Inferring aquifer storage parameters using satellite and in situ measurements: Estimation under uncertainty

Alexander Y. Sun,1 Ronald Green,1 Matthew Rodell,2 and Sean Swenson3

Received 11 March 2010; revised 12 April 2010; accepted 20 April 2010; published 19 May 2010.

[1] We present a robust optimization method for estimating aquifer storage parameters (specific yield or storativity) using the Gravity Recovery and Climate Experiment (GRACE) data, in situ well level observations, and other ancillary information. Uncertainty inherent in the remotely sensed and in situ time series can adversely affect the parameter estimation process and, in the worse case, make the solution completely meaningless. Our estimation problem is formulated to directly minimize the negative impact of data uncertainty by incorporating bounds on data variations. We demonstrate our method for the interconnected Edwards-Trinity Plateau and Pecos Valley aquifers in central Texas. The study area is divided into multiple zones based on the geology and monitor well coverage. Our estimated aquifer storage parameters are consistent with previous results obtained from pumping tests and model calibration, demonstrating the potential of using GRACE data for validating regional groundwater model parameters. Citation: Sun, A. Y., R. Green, M. Rodell, and S. Swenson (2010), Inferring aquifer storage parameters using satellite and in situ measurements: Estimation under uncertainty, Geophys. Res. Lett., 37, L10401, doi:10.1029/2010GL043231.

1. Introduction

[2] Although the fast advance of remote and in situ sensing technologies has significantly improved our capability to characterize hydrologic processes, the inherent uncertainty and limitations associated with various Earth observation products remain obstacles to those who wish to incorporate these data in their decision making processes. In this work, a robust optimization framework is introduced for estimating aquifer storage parameters using satellite and in situ measurements while minimizing the adverse impact of data uncertainty.

[3] GRACE, a joint satellite mission of the U.S. National Aeronautics and Space Administration (NASA) and German Space Agency, has been making detailed measurements of Earth’s gravity field since March 2002. GRACE provides the first opportunity for hydrologists to estimate terrestrial water storage (TWS) variations, or \( \Delta TWS \), over regional watersheds. To trace variations of individual hydrological components such as groundwater, the GRACE \( \Delta TWS \) must be disaggregated using ancillary information to remove the contributions of all other significant water storage components (e.g., snow and surface waters). A major challenge, as described next, is that all remotely sensed and in situ time series used in these analyses can be uncertain.

[4] For example, the accuracy of GRACE \( \Delta TWS \) is limited by (a) instrumentation error, (b) inaccuracies in atmospheric and ocean fields used to remove the effects of atmosphere and ocean from GRACE observations, and (c) leakage error arising from using a limited range of spherical harmonics to represent the gravity field variations [Seo et al., 2006]. Decorrelation and smoothing are commonly applied to improve the signal-to-noise ratios of the GRACE \( \Delta TWS \) estimates; however, these operations may introduce further uncertainty in \( \Delta TWS \) [Swenson et al., 2003; Swenson and Wahr, 2006a, 2006b; Seo et al., 2006]. For near-surface soil moisture (SM), the North American Land Data Assimilation System (NLDAS) has been used as independent estimates for removing soil moisture storage changes (\( \Delta SM \)) from the GRACE \( \Delta TWS \) [e.g., Strassberg et al., 2007]. Uncertainty in NLDAS data may result from uncertainties in land surface models and in atmospheric forcing variables [Luo et al., 2003]. The estimated groundwater level variations can be uncertain because of the sparse coverage of groundwater monitor wells. In many situations, one can have access to only the estimated bounds of these uncertain time series. A frequent need is, therefore, to integrate these bounds on uncertainty while performing parameter estimation.

[5] In the following, we first present a general robust optimization approach for inferring aquifer storage parameters and then demonstrate it using data related to the interconnected Edwards Trinity Plateau (ETP) and Pecos Valley (PV) aquifers (ETP-PV hereinafter).

2. Problem Formulation

[6] The groundwater storage change (in terms of equivalent water thickness), \( \Delta GWS \), can be related to average water level or hydraulic head changes (\( \Delta h \)) as [Fetter, 1994]

\[
\Delta GWS = \frac{1}{C} \sum_{j} S_{j} C_{j} \Delta h_{j}, \quad C = \sum_{j} C_{j}
\]

(1)

where \( S \) is the specific yield for unconfined aquifers or storativity for confined aquifers; \( N \) is the number of subareas or zones defined for a study region; \( C_{j} \) are the sizes of subareas; and \( C \) is the total aquifer area. \( \Delta GWS \) can also be estimated by subtracting the sum of all other storage components (\( \Delta V_{k} \)) from the GRACE \( \Delta TWS \)

\[
\Delta GWS = \Delta TWS - \sum_{k} \Delta V_{k}
\]

(2)
where $K$ is the number of other storage components. Combining Eqs (1) and (2), we obtain a linear system with spatially distributed $S_j$ as unknowns

$$
\frac{1}{C} \sum_{j} S_j C h_j \approx \left( \Delta TWS_i - \sum_{k} \Delta V_{ik} \right), i = 1, \ldots, M
$$

or equivalently

$$
Ax \approx b
$$

where $M$ is the number of periods where data are available; rows of the $M \times N$ matrix $A$ correspond to the average $\Delta h$ (area-weighted) for the $i$-th period; columns of $A$ correspond to average $\Delta h$ for the $j$-th zone; elements of $x$ are the unknown $S_j$ values for different zones; and elements of $b$ correspond to the $D_{GWS}$ values estimated from GRACE and NLDAS time series [i.e., Eqn (2)]. The number of zones to be used for a certain study region may be determined based on the knowledge of regional hydrostratigraphy and coverage of monitor networks.

The solution to Eqn (4) is usually sought in a least squares sense:

$$
\min_{x} \| Ax - b \|
$$

where $\| \cdot \|$ represents the Euclidean norm. For reasons mentioned in Section 1, $A$ and $b$ are both uncertain and the errors are not necessarily Gaussian distributed. As a result, the ordinary least square method is no longer the optimal estimator. We formulate a robust optimization problem (robust least squares to be more specific), in which a robust solution to Eqn (5) is sought that satisfies constraints for every possible realization of the uncertain parameters within the user-defined uncertainty bounds [El Ghaoui and Lebret, 1997; Lobo et al., 1998; Sun et al., 2006; Ben-Tal et al., 2009]. The robust optimization paradigm only requires knowing the uncertainty bounds of parameters, but not their actual probability distributions.

Let $a_i \in \mathbb{R}^N$ represent rows in $A$ and $b_i$ represents elements of $b$. We now assume that $a_i$ is in an ellipsoidal uncertainty region and $b_i$ is subject to interval uncertainty [Ben-Tal et al., 2009]

$$
a_i \in \Phi_i, \Phi_i = \{ \bar{a}_i + P_i u : \| u \| \leq 1 \} \quad (6a)
$$

$$
b_i \in [\bar{b}_i - p_i, \bar{b}_i + p_i] \quad (6b)
$$

where $\bar{a}_i$ and $\bar{b}_i$ represent nominal values of $a_i$ and $b_i$, respectively; $P_i u$ and $p_i$ are user-defined uncertainty bounds. The nominal values represent the best estimates one can obtain to support the “here-and-now" decision making. The ellipsoidal uncertainty structure used in Eqn (6a) is quite general. For example, it can be used to represent an $N$-dimensional confidence ellipsoid centered at $\bar{a}_i$, where the confidence intervals associated with each element of $a_i$ are used to specify the uncertainty bounds. If the errors are independent, $P_i$ is simply a diagonal matrix with standard deviations of $a_i$’s elements in its diagonal.

For the uncertainty structures specified in Eqs (6a) and (6b), a robust optimization problem may be formulated to minimize the worst case bounds [Lobo et al., 1998]

$$
\min \left[ \sum_{i} (|\bar{a}_i^T x - \bar{b}_i| + \| P_i x \| + p_i) \right]^{1/2}, i = 1, \ldots, M
$$

Figure 1. The ETP-PV study region and locations of groundwater monitor wells. The wells are divided into four zones (dark solid lines) based on the regional hydrostratigraphy and well locations.
The norm $\|P_x\|$ can be considered a regularization term that penalizes large $x$ in directions with considerable uncertainty. Eqn (7) is equivalent to the following form

$$\begin{align*}
    \min_s & \quad s \\
    \text{s.t.} & \quad \|t\| \leq s \\
    & \quad a_i^T x - b_i + ||P_x|| + \rho_i \leq t_i, i = 1, \ldots, M \\
    & \quad L \leq x \leq U
\end{align*}$$

where $t$ and $s$ are introduced to convert the objective function in Eqn (7) into a number of constraints to facilitate problem solving, and $l:1$ represents the absolute value. A linear constraint is added for the user to impose lower and upper limits ($L$ and $U$) on $x$. Eqn. (8) defines our robust optimization problem and it is solved using the Matlab toolbox SeDuMi [Sturm, 1998].

3. Demonstration

3.1. ETP-PV Study Area

[10] The interconnected ETP-PV aquifers, extending over an area of 115,000 km$^2$, provide water supplies to all or part of 39 counties in the semiarid to arid western United States (Figure 1). The ETP aquifer is one of the largest contiguous karst regions in the United States [Kastning, 1984]. The extreme annual variability of precipitation over ETP-PV and the expected increase in well pumpage have raised concerns over future groundwater availability in the region, especially under drought conditions [Anaya and Jones, 2009]. Analyses of water availability are directly affected by aquifer storage parameters. Measuring these parameters using in situ methods usually involves considerable uncertainty [Scanlon et al., 2002] and the results may not be valid at the regional level. Our main interest is to investigate how the integrated use of satellite data, in situ observations, and the robust optimization techniques may improve the reliability of such water availability analyses.

3.2. Data and Data Processing

[11] We assume that $\Delta SM$ and $\Delta GWS$ account for most of the $\Delta TWS$ variability in the semiarid to arid ETP-PV region, where there is no significant presence of other hydrologic components such as surface water.

3.2.1. Water Level Measurements

[12] Water level measurements were extracted from a Texas Water Development Board (TWDB) groundwater database (available at http://www.twdb.state.tx.us/publications/reports/GroundWaterReports/GWDatabaseReports/GWdatabasept.htm), which includes periodic measurements collected by TWDB and its cooperators from a statewide monitoring network consisting of more than 6,500 wells [Boghici, 2008]. Active pumping wells and wells that are potentially affected by active pumping were excluded from consideration (R. Boghici, personal communication, October 2009), resulting in only 30 wells satisfying our search criteria (Figure 1). The wells were grouped into four zones (Figure 1) on the basis of major hydrostratigraphic units identified for the region [Anaya and Jones, 2009, Figure 5–1]. The areas of the four zones are approximately 15,000, 24,000, 46,000, and 30,000 km$^2$, respectively. A linear trend was removed from each of the water level time series to obtain $\Delta h$. Seasonal averages of $\Delta h$ and the associated standard deviations were calculated first for each of the wells and then for each of the four zones. The area-weighted, seasonal $\Delta h$ averages are used as nominal values for elements of $A$.

3.2.2. GRACE $\Delta TWS$

[13] The monthly $\Delta TWS$ was obtained for the study region for the period from August 2002 to December 2007. Briefly, a decorrelation filter was first applied to the GRACE data produced by the Center for Space Research (Release 4), followed by Gaussian smoothing using a radius of 300 km. Filtering naturally attenuates signal strength. A scaling parameter was calculated for the study area to restore the attenuated GRACE signals by using the Global Land Data Assimilation System (GLDAS) data [Rodell et al., 2004]. Because of its coarse resolution, GRACE does not detect $\Delta TWS$ of an aquifer in isolation, but a region centered about the aquifer. So the second purpose of scaling is to use the GLDAS model simulation to relate the regional average (sampled in a manner consistent with GRACE processing) to the aquifer average. This so-called basin-specific averaging kernel was introduced by Swenson and Wahr [2002, 2006b] to minimize the effects of GRACE uncertainty while faithfully representing the water storage changes in a region. Previous simulations show that the aquifer average can be well represented by applying this scaling procedure [Swenson and Wahr, 2006b]. Neither $\Delta SM$ nor $\Delta GWS$ data is filtered. So the side effect of scaling is to restore the $\Delta TWS$ to a state that can be directly disaggregated by other data series. The resulting scaling parameter of 1.5 was applied to the monthly $\Delta TWS$. The root-mean-square (RMS) error associated with the estimated monthly $\Delta TWS$ is 1.9 cm. A smoothed seasonal $\Delta TWS$ time series was constructed by applying a low-pass filter consisting of six terms (mean, linear trend, annual sine and cosine, semianurnal sine and cosine) to the monthly series [Yeh et al., 2006]. From the monthly RMS, the seasonal $\Delta TWS$ RMS was estimated to be 1.1 cm [Strassberg et al., 2007, Eqn. 2].

3.2.3. NLDAS $\Delta SM$

[14] The monthly $\Delta SM$ was obtained from the NLDAS, which uses hourly observation-based precipitation and solar radiation, and high-quality soil and vegetation parameter fields as inputs [Cosgrove et al., 2003]. A smoothed seasonal $\Delta SM$ time series was then constructed in the same way as for the monthly $\Delta TWS$ series. Uncertainty in $\Delta SM$ may be estimated by comparing outputs of different land surface model outputs with ground truth and calculating a coefficient of variation (CV) [e.g., Kato et al., 2007]. Without in situ SM measurements, we chose a conservative CV of 1.0 in our study. This value is much larger than the CV of 0.3 calculated by Kato et al. [2007] for their Tongyu reference site, which has similar climatology as our site. Multiplying the CV with the mean seasonal $\Delta SM$ yields an uncertainty bound of 0.5 cm for $\Delta SM$.

4. Results and Discussion

[15] Figure 2 compares the original and smoothed monthly GRACE $\Delta TWS$ (Figure 2a) and NLDAS $\Delta SM$ (Figure 2b) series. The $\Delta TWS$ of the study region did not exhibit significant linear trend during the 5-year study period. The $\Delta TWS$ and $\Delta SM$ series have a Pearson’s correlation coefficient of 0.77. We used the robust optimization in Section 2 to estimate $S$ values for the four zones defined in Figure 1 by solving Eqn (8). The uncertainty bounds on elements of $A$ and $b$ were established as follows: (a) the
reflecting the confined nature of that portion of the aquifer.

corresponding to Zone 2 exhibit the largest variability, seasonal

There is a phase lag of about one to two seasons between the seasonal precipitation (bar plot) and precipitation, indicating that the ordinary least squares estimates can vary over several orders of magnitudes for some zones. Thus, the ordinary least squares estimates are highly unreliable without regularizing the impact of uncertainty.

5. Conclusion

[19] Aquifer storage parameters (S) play an important role in transient groundwater flow simulations and in water resources planning. However, determining representative S values using traditional techniques (e.g., pumping test) is challenging, especially at the regional scale. A robust optimization method is formulated for obtaining distributed S values using the additional information from GRACE, essentially enabling a downscaling process. Previous GRACE studies highlight the importance of identifying distributed S values, instead of assuming a uniform global value to all wells in the study area [Rodell et al., 2007]. Our work complements such a goal. Finally, we emphasize the importance of maintaining high-quality in situ monitoring networks, which will significantly help extend the benefits of GRACE data for groundwater resources analyses.
References


R. Green and A. Y. Sun, Geosciences and Engineering Division, Southwest Research Institute, San Antonio, TX 78238, USA. (asun@cnwra.swri.edu)

M. Rodell, Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.

S. Swenson, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309, USA.