater Elevation Monitoring (DWR CASGEM), 2021. Periodic groundwater level measurements. [Data file]. https://data.cnra.ca.gov/dataset/periodic-groundwater-level-measurements.

Diffenbaugh, N.S., Swain, D.L., Touma, D., 2015. Anthropogenic warming has increased drought risk in California. Proc. Natl. Acad. Sci. 112 (13).

Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., Johnson, J., Kimball, J., 2010. The soil moisture active passive (SMAP) mission. Proc. IEEE 98 (5), 704–716.

Famiglietti, J.S., Lo, M., Ho, S.L., Bethune, J., Anderson, K.J., Syed, T.H., Swenson, S.C., De Linage, C.R., Rodell, M., 2011. Satellites measure recent rates of groundwater depletion in California's Central Valley. Geophys.Res.Lett. 38 (3).

Famiglietti, J.S., 2014. The global groundwater crisis. Nat.Clim.Chang. 4 (11), 945.

Faunt, C.C., Hanson, R.T., Belitz, K., Schmid, W., Predmore, S.P., Rewis, D.L., McPherson, K., 2009. Chapter C—numerical model of the hydrologic landscape and groundwater flow in California's Central Valley. Groundwater Availability of the Central Valley Aquifer, California: US Geological Survey Professional Paper. 1766, pp. 121–212.

Faunt, C.C., Sneed, M., 2015. Water availability and subsidence in California's Central Valley. San Francisco Estuary and Watershed. Science 13 (3).

Fisher, J.B., Tu, K.P., Baldocchi, D.D., 2008. Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. Remote Sens.Environ. 112 (3), 901–919.

Foken, T., 2008. The energy balance closure problem: an overview. Ecol.Appl. 18 (6), 1351–1367.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Sci. data 2, 150066.

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., 2017. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). J.Clim. 30 (14), 5419–5454.

Goodrich, J.P., Cayan, D.R., Pierce, D.W., 2020. Climate and land-use controls on surface water diversions in the Central Valley, California. San Franc. Estuary Watershed Sci. 18 (1).

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earthengine:planetary-scale geospatial analysis for everyone. Remote Sens.Environ. 202, 18–27.

Gruber, A., Su, C.H., Zwieback, S., Crow, W., Dorigo, W., Wagner, W., 2016. Recent advances in (soil moisture) triple collocation analysis. Int.J.Appl. Earth Obs.Geoinf. 45, 200–211.

Hanak, E., Escriva-Bou, A., Gray, B., Green, S., Harter, T., Jezdimirovic, J., Lund, J., Medellín-Azuara, J., Moyle, P., Seavy, N., 2019. Water and the Future of the San Joaquin Valley. Public Policy Institute of California, p. 100.

Hanson, R.T., Flint, L.E., Flint, A.L., Dettinger, M.D., Faunt, C.C., Cayan, D., Schmid, W., 2012. A method for physically based model analysis of conjunctive use in response to potential climate changes. Water Resour. Res. 48, 1–23. https://doi.org/10.1029/2011WR010774.

Hart, Q.J., Brugnach, M., Temesgen, B., Rueda, C., Ustin, S.L., Frame, K., 2009. Daily reference evapotranspiration for California using satellite imagery and weather station measurement interpolation. Civ. Eng. Environ. Syst. 26 (1), 19–33.

Huffman, G.J., Stocker, E.F., Bolvin, D.T., Nelkin, E.J., Tan, Jackson, 2019. GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06, Greenbelt, MD. Goddard Earth Sciences Data and Information Services Center (GES DISC). https://doi.org/10.5067/GPM/IMERG/3B-HH/06. (Accessed  July 2021).

Kumar, S., Holmes, T., Mocko, D., Wang, S., Peters-Lidard, C., 2018. Attribution of flux partitioning variations between land surface models over the continental us. Remote Sens. 10 (5), 751.

Lakshmi, V., 2016. Beyond GRACE: using satellite data for groundwater investigations. Groundwater 54 (5), 615–618.

Liou, Y.A., Kar, S., 2014. Evapotranspiration estimation with remote sensing and various surface energy balance algorithms—a review. Energies 7 (5), 2821–2849.

McNally, A., Arsenault, K., Kumar, S., Shukla, S., Peterson, P., Wang, S., Funk, C., Peters-Lidard, C.D., Verdin, J.P., 2017. A land data assimilation system for sub-Saharan Africa food and water security applications. Sci. Data 4, 170012.

Melton, F.S., Johnson, L.F., Lund, C.P., Pierce, L.L., Michaelis, A.R., Hiatt, S.H., Guzman, A., Adhikari, D.D., Purdy, A.J., Rosevelt, C., Votava, P., 2012. Satellite irrigation management support with the terrestrial observation and prediction system: a framework for integration of satellite and surface observations to support improvements in agricultural water resource management. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 5 (6), 1709–1721.

Mendiguren, G., Koch, J., Stisen, S., 2017. Spatial pattern evaluation of a calibrated national hydrological model–a remote-sensing-based diagnostic approach. Hydrol. Earth Syst. Sci. 21 (12), 5987–6005.

Monteith, J.L., 1965. Evaporation and environment. Symposia of the Society for Experimental Biology. 19. Cambridge University Press (CUP), Cambridge, pp. 205–234.

Mu, Q., Zhao, M., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sens.Environ. 115 (8), 1781–1800.

Mu, Q., Zhao, M., Running, S.W., 2013. MODIS Global Terrestrial Evapotranspiration (ET) Product (NASA MOD16A2/A3). Algorithm Theoretical Basis Document, Collection, p. 5.

Nady, P. and Larragueta, L.L., 1983. Estimated average annual streamflow into the Central Valley of California (No. 657).

National Aeronautics and Space Administration (NASA), 2021. Monthly Mass Grids – Global Mascons (JPL RL06). Retrieved from: https://grace.jpl.nasa.gov/data/get-data/jpl_global_mascons/ Accessed June 2021.

National Operational Hydrologic Remote Sensing Center (NOHRC), 2004. Snow Data Assimilation System (SNODAS) Data Products at NSIDC, Version 1. [Subset tS__ssmv11034]. NSIDC: National Snow and Ice Data Center, Boulder, Colorado USA https://doi.org/10.7265/N5TB14TC Accessed June 2021.

Niu, G.Y., Yang, Z.L., Mitchell, K.E., Chen, F., Ek, M.B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., 2011. The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. J. Geophys. Res. Atmos. 116 (D12).

Pan, M., Sahoo, A.K., Troy, T.J., Vinukollu, R.K., Sheffield, J., Wood, E.F., 2012. Multisource estimation of long-term terrestrial water budget for major global river basins. J. Clim. 25 (9), 3191–3206.

Pearson, K., 1895. VII. Note on regression and inheritance in the case of two parents. Proc. R. Soc. Lond. 58 (347-352), 240–242.

Pereira, L.S., Paredes, P., Melton, F., Johnson, L., Wang, T., López-Urrea, R., Cancela, J.J., Allen, R.G., 2020. Prediction of crop coefficients from fraction of ground cover and height. background and validation using ground and remote sensing data. Agric. Water Manag. 241, 106197.

Pool, D.R., Eychaner, J.H., 1995. Measurements of aquifer-storage change and specific yield using gravity surveys. Groundwater 33 (3).

Rodell, M., Houser, P.R., Jambor, U.E.A., Gottschalck, J., Mitchell, K., Meng, C.J., Entin, J.K., 2004. The global land data assimilation system. Bull.Am.Meteorol.Soc. 85 (3), 381–394.

Rodell, M., Velicogna, I., Famiglietti, J.S., 2009. Satellite-based estimates of groundwater depletion in India. Nature 460 (7258), 999.

Sahoo, A.K., Pan, M., Troy, T.J., Vinukollu, R.K., Sheffield, J., Wood, E.F., 2011. Reconciling the global terrestrial water budget using satellite remote sensing. Remote Sens. Environ. 115 (8), 1850–1865.

Schneider, D., Molotch, N.P., 2016. Real-time estimation of snow water equivalent in the upper Colorado River basin using MODIS-based SWE reconstructions and SNOTEL data. Water Resour.Res. 52 (10), 7892–7910.

Senay, G.B., Bohms, S., Singh, R.K., Gowda, P.H., Velpuri, N.M., Alemu, H., et al., 2013. Operational evapotranspirationmapping using remote sensing andweather datasets: a new parameterization for the SSEB approach. J.Am. Water Resour.Assoc. 1–15. https://doi.org/10.1111/jawr.12057.

Senay, G.B., Kagone, S., Velpuri, N.M., 2020. Operational global actual evapotranspiration: development, evaluation, and dissemination. Sensors 20 (7), 1915.

Scanlon, B.R., Longuevergne, L., Long, D., 2012. Ground referencing grace satellite estimates of groundwater storage changes in the California central valley, Usa. Water Resour. Res. 48.

Schmid, W., Hanson, R.T., Maddock III, T., Leake III, S.A., 2004. User guide for the farm process (FMP1) for the US Geological Survey's modular three-dimensional finite-difference ground-water flow model, MODFLOW-2000. US Geological Survey Techniques and Methods, 6-A17.

Scipal, K., Holmes, T., De Jeu, R., Naeimi, V., Wagner, W., 2008. A possible solution for the problem of estimating the error structure of global soil moisture data sets. Geophys. Res.Lett. 35 (24).

Siebert, S., Burke, J., Faures, J.M., Frenken, K., Hoogeveen, J., Döll, P., Portmann, F.T., 2010. Groundwater use for irrigation–a global inventory. Hydrol.EarthSyst.Sci. 14 (10), 1863–1880.

Smith, R.G., Knight, R., Chen, J., Reeves, J.A., Zebker, H.A., Farr, T., Liu, Z., 2017. Estimating the permanent loss of groundwater storage in the southern San Joaquin Valley, California. Water Resour. Res. 53 (3), 2133–2148.

Stoffelen, A., 1998. Toward the true near-surface wind speed: error modeling and calibration using triple collocation. J. Geophys. Res. Oceans 103 (C4), 7755–7766.

Swain, D.L., Langenbrunner, B., Neelin, J.D., Hall, A., 2018. Increasing precipitation volatility in twenty-first-century California. Nat.Clim.Chang. 8 (5), 427.

Swenson, S.C., 2012. GRACE Monthly Land Water Mass Grids NETCDF RELEASE 5.0. Ver. 5.0. PO.DAAC, CA, USA. https://doi.org/10.5067/TELND-NC005 Dataset accessed [2021-05-20] at:.

Tian, Y., Peters-Lidard, C.D., 2010. A global map of uncertainties in satellite-based precipitation measurements. Geophys.Res.Lett. 37 (24).

Thornton, P.E., Thornton, M.M., Mayer, B.W., Wei, Y., Devarakonda, R., Vose, R.S., Cook, R.B., 2018. Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 3. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1328.

USDA National Agricultural Statistics Service Cropland Data Layer, 2021. Published crop-specific data layer.[Online]. Available at:USDA-NASS, Washington, DC (accessed June 10, 2019; verified August 15, 2019) https://nassgeodata.gmu.edu/CropScape/.

U.S. Geological Survey, 2021. National Water Information System data available on the World Wide Web.  (USGS Water Data for the Nation), accessed July, 2021, at URL https://waterdata.usgs.gov/nwis/.

Velpuri, N.M., Senay, G.B., Singh, R.K., Bohms, S., Verdin, J.P., 2013. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: using point and gridded FLUXNET and water balance ET. Remote Sens.Environ. 139, 35–49.

Wada, Y., Van Beek, L.P., Van Kempen, C.M., Reckman, J.W., Vasak, S., Bierkens, M.F., 2010. Global depletion of groundwater resources. Geophys.Res.Lett. 37 (20).

Wang, J., Petersen, W.A., Wolff, D.B., 2021. Validation of satellite-basedprecipitationproducts from TRMM to GPM. Remote Sens. 13 (9), 1745.

Wang-Erlandsson, L., Bastiaanssen, W.G., Gao, H., Jagermeyr, J., Senay, G.B., Van Dijk, A., Guerschman, J.P., Keys, P.W., Gordon, L.J., Savenije, H.H., 2016. Global root zone storage capacity from satellite-based evaporation. Hydrol. Earth Syst. Sci. 20, 1459–1481. https://doi.org/10.5194/hess-20-1459-2016.

Watkins, M.M., Wiese, D.N., Yuan, D.-N., Boening, C., Landerer, F.W., 2015. Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. J. Geophys. Res. Solid Earth 120. https://doi.org/10.1002/2014JB011547.

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

Wiese, D.N., 2015. GRACE Monthly Global Water Mass Grids NETCDF RELEASE 5.0. Ver. 5.0. PO.DAAC, CA, USA. https://doi.org/10.5067/TEMSC-OCL05 D05109.

Wrzesien, M.L., Durand, M.T., Pavelsky, T.M., Howat, I.M., Margulis, S.A., Huning, L.S., 2017. Comparison of methods to estimate snow water equivalent at the mountain range scale: a case study of the California Sierra Nevada. J.Hydrometeorol. 18 (4), 1101–1119.

Xiao, M., Koppa, A., Mekonnen, Z., Pagán, B.R., Zhan, S., Cao, Q., Lettenmaier, D.P., 2017. How much groundwater did California's Central Valley lose during the 2012–2016 drought? Geophys.Res.Lett. 44 (10), 4872–4879.

Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S.M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., 2018. A new generation of the United States National Land Cover Database: requirements, research priorities, design, and implementation strategies. ISPRS J.Photogramm. Remote Sens. 146, 108–123.

Yin, J., Medellín-Azuara, J., Escriva-Bou, A., Liu, Z., 2021. Bayesian machine learning ensemble approach to quantify model uncertainty in predicting groundwater storage change. Sci. Total Environ. 769, 144715.