

GRACE Improves Seasonal Groundwater Forecast Initialization over the United States

AUGUSTO GETIRANA,^{a,b} MATTHEW RODELL,^a SUJAY KUMAR,^a HIROKO KATO BEAUDOING,^{a,b} KRISTI ARSENAULT,^{a,c} BENJAMIN ZAITCHIK,^d HIMANSHU SAVE,^c AND SRINIVAS BETTADPUR^e

^a *Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, Maryland*

^b *Earth System Science Interdisciplinary Center, University of Maryland, College Park, College Park, Maryland*

^c *Science Applications International Corporation, Reston, Virginia*

^d *Department of Earth and Planetary Science, The Johns Hopkins University, Baltimore, Maryland*

^e *Center for Space Research, The University of Texas at Austin, Austin, Texas*

(Manuscript received 25 April 2019, in final form 29 October 2019)

ABSTRACT

We evaluate the impact of Gravity Recovery and Climate Experiment data assimilation (GRACE-DA) on seasonal hydrological forecast initialization over the United States, focusing on groundwater storage. GRACE-based terrestrial water storage (TWS) estimates are assimilated into a land surface model for the 2003–16 period. Three-month hindcast (i.e., forecast of past events) simulations are initialized using states from the reference (no data assimilation) and GRACE-DA runs. Differences between the two initial hydrological condition (IHC) sets are evaluated for two forecast techniques at 305 wells where depth to water table measurements are available. Results show that using GRACE-DA-based IHC improves seasonal groundwater forecast performance in terms of both RMSE and correlation. While most regions show improvement, degradation is common in the High Plains, where withdrawals for irrigation practices affect groundwater variability more strongly than the weather variability, which demonstrates the need for simulating such activities. These findings contribute to recent efforts toward an improved U.S. drought monitoring and forecast system.

1. Introduction

In the past decade, the United States has faced a number of severe droughts (e.g., Famiglietti and Rodell 2013), affecting many sectors, such as agriculture, ecosystem services, energy, human health, and water resources, and costing the country's economy billions of dollars per year (NCDC 2014). Drought is a recurrent climatic feature whose impacts are anticipated to worsen both in the United States and globally as a result of climate change and population increase. The Gravity Recovery and Climate Experiment (GRACE) mission (Tapley et al. 2004) enabled satellite-based monitoring of global water storage trends and extreme events, giving the scientific community unprecedented insights into terrestrial water storage (TWS) variations around the world (Rodell et al. 2018). GRACE-based TWS estimates have been used in many drought-related studies

worldwide (e.g., Long et al. 2013; Thomas et al. 2014; Getirana 2016), which have shown it to be capable of detecting water storage deficits and trends related to drought and human consumption (e.g., Giroto et al. 2017). Recent studies using GRACE data have proven that assimilating GRACE-based TWS into land surface models (LSMs) has an important and lasting impact on modeled state and flux variables (e.g., Giroto et al. 2016; Kumar et al. 2016), while GRACE data themselves are enhanced through synthesis with high-resolution meteorological data constrained by LSM physics (Zaitchik et al. 2008). In terms of extreme events, GRACE data assimilation (DA) has been shown to improve identification of droughts in parts of North America (Houborg et al. 2012; Kumar et al. 2016), Europe (Li et al. 2012), Asia (Giroto et al. 2017) and globally (Li et al. 2019). Such studies demonstrated that GRACE-DA improves the simulations of water storage variability, in particular, groundwater.

A series of major drought events in the United States in the early 2000s made the development of an integrated

Corresponding author: Augusto Getirana, augusto.getirana@nasa.gov

drought early warning system for the United States a high priority (NIDIS 2004). The flagship product of that effort, the U.S. Drought Monitor (USDM; Svoboda et al. 2002), has become the nation's premier drought monitoring tool. However, the development of operational drought forecasts has been very limited. Notable exceptions are the NOAA Climate Prediction Center's (CPC) U.S. Seasonal and Monthly Drought Outlooks, which are based on precipitation and temperature outlooks and modeled soil moisture. While these are a good start toward a national drought early warning system, the outlooks are unable to forecast groundwater storage (GWS) and deep soil moisture, which is a serious shortcoming, considering the importance of those resources for people, plants, and ecosystems and their value as indicators of drought (van Lanen and Peters 2000; Thomas et al. 2014).

Numerous studies have explored the impact of initial hydrological conditions (IHCs) on seasonal forecasts (e.g., Shukla and Lettenmaier 2011; Li et al. 2009). They have shown that IHC uncertainties generally outweigh forcing uncertainties thus dominating forecast errors up to about 1-month lead time. In a recent study, Wanders et al. (2019) showed that IHCs, in particular groundwater, can explain at least half of the variance in hydrological forecasts at lead times of up to three months. At longer lead times, forcing uncertainties become a more important contributor. Because deep soil moisture and GWS slowly change in response to meteorological conditions due to their substantial "memory," they have great potential to inform seasonal predictions. Recognizing this potential, and building upon recent GRACE-DA applications to improve our understanding of physical processes related to water storage variability, as well as studies showing the importance of IHC for seasonal forecasts, the goal of this study is to evaluate that potential by assessing the impact of GRACE-DA on seasonal drought forecasts over the contiguous United States (CONUS). Our hypothesis is that assimilating GRACE into an offline LSM (i.e., not coupled to the atmosphere) results in more accurate water storage states, in particular deep soil moisture and groundwater, as represented in an LSM, and that these assimilated states will then improve the IHC of seasonal drought forecast simulations with the same model. To test our hypothesis, we assimilate GRACE-based TWS into an LSM, then use the updated states derived from GRACE-DA as the IHC for seasonal drought hindcasts (i.e., forecasts of past events, or historical forecasts) up to 90 lead days in the future using near surface meteorological data from two different established seasonal forecasting techniques. Hindcasts are evaluated in terms of improvement to forecasted groundwater,

when compared with those initialized by the reference simulation, also called the open loop (OL) simulation (i.e., no perturbation or assimilation applied).

2. Data and methods

GRACE-based TWS is assimilated into the Catchment land surface model (CLSM; Koster et al. 2000), which is a physically based numerical model that simulates the water and energy balance at and below the land surface in response to surface meteorology. Unlike models that simulate soil moisture content within a series of soil layers down to about 2-m depth, CLSM simulates equilibrium water storage within the unsaturated and saturated zone down to the bedrock, with a variable saturation curve that describes the vertical distribution. The equilibrium vertical distribution of soil moisture includes an implicit water table, located at the depth of the equilibrium saturation and implying the presence of an unconfined aquifer (groundwater) that has time-varying water storage. Bedrock depths were derived from the Second Global Soil Wetness project (GSWP-2; Dirmeyer and Oki 2002). In CLSM, TWS is defined as the sum of groundwater, soil moisture, snow water equivalent, and canopy interception.

We assimilated monthly gridded 0.5° GRACE-based TWS anomalies derived from the University of Texas at Austin's Center for Space Research (CSR) Mascon solution (Save et al. 2016). Even though the CSR GRACE mascons are natively estimated at 120-km-wide mascon blocks and resampled to a 0.5° grid, the actual resolution of its solution is about 250–300 km along the equator line. Mascon-based products have been shown to have a somewhat higher signal-to-noise ratio and reduced errors than products based on spherical harmonics (e.g., Rowlands et al. 2010; Save et al. 2016). GRACE data uncertainty has been estimated to be 1 cm equivalent water height, when averaged over areas larger than about $4 \times 10^5 \text{ km}^2$, and errors increase as the area observed decreases (Swenson et al. 2003). GRACE data were linearly interpolated to match the 0.125° model grid space and bias-corrected to match the mean TWS simulated in the OL run. GRACE data processing was performed with the NASA Land Surface Data Toolkit (LDT; Arsenault et al. 2018) and all model runs, as described below, were performed within the NASA Land Information System (LIS; Kumar et al. 2006).

a. Modeling configuration

As illustrated in the modeling flowchart provided in Fig. 1, we used two meteorological forcing datasets for the retrospective part of the analysis. One was the Princeton Meteorological Dataset (Sheffield et al. 2006),

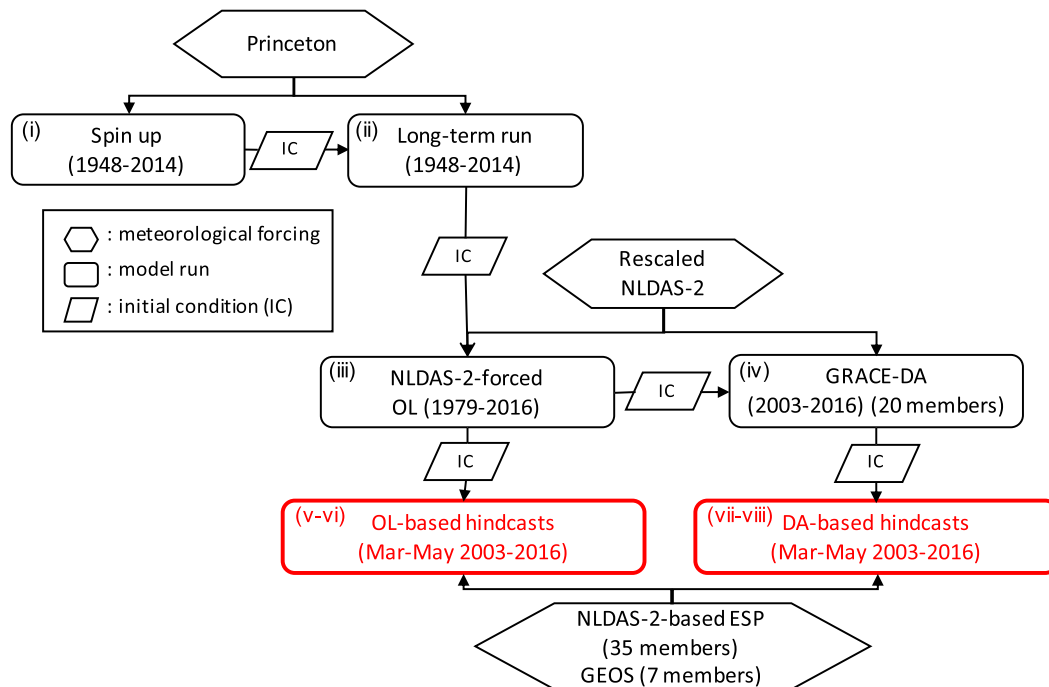


FIG. 1. Modeling flowchart. Experiments are ordered with roman numerals.

which is 1° gridded and spans from 1948 to 2014 with 3-hourly temporal resolution. The second was the North American Land Data Assimilation System phase 2 (NLDAS-2; Xia et al. 2012) dataset, which is 0.125° gridded and covers Central North America from 1979 to present, with hourly temporal resolution. NLDAS-2 was the primary dataset used in this study to force CLSM from 1979 to 2016. To compute drought indicator percentiles, for which a longer record is preferable, we generated a 67-yr climatology of soil moisture and GWS from a CLSM simulation forced by the Princeton dataset. CLSM was spun up using the Princeton dataset for the entirety of its period, and initial conditions for the subsequent 1948–2014 Princeton-forced long-term run were generated by averaging simulated states from the final 10 spinup years, as suggested by Rodell et al. (2005). We removed inconsistencies between the two datasets by scaling the NLDAS-2 meteorological variables (precipitation, air temperature and humidity, solar radiation, wind speed, and pressure) to match the Princeton monthly climatology using the overlapping period, from 1979 to 2014. The OL run was initialized from the Princeton-forced long-term run using rescaled NLDAS-2 data for the 1979–2016 period. In turn, the initial conditions for the 2003–16 GRACE-DA run were taken from the OL, and the rescaled NLDAS-2 forcing data were used again. Houborg et al. (2012) showed that, in order to better represent TWS dynamics and to accommodate

significant GWS decline during exceptional droughts, CLSM depth to bedrock should be increased beyond the model default values. In this study, bedrock depths were increased by three meters throughout the domain. All model runs were performed at a 15-min time step.

b. Data assimilation procedure

To deal with GRACE’s coarse spatial and temporal resolutions, a 3D-based ensemble Kalman smoother (EnKS; e.g., Evensen and Van Leeuwen 2000) approach, as described in Zaitchik et al. (2008), was used to assimilate GRACE-based TWS into CLSM. The EnKS, as applied here, includes a 2° spatial correlation window and a monthly temporal window within which two passes are performed: 1) the first pass integrates a forecast step to generate the ensemble of CLSM-based TWS state terms (no assimilation); and 2) the second pass performs the assimilation update based on the relative weights of the model estimates and observations (i.e., in terms of their error covariance matrices), which are determined by the Kalman gain matrix. In the first pass, the CLSM TWS-based states are stored on the 5th, 15th, and 25th of each month (approximately related to the overpass frequency of GRACE). In the second pass, the ensemble is reinitialized, and the monthly analysis increments are applied evenly across the month. Two CLSM prognostic soil moisture variables are perturbed with normally distributed additive perturbations, and

three snow water equivalent state layers are perturbed with a lognormal, mean 1, multiplicative perturbation. We assumed that the contribution of surface water storage (SWS) variability to TWS over the United States is negligible (Rodell and Famiglietti 2001). Getirana et al. (2017) concluded that this is true for the most part, however, in certain parts of the United States such as Minnesota and southern Florida surface water storage changes are a substantial component of TWS changes. The effect of not modeling surface water in CLSM is that TWS changes (the residual of the water budget or from GRACE-DA) that would be attributed to SWS in the real world are, in effect, integrated into and simulated as all other TWS components, that is, snow, GWS and soil moisture. While this is not ideal, for the purposes of this study, whose primary goal is to improve drought/wetness forecasts, it is sufficient. Perturbations were also applied to three of the meteorological forcing fields: incoming longwave radiation (additive type), and incoming shortwave radiation and precipitation (both multiplicative perturbation type).

For the TWS observational standard error covariance, we applied a spatially uniform scalar value of 10mm (Zaitchik et al. 2008). GRACE-DA was performed with an ensemble of 20 members. Previous studies have shown that this ensemble size is enough to represent model uncertainty within a GRACE-DA framework (Zaitchik et al. 2008; Li et al. 2012; Kumar et al. 2016). It is important to note that our 3D-based EnKS does not account for the inherently low spatial resolution of the GRACE products used in this study. Our data assimilation simulations cannot, at this time, completely distill the spatially smooth GRACE TWS's estimates to capture all local variability. More details on the GRACE-DA configuration and specific perturbation settings used in this study can be found in Kumar et al. (2016).

c. Seasonal forecasts

Daily hydrological hindcasts with up to 90 lead days were performed using near surface meteorological data from two data sources to force the model: NASA's GEOS Seasonal-to-Interannual Forecast System (Borovikov et al. 2019) version 1 and the Ensemble Streamflow Prediction (ESP; Day 1985). The purpose of including GEOS and ESP was to determine the applicability and robustness of using GRACE-DA for seasonal groundwater forecasts combined with different atmospheric forecast techniques. GEOS is a dynamically based seasonal forecast system composed of multiple members designed to generate skillful meteorological predictions. The GEOS hindcast ensemble

used in this study consists of its first seven members. ESP is a statistical technique based on historical meteorological data, and is intended to provide a "null" atmospheric forecast, in which the ensemble of meteorological fields represents a probability-weighted sampling from the historic record. NLDAS-2 historical meteorological data, from 1982 to 2016 (excluding the particular hindcasted year), are used to generate ESP hindcasts, consisting of a 35-member ensemble. Hindcasts were initialized every 1 March and May over the 2003–16 period using states from the OL and GRACE-DA runs as the IHCs. Both months were selected because in much of the United States 1) they represent the end of the wet period, 2) they are near the start of the growing season, and 3) March is an important starting point for streamflow and flood predictions. GRACE-DA-based IHCs were generated by first averaging the 20-member DA ensemble, then using that average to initialize each member used in each forecast technique. This means that, for each hindcast run, all members were initialized with the same IHC. IHCs are composed of water storage components (in the surface and in the different soil layers) and soil temperature. Hindcasts were evaluated by deterministic means, that is, using the forecasted ensemble means. Briefly, as illustrated in Fig. 1, a total of eight model runs were performed: (i) spin up (1948–2014) and (ii) long-term (1948–2014) runs forced with the Princeton meteorological dataset; (iii) OL (1979–2016) and (iv) GRACE-DA (2003–16) runs forced with rescaled NLDAS-2 meteorological dataset; and NLDAS-2-based ESP and GEOS hindcasts initialized with states from the (v) and (vi) OL and (vii) and (viii) GRACE-DA runs.

d. Groundwater observations

The impact of GRACE-DA-based IHC on seasonal forecasts was evaluated at 305 wells drawn from records of daily measurements of depth-to-water-table were available. These data were collected and provided by the U.S. Geological Survey (USGS) and the Illinois State Water Survey (<http://www.isws.illinois.edu/warm>). The 305 well locations used here were culled from a larger dataset following the criteria that the wells must be installed in unconfined aquifers and not directly affected by pumping or injections (Rodell et al. 2007; Houborg et al. 2012; Girotto et al. 2016; Kumar et al. 2016). In addition to using available metadata, relevant literature, geological maps, and satellite imagery to make these determinations, we chose groundwater time series that displayed a clear seasonal cycle (which is often not present in confined aquifer time series) and lacked sudden water table declines that

might be associated with pumping (Giroto et al. 2016). Specific yield values identified in previous studies (Rodell et al. 2007; Houborg et al. 2012; Li and Rodell 2015; Giroto et al. 2016; Kumar et al. 2016) were used to convert the depth to water to equivalent height of water, in millimeters, which can be directly compared to simulated groundwater storage.

e. Evaluation procedure

The accuracy of simulated GWS time series is quantified using the temporal correlation r and the root-mean-square error (RMSE) between simulation s and observation o . RMSE is defined as follows:

$$RMSE = \left[\frac{\sum_{t=1}^{nt} (s_t - o_t)^2}{nt} \right]^{1/2}, \quad (1)$$

where t is the time step and nt is the period length.

Following Kumar et al. (2014), we used the normalized information contribution (NIC) metric applied to the RMSE and r in order to determine any improvement that GRACE-DA-based IHC contributes to seasonal GWS hindcasts compared to those initialized with the OL run. Their respective NIC values are defined below:

$$RMSE_{NIC} = \frac{(RMSE_{OL} - RMSE_{DA})}{RMSE_{OL}}, \quad (2)$$

$$r_{NIC} = \frac{(r_{DA} - r_{OL})}{(1 - r_{OL})}. \quad (3)$$

Both metrics range from $-\infty$ to 1, where values above zero indicate improvement, below zero indicates degradation, and zero means no added skill.

Drought detection skills were spatially evaluated using the probability of detection (POD), false alarm rate (FAR), and spatial correlation r_s . POD and FAR are defined as

$$POD = \frac{a}{a + c}, \quad (4)$$

$$FAR = \frac{b}{b + d}, \quad (5)$$

where a is the number of drought pixels correctly detected by forecasts, b stands for the number of false alarms (drought pixels detected but not observed), c is the number of drought pixels not detected, and d is the sum of pixels when neither observations nor forecasts occurred.

3. Results and discussion

a. GRACE-DA impact on groundwater variability

The standard deviation ratio, translating changes in average GWS amplitudes simulated by both GRACE-DA and OL, shows impacts of data assimilation on groundwater variability, as shown in Fig. 2. Most changes are due to lower TWS amplitudes observed by GRACE, as a result of a coarser spatial resolution that smoothens out high variability at finer scales. Decreased amplitudes are mostly noticeable in the East, the South, and parts of the Midwest. Increased amplitudes are observed in parts of the western United States, Texas, and the Great Plains.

b. Added skill with GRACE-DA-based initialization

Figures 3 and 4 map and summarize the differences in groundwater hindcast skill between the GRACE-DA and OL initialized experiments, evaluated at the 305 well locations. This is illustrated in terms of normalized improvements in RMSE ($RMSE_{NIC}$, in Fig. 3) and correlation coefficient r (r_{NIC} , in Fig. 4) for 3-month hindcasts initialized on 1 March and 1 May for the 2003–16 period. An overall improvement is observed at a large majority of wells where in situ observations are available. Improvements in RMSE values for ESP and GEOS hindcasts ranged between 74% and 77% of wells when initialized with GRACE-DA. Half of the wells had an $RMSE_{NIC}$ equal or above 0.19–0.25, depending on the forecast technique. Correlation generally improves, but at a lower rate (58%–60% of wells), with a median improvement of 0.03–0.09. As one can see in the figure, RMSE improves quite homogeneously throughout the United States, but r improvements are inconsistent, and some locations show substantial degradation.

To further investigate how GRACE-DA-based IHC impact hindcasts over the United States, we divided the domain into six regions, as delineated in the maps in Figs. 3 and 4. The regions are: Northeast, Southeast, Midwest, Great Plains, Northwest, and Southwest. Scatterplots in Figs. 3 and 4 show median values of $RMSE_{NIC}$ and r_{NIC} (x axis) as a function of percentage of locations with improved skill (y axis) within each region for hindcasts initialized on 1 March and 1 May. Hence, a symbol representing a region with overall improved skill appears in the upper-right quadrant. Regions located in other quadrants have either negative median metrics, or fewer than half the wells improved, or both. All hindcast experiments (i.e., both forecast techniques and initialization months) present $RMSE_{NIC}$ values located in the upper-right quadrant for all regions, representing overall improvement in

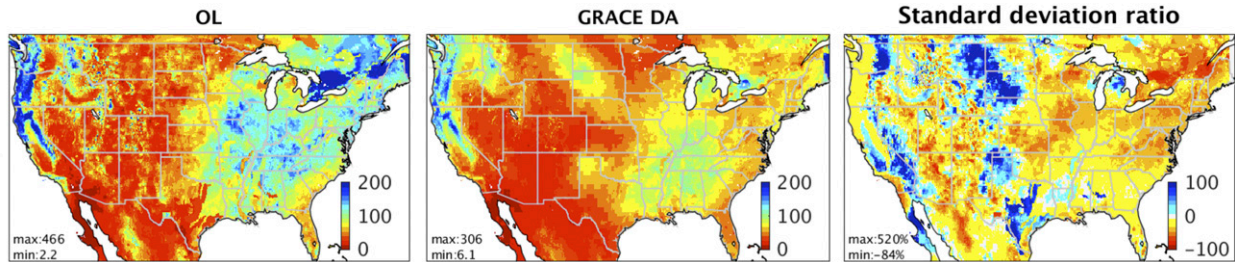


FIG. 2. Groundwater storage standard deviation (mm) derived from CLSM open loop (OL) and GRACE data assimilation (GRACE-DA) runs for the 2003–16 period and the standard deviation ratio (%), represented as $[(\text{GRACE-DA}/\text{OL}) - 1] \times 100$.

seasonal groundwater hindcasts. The exception is the Great Plains, where little or no improvement was shown, as evidenced by the low r_{NIC} values. A likely explanation is that groundwater pumping to support irrigated agriculture, which is widespread in the Great Plains but not represented in the model, exerts substantial control over the water table variations.

We note that DA improves RMSE relative to OL to a much larger extent than it improves correlation. This is because the range of TWS variability is conserved in GRACE observations, resulting in significant DA updates to the simulated TWS amplitudes and trends. This benefit could, conceivably, be achieved through other methods with the aid of GRACE data, such as model calibration or rescaling. Rescaling techniques such as cumulative distribution function (CDF) matching between the simulated and observed TWS have previously

been used with GRACE-DA (e.g., Girotto et al. 2016). The key motivation for using that approach was to preserve the soil moisture climatology of the model, so that coupled land–atmosphere applications are not impacted by climatological changes to soil moisture and land–atmosphere fluxes. The downside of such CDF matching is that the utility from assimilation is limited to the corrections to temporal anomaly information, in other words, the assimilated observations are not allowed to improve the range of variability of the simulated soil moisture or groundwater, which could be important for water resources and other applications. Indeed, prior studies (Draper et al. 2009; Kumar et al. 2015; Lee and Im 2015) have shown that this approach severely limits the potential benefit gained from assimilating soil moisture, not to mention the fact that it eliminates the potential of DA to

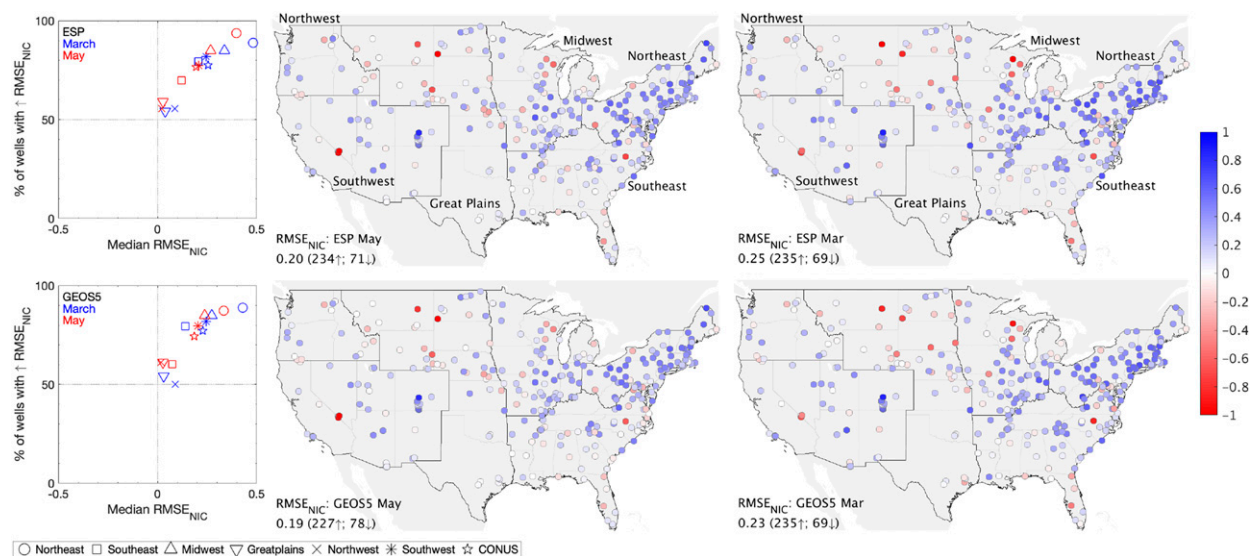


FIG. 3. GRACE-DA impact on the seasonal hindcast initialization using surface meteorological data from ESP and GEOS historical forecasts. (right) Spatial distribution of RMSE_{NIC} values derived from daily 3-month groundwater hindcasts initialized in May and March over the 2003–16 period. At the bottom left of each map, the median value and number of wells with improved (\uparrow) and degraded (\downarrow) metrics. (left) Median (x axis) and percentage of wells with positive NIC values (y axis) computed for daily 3-month hindcasts initialized in May and March for each U.S. region (Northeast, Southeast, Midwest, Great Plains, Northwest, and Southwest).

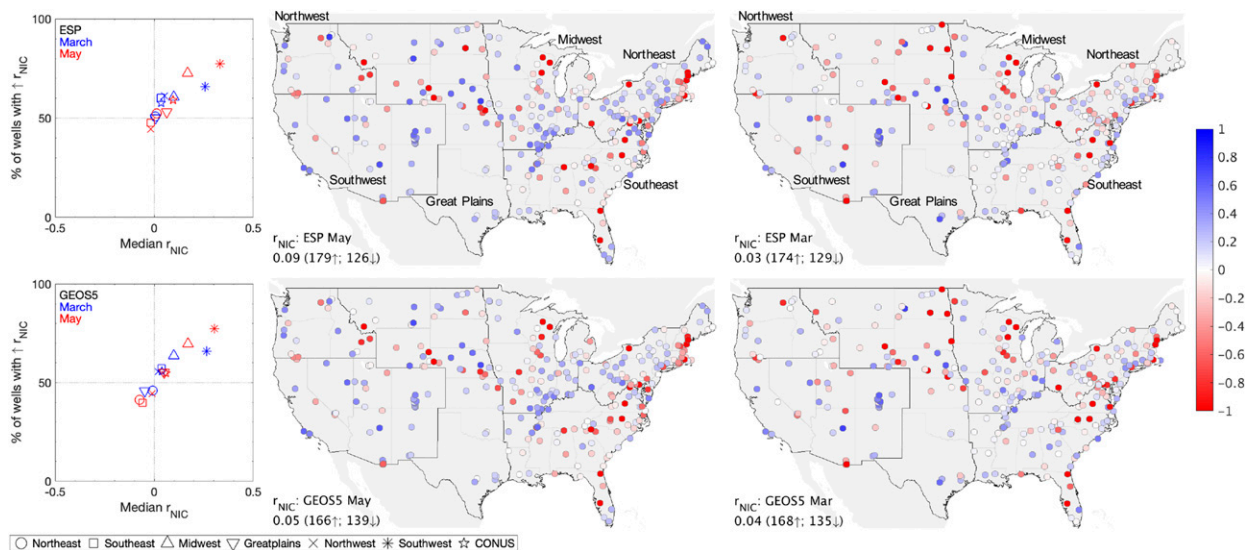


FIG. 4. As in Fig. 3, but for r_{NIC} values.

diagnose, quantify and correct biases. In addition, CDF matching was also shown to introduce large statistical errors when unmodeled features are present in the observations and if sufficient temporal specificity is not included in the computation of the CDFs (Kumar et al. 2015; Yin and Zhan 2018). While CDF scaling is commonly used as an a priori bias correction approach, these limitations highlight that it is not necessarily a technically more correct approach than the method we use. Further, other studies (Zaitchik et al. 2008; Houborg et al. 2012; Li et al. 2012, 2019; Kumar et al. 2016) have established that our is valid and is useful in developing meaningful and physically consistent improvements from the assimilation of GRACE observations.

In this application, groundwater trends and variability are significantly improved in all U.S. regions. Figure 5 plots the OL and DA groundwater time series along with hindcasts initialized on 1 May from both of those runs and using the original and ESP-downscaled GEOS forecast meteorological forcing. The largest changes owing to DA are seen in the Northeast, Midwest, and the Great Plains, while changes are smaller or less consistent in the other three regions. The types of changes include increases or decreases in amplitude and/or increases or decreases in the intensity of the wet or dry extremes. For example, DA generally increases the amplitude of both seasonal and interannual groundwater variability in the Great Plains, while it tempers the extremes in the Northeast. In all cases the hindcasts initialized by the DA simulation remain close to the DA simulation throughout the three months, which is positive, given that the DA

simulation is the closest thing we have to “truth,” though not surprising.

c. Groundwater memory in seasonal forecasts

We also examined $RMSE_{NIC}$ and r_{NIC} values for individual lead months and regions, for each forecast technique and initialization month, as shown in Fig. 6. In general, the benefits of GRACE-DA-based IHC decline from hindcast lead month zero to one and from one to two, particularly in regions with high $RMSE$ or correlation values. That makes intuitive sense, because the groundwater information imparted by GRACE-DA fades over time, causing the hindcasts initialized with GRACE-DA and OL to converge. However, groundwater’s substantial memory allows GRACE-DA initialization to improve the hindcasts for at least three months. Both $RMSE_{NIC}$ and r_{NIC} values are positive at the end of the third lead month for most regions. Once again, little skill is seen in the Great Plains region for all forecast techniques, metrics and lead months. The Northeast, Southeast, and Northwest regions display positive $RMSE_{NIC}$, but negative r_{NIC} in some scattered months. This could be a result of insufficient frequency of observations (at some wells in these regions there is only one observation per month) or a systematic bias related to the simplified groundwater dynamics used in CLSM.

d. Drought severity forecast skill

Finally, we evaluated how GRACE-DA-based IHC impacts forecasts of drought severity indexes in terms of improvements in detecting extreme droughts relative to the USDM reference. USDM drought severity index

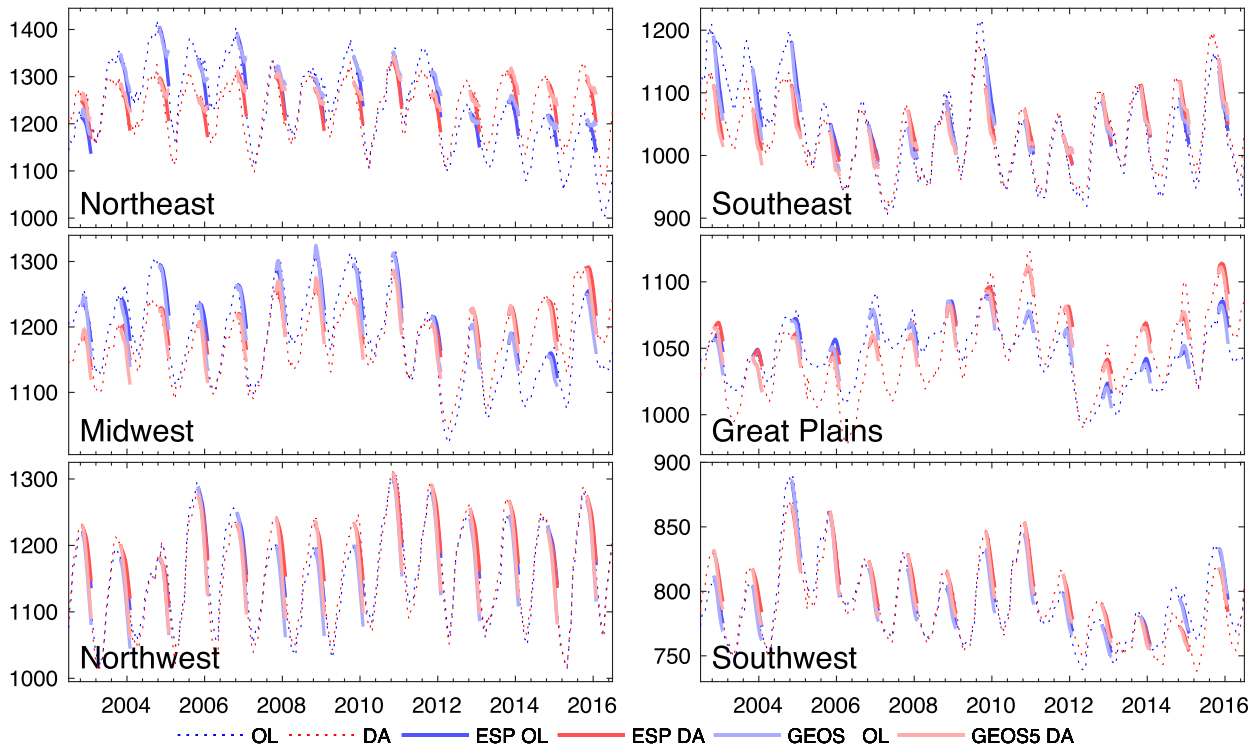


FIG. 5. Daily groundwater variability averaged for six U.S. regions. Dotted lines represent OL and GRACE-DA retrospective runs, and bold lines are 3-month ESP and GEOS hindcasts initialized on 1 May.

maps are drawn by a team of authors using observation and model-based information on precipitation, temperature, soil moisture, surface water, snowpack, vegetation, and other indicators as well as drought impact reports. Outputs from a similar GRACE-DA system are also made available for the generation of the USDM maps and, as a consequence, this is not a completely independent evaluation. USDM maps droughts in five categories of varying intensity: D_0 (abnormally dry, percentile $\leq 30\%$), D_1 (moderate drought, percentile $\leq 20\%$), D_2 (severe drought, percentile $\leq 10\%$), D_3 (extreme drought, percentile $\leq 5\%$), and D_4 (exceptional drought, percentile $\leq 2\%$). Maps are published online each week, typically on Wednesdays, and represent drought conditions from one to two days prior to the release date. Drought severity maps, following the same USDM categories, were derived from our model runs using the climatology derived from the Princeton-based long-term run.

Figure 7 shows drought intensity maps at the end of three consecutive months (May, June, and July) in 2014, during which a major drought was developing in California and parts of Nevada, and another one weakening in the southern United States, covering parts of Texas, Oklahoma, Colorado, and Kansas, as reported by USDM. Because USDM maps are largely based on

precipitation and soil moisture indexes, they are not perfectly analogous to our GRACE-DA-based groundwater drought indexes perfectly match with them, but there should be large scale, general agreement. Although the OL is capable of detecting both of the D_4 exceptional droughts, it overestimates their extents throughout the Northwest. It also shows an extended drought of intensities varying from D_0 to D_3 over a large portion of the central and northeastern United States. Also, the OL mistakenly shows the intensification of a large exceptional drought taking place in the central United States, not reported by USDM. On the other hand, assimilating GRACE data results in a much better agreement with USDM, when compared to OL, eliminating both unreported exceptional droughts in the central and northeastern United States. That improvement directly impacts the forecasted drought maps. Both ESP and GEOS are positively affected by the GRACE-DA-based initialization, with impacts observed throughout the 3-month hindcasts. However, one can notice that GEOS diverges at the end of the second and third months, showing an expanding exceptional drought in the Northwest, which could be a result of an inaccurate meteorological forecast in that year, or a product of climatological drift in the GEOS forecast system. Simulation drift is often observed in seasonal forecasts,

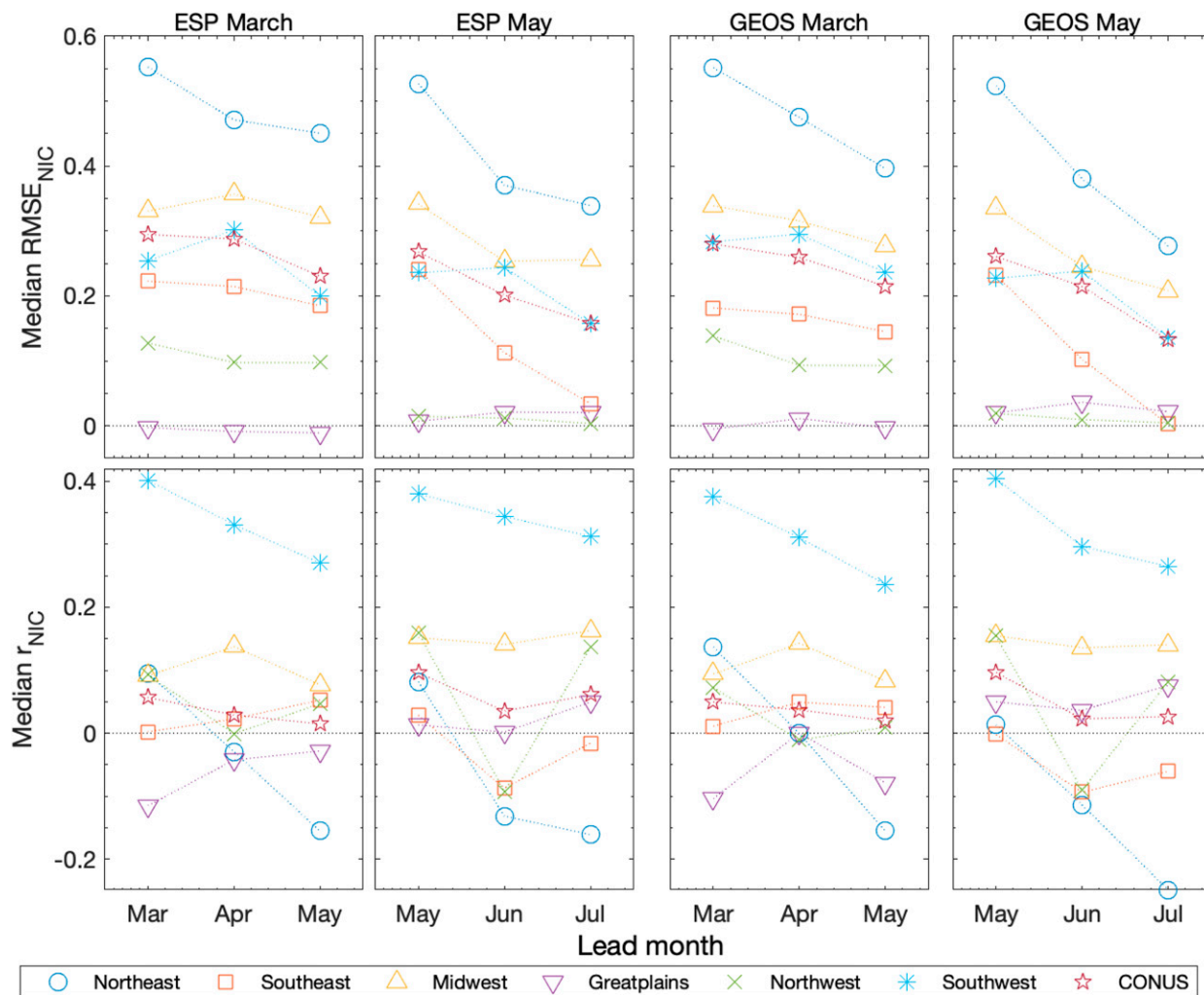


FIG. 6. Impact of initializing groundwater storage seasonal forecasts with GRACE-DA, for each lead month and region, for the 2003–16 period. The impact is defined as the normalized information contribution (NIC) of the RMSE and correlation r .

and can be a product of model parameterizations, integrating errors in atmospheric dynamics, ocean response to an error in initial atmospheric conditions, or other sources (e.g., Smith et al. 2013; Hermanson et al. 2018).

Drought detection skill was quantitatively evaluated for the three dates shown in Fig. 7 using POD, FAR, and r_s . Extreme (D_3) and exceptional (D_4) drought events combined, corresponding to the 5th percentile, were used as threshold for computing POD and FAR. Spatial correlation was computed using all drought severity indexes. GRACE-DA-based hindcasts show a positive impact on spatial correlation for all three dates, with an r_s increase varying from 0.14 to 0.23, and averaging 0.18 for all three dates. The probability of detection of extreme and exceptional drought events (indices D_3 and D_4) slightly degrades with both forecast techniques at the end of the first month, but improves later on, with

an average POD improvement of 0.03. Hindcasts initialized with GRACE-DA show little reduction in false alarms. Figure 8 summarizes the improvements in skill detection for the three dates. For comparison reasons, results from the retrospective runs (GRACE-DA – OL) are also provided.

4. Final discussion

Herein, we leverage the use of the information contained in GRACE to improve IHCs for seasonal forecasts through data assimilation techniques. Our hypothesis was that 0.125° gridded groundwater storage states resulting from this new GRACE-DA-based enhanced prediction system would better match observations, thus enhancing the skill and the potential applications value of seasonal drought forecasts.

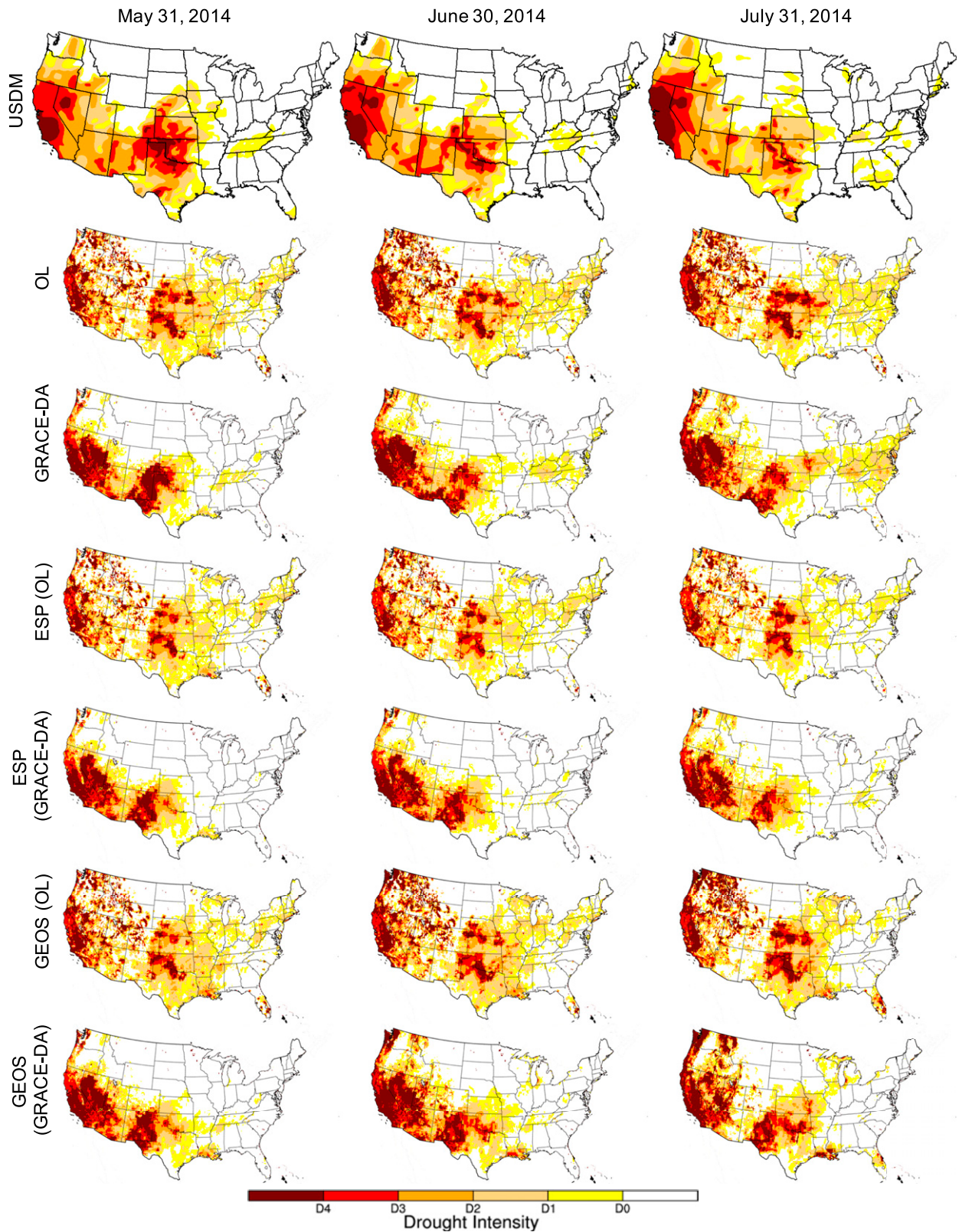


FIG. 7. U.S. drought intensity maps during major droughts in the western and southern United States between May and July 2014. Shown from top to bottom are U.S. Drought Monitor, retrospective OL and GRACE-DA runs, and ESP and GEOS hindcasts initialized with the 1 May GRACE-DA states. USDM estimates correspond to drought intensity on 3 Jun, 1 Jul, and 5 Aug, and the other experiments are snapshots at the dates displayed on the top of each column.

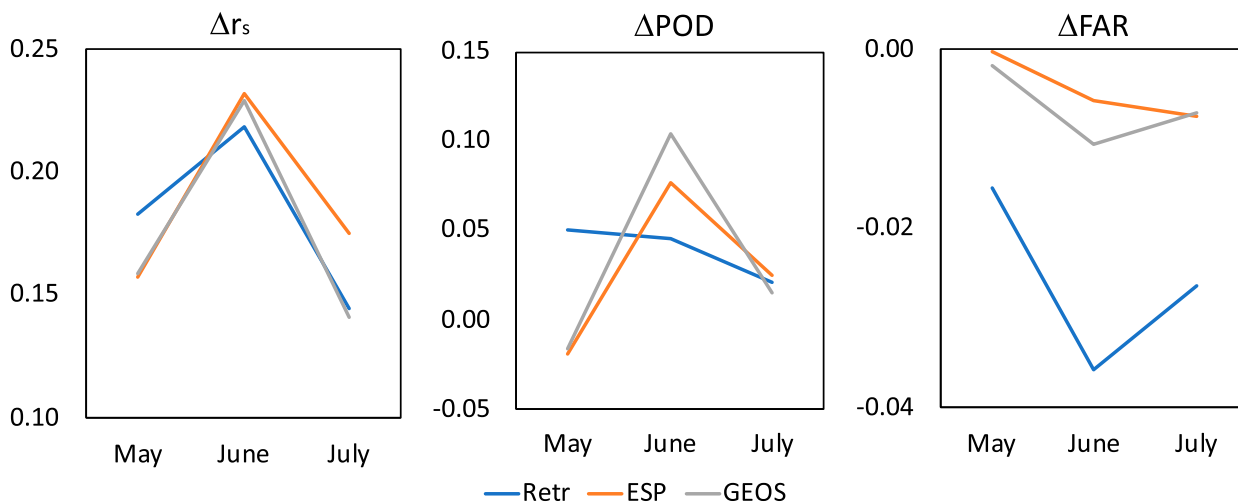


FIG. 8. Differential spatial correlation (Δr_s), probability of detection (ΔPOD) and false alarm rate (ΔFAR) of drought severity index maps for three dates in 2014, as shown in Fig. 7. POD and FAR thresholds were defined as the extreme (D_3) and exceptional (D_4) drought events combined, corresponding to the 5th percentile. Values correspond to the improvement in skills when ESP and GEOS forecast techniques are initialized with GRACE-DA (i.e., GRACE-DA – OL). Differential values for the retrospective runs (Retr) are also shown for comparison purposes. Improvement is represented by positive values for r_s and POD, and negative values for FAR.

Our results confirm this hypothesis, as quantified in terms of improvements in both seasonal forecast skill metrics at hundreds of wells throughout the United States and spatial agreement with the U.S. Drought Monitor. We also found that a GRACE-DA-based IHC improves both forecast techniques (ESP and GEOS), which could be evidence that the skill added by GRACE-DA, as observed in this study, could be generalized for a wider range of forecast techniques.

We acknowledge that neglecting anthropogenic impacts, such as irrigation, is a major limitation in the GRACE-DA scheme applied to CLSM in this study. Such activities may cause major changes in water cycle components, in particular, soil moisture and GWS variability, which are observed by GRACE, but not represented by the model (Giroto et al. 2017). Although we have made a significant effort to remove wells impacted by water pumping in order to perform an unbiased evaluation, we recognize that some groundwater observations might not have been filtered out. This is likely one of the causes for the poor results in the Great Plains, where groundwater is the major source for irrigation. Nie et al. (2018) have shown that accounting for groundwater pumping in the Great Plains improves the representation of TWS variability by LSMs and agreement with GRACE. Another plausible explanation for these differences in irrigated areas is that GRACE-DA might not properly attribute mass change to the different soil layers (i.e., surface soil moisture, root-zone soil moisture and groundwater). That issue could be addressed with multisensor DA, taking into account soil

moisture (e.g., Giroto et al. 2019). Neglecting surface water storage variability is another limiting factor in our GRACE-DA scheme. For smaller rivers, groundwater and surface water can be treated as a single TWS component, since surface water occurs where the water table intersects the land surface (Winter et al. 1998; Rodell et al. 2007). As demonstrated by Getirana et al. (2017), SWS may impact TWS in the tropics and large rivers flowing in high latitudes and arid regions. Although the authors show little SWS impact on TWS variability in the United States, regions located near major rivers, such as the Mississippi River, have a nonnegligible impact from river dynamics. In this sense, further investigation considering surface water in a GRACE-DA application is required.

Although the long latency was an issue for generating GRACE Level-3 products (typically 2–5-month latency), one of our assumptions is that the long memory of groundwater and deep soil moisture allows GRACE-DA to be used to generate improved initial conditions for seasonal forecasts. The longer memory is demonstrated by the nonzero GRACE-DA impact on groundwater hindcasts at the end of the third lead month of our experiments, as shown in Fig. 6. With GRACE Follow On (GRACE-FO) data available from May 2018 to present and the development of new low-latency GRACE hydrology products (Sakumura et al. 2016) that would make TWS estimates available within 2–6 weeks from the raw data acquisition, we should be able to overcome the long latency issue. The GRACE-FO-based extreme event monitoring and forecast system is underway and

will provide seasonal forecasts operationally (i.e., once a month), taking advantage of these low latency TWS products. We expect that these advances will have immediate impact on continuing and future work developed by U.S. institutions, such as the Army Corps of Engineers, the National Drought Mitigation Center, the U.S. Drought Monitor, and NOAA's National Weather Service, being used as an additional source of information for drought monitoring and seasonal hydrological forecasts.

Acknowledgments. This study was funded by the NASA's Applied Sciences Program–Water Resources, and the GRACE and GRACE Follow On Science Team. We thank Manuela Giroto for her support in acquiring USGS groundwater data (<https://water.usgs.gov/ogw/>) and Himanshu Save for making GRACE data available. The NLDAS-2 meteorological dataset was acquired as part of the activities of NASA's Science Mission Directorate and is archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC) on <https://earthdata.nasa.gov/about/daacs/daac-ges-disc>. LIS is open source and available on <https://lis.gsfc.nasa.gov>.

REFERENCES

- Arsenault, K. R., and Coauthors, 2018: The Land surface Data Toolkit (LDT v7.2) – A data fusion environment for land data assimilation systems. *Geosci. Model Dev.*, **11**, 3605–3621, <https://doi.org/10.5194/gmd-11-3605-2018>.
- Borovikov, A., and Coauthors, 2019: GEOS-5 seasonal forecast system. *Climate Dyn.*, **53**, 7335–7361, <https://doi.org/10.1007/s00382-017-3835-2>.
- Day, G., 1985: Extended streamflow forecasting using NWSRFS. *J. Water Resour. Plann. Manage.*, **111**, 157–170, [https://doi.org/10.1061/\(ASCE\)0733-9496\(1985\)111:2\(157\)](https://doi.org/10.1061/(ASCE)0733-9496(1985)111:2(157)).
- Dirmeyer, P., and T. Oki, 2002: The Second Global Soil Wetness Project (GSWP-2) Science and Implementation Plan. IGPO Publication Series 37, 64 pp.
- Draper, C., J.-F. Mahfouf, and J. Walker, 2009: An EKF assimilation of AMSR-E soil moisture into the ISBA land surface scheme. *J. Geophys. Res.*, **114**, D20104, <https://doi.org/10.1029/2008JD011650>.
- Evensen, G., and P. J. van Leeuwen, 2000: An ensemble Kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.*, **128**, 1852–1867, [https://doi.org/10.1175/1520-0493\(2000\)128<1852:AEKSFN>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2).
- Famiglietti, J. S., and M. Rodell, 2013: Water in the balance. *Science*, **340**, 1300–1301, <https://doi.org/10.1126/science.1236460>.
- Getirana, A., 2016: Extreme water deficit in Brazil detected from space. *J. Hydrometeorol.*, **17**, 591–599, <https://doi.org/10.1175/JHM-D-15-0096.1>.
- , S. Kumar, M. Giroto, and M. Rodell, 2017: Rivers and floodplains as key components of global terrestrial water storage variability. *Geophys. Res. Lett.*, **44**, 10 359–10 368, <https://doi.org/10.1002/2017GL074684>.
- Giroto, M., G. J. M. De Lannoy, R. H. Reichle, and M. Rodell, 2016: Assimilation of gridded terrestrial water storage observations from GRACE into a land surface model. *Water Resour. Res.*, **52**, 4164–4183, <https://doi.org/10.1002/2015WR018417>.
- , —, —, —, C. Draper, S. N. Bhanja, and A. Mukherjee, 2017: Benefits and pitfalls of GRACE data assimilation: A case study of terrestrial water storage depletion in India. *Geophys. Res. Lett.*, **44**, 4107–4115, <https://doi.org/10.1002/2017GL072994>.
- , R. H. Reichle, M. Rodell, Q. Liu, S. Mahanama, and G. J. De Lannoy, 2019: Multi-sensor assimilation of SMOS brightness temperature and GRACE terrestrial water storage observations for soil moisture and shallow groundwater estimation. *Remote Sens. Environ.*, **227**, 12–27, <https://doi.org/10.1016/j.rse.2019.04.001>.
- Hermanson, L., and Coauthors, 2018: Different types of drifts in two seasonal forecast systems and their dependence on ENSO. *Climate Dyn.*, **51**, 1411–1426, <https://doi.org/10.1007/s00382-017-3962-9>.
- Houborg, R., M. Rodell, B. Li, R. Reichle, and B. F. Zaitchik, 2012: Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations. *Water Resour. Res.*, **48**, W07525, <https://doi.org/10.1029/2011WR011291>.
- Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2000: A catchment based approach to modeling land surface processes in a general circulation model: 1. Model structure. *J. Geophys. Res.*, **105**, 24 809–24 822, <https://doi.org/10.1029/2000JD900327>.
- Kumar, S. V., and Coauthors, 2006: LIS – An interoperable framework for high resolution land surface modeling. *Environ. Modell. Software*, **21**, 1402–1415, <https://doi.org/10.1016/j.envsoft.2005.07.004>.
- , and Coauthors, 2014: Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation. *J. Hydrometeorol.*, **15**, 2446–2469, <https://doi.org/10.1175/JHM-D-13-0132.1>.
- , C. Peters-Lidard, J. Santanello, R. Reichle, C. Draper, R. Koster, G. Nearing, and M. Jasinski, 2015: Evaluating the utility of satellite soil moisture retrievals over irrigated areas and the ability of land data assimilation methods to correct for unmodeled processes. *Hydrol. Earth Syst. Sci.*, **19**, 4463–4478, <https://doi.org/10.5194/hess-19-4463-2015>.
- , and Coauthors, 2016: Assimilation of gridded GRACE terrestrial water storage estimates in the North American Land Data Assimilation System. *J. Hydrometeorol.*, **17**, 1951–1972, <https://doi.org/10.1175/JHM-D-15-0157.1>.
- Lee, J. H., and J. Im, 2015: A novel bias correction method for Soil Moisture and Ocean Salinity (SMOS) soil moisture: Retrieval ensembles. *Remote Sens.*, **7**, 16 045–16 061, <https://doi.org/10.3390/rs71215824>.
- Li, B., and M. Rodell, 2015: Evaluation of a model-based groundwater drought indicator in the conterminous U.S. *J. Hydrol.*, **526**, 78–88, <https://doi.org/10.1016/j.jhydrol.2014.09.027>.
- , —, B. F. Zaitchik, R. H. Reichle, R. D. Koster, and T. M. van Dam, 2012: Assimilation of GRACE terrestrial water storage into a land surface model: Evaluation and potential value for drought monitoring in western and central Europe. *J. Hydrol.*, **446–447**, 103–115, <https://doi.org/10.1016/j.jhydrol.2012.04.035>.
- , and Coauthors, 2019: Global GRACE data assimilation for groundwater and drought monitoring: Advances and challenges. *Water Resour. Res.*, **55**, 7564–7586, <https://doi.org/10.1029/2018WR024618>.
- Li, H., L. Luo, E. F. Wood, and J. Schaake, 2009: The role of initial conditions and forcing uncertainties in seasonal hydrologic

- forecasting. *J. Geophys. Res.*, **114**, D04114, <https://doi.org/10.1029/2008JD010969>.
- Long, D., B. R. Scanlon, L. Longuevergne, A. Y. Sun, D. N. Fernando, and H. Save, 2013: GRACE satellite monitoring of large depletion in water storage in response to the 2011 drought in Texas. *Geophys. Res. Lett.*, **40**, 3395–3401, <https://doi.org/10.1002/grl.50655>.
- NCDC, 2014: Billion dollar U.S. weather disasters. Accessed 7 January 2020, <https://www.ncdc.noaa.gov/billions/overview>.
- NIDIS, 2004: Creating a drought early warning system for the 21st century: The National Integrated Drought Information System (NIDIS). Western Governor's Association Rep., 13 pp., https://www.drought.gov/drought/sites/drought.gov.drought/files/media/whatisnidis/Documents/200406_WGA_NIDIS_Report.pdf.
- Nie, W., B. F. Zaitchik, M. Rodell, S. V. Kumar, M. C. Anderson, and C. Hain, 2018: Groundwater withdrawals under drought: Reconciling GRACE and Land Surface Models in the United States High Plains Aquifer. *Water Resour. Res.*, **54**, 5282–5299, <https://doi.org/10.1029/2017WR022178>.
- Rodell, M., and J. S. Famiglietti, 2001: An analysis of terrestrial water storage variations in Illinois with implications for the Gravity Recovery and Climate Experiment (GRACE). *Water Resour. Res.*, **37**, 1327–1339, <https://doi.org/10.1029/2000WR900306>.
- , P. R. Houser, A. A. Berg, and J. S. Famiglietti, 2005: Evaluation of 10 methods for initializing a land surface model. *J. Hydrometeor.*, **6**, 146–155, <https://doi.org/10.1175/JHM414.1>.
- , J. Chen, H. Kato, J. S. Famiglietti, J. Nigro, and C. R. Wilson, 2007: Estimating ground water storage changes in the Mississippi River basin (USA) using GRACE. *Hydrogeol. J.*, **15**, 159–166, <https://doi.org/10.1007/s10040-006-0103-7>.
- , J. S. Famiglietti, D. N. Wiese, J. T. Reager, H. K. Beaudoin, F. W. Landerer, and M. H. Lo, 2018: Emerging trends in global freshwater availability. *Nature*, **37**, 1327, <https://doi.org/10.1038/s41586-018-0123-1>.
- Rowlands, D. D., S. B. Luthcke, J. J. McCarthy, S. M. Klosko, D. S. Chinn, F. G. Lemoine, J.-P. Boy, and T. J. Sabaka, 2010: Global mass flux solutions from GRACE: A comparison of parameter estimation strategies—Mass concentrations versus Stokes coefficients. *J. Geophys. Res.*, **115**, B01403, <https://doi.org/10.1029/2009JB006546>.
- Sakumura, C., S. Bettapur, H. Save, and C. McCullough, 2016: High-frequency terrestrial water storage signal capture via a regularized sliding window mascon product from GRACE. *J. Geophys. Res. Solid Earth*, **121**, 4014–4030, <https://doi.org/10.1002/2016JB012843>.
- Save, H., S. Bettapur, and B. D. Tapley, 2016: High resolution CSR GRACE RL05 mascons. *J. Geophys. Res. Solid Earth*, **121**, 7547–7569, <https://doi.org/10.1002/2016JB013007>.
- Sheffield, J., G. Goteti, and E. F. Wood, 2006: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Climate*, **19**, 3088–3111, <https://doi.org/10.1175/JCLI3790.1>.
- Shukla, S., and D. P. Lettenmaier, 2011: Seasonal hydrologic prediction in the United States: understanding the role of initial hydrologic conditions and seasonal climate forecast skill. *Hydrol. Earth Syst. Sci.*, **15**, 3529–3538, <https://doi.org/10.5194/hess-15-3529-2011>.
- Smith, D. M., and Coauthors, 2013: Real-time multi-model decadal climate predictions. *Climate Dyn.*, **41**, 2875–2888, <https://doi.org/10.1007/s00382-012-1600-0>.
- Svoboda, M., and Coauthors, 2002: The Drought Monitor. *Bull. Amer. Meteor. Soc.*, **83**, 1181–1190, <https://doi.org/10.1175/1520-0477-83.8.1181>.
- Swenson, S., J. Wahr, and P. C. D. Milly, 2003: Estimated accuracies of regional water storage variations inferred from the gravity recovery and climate experiment (GRACE). *Water Resour. Res.*, **39**, 1223, <https://doi.org/10.1029/2002WR001808>.
- Tapley, B. D., S. Bettapur, M. Watkins, and C. Reigber, 2004: The gravity recovery and climate experiment: Mission overview and early results. *Geophys. Res. Lett.*, **31**, L09607, <https://doi.org/10.1029/2004GL019920>.
- Thomas, A. C., J. T. Reager, J. S. Famiglietti, and M. Rodell, 2014: A GRACE-based water storage deficit approach for hydrological drought characterization. *Geophys. Res. Lett.*, **41**, 1537–1545, <https://doi.org/10.1002/2014GL059323>.
- van Lanen, H. A. J., and E. Peters, 2000: Definition, effects and assessment of groundwater droughts. *Drought and Drought Mitigation in Europe*, J. V. Vogt and F. Somma, Eds., Kluwer Academic Publishers, 49–61.
- Wanders, N., S. Thober, R. Kumar, M. Pan, J. Sheffield, L. Samaniego, and E. F. Wood, 2019: Development and evaluation of a pan-European multimodel seasonal hydrological forecasting system. *J. Hydrometeor.*, **20**, 99–115, <https://doi.org/10.1175/JHM-D-18-0040.1>.
- Winter, T. C., J. W. Harvey, Q. L. Franke, and W. M. Alley, 1998: Ground water and surface water: a single resource. USGS Circular 1139, 79 pp., <https://doi.org/10.3133/cir1139>.
- Xia, Y., and Coauthors, 2012: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res.*, **117**, D03109, <https://doi.org/10.1029/2011JD016048>.
- Yin, J., and X. Zhan, 2018: Impact of bias-correction methods on effectiveness of assimilating SMAP soil moisture data into NCEP Global Forecast System using the ensemble Kalman filter. *IEEE Geosci. Remote Sens. Lett.*, **15**, 659–663, <https://doi.org/10.1109/LGRS.2018.2806092>.
- Zaitchik, B. F., M. Rodell, and R. H. Reichle, 2008: Assimilation of GRACE terrestrial water storage data into a land surface model: Results for the Mississippi River basin. *J. Hydrometeor.*, **9**, 535–548, <https://doi.org/10.1175/2007JHM951.1>.