Improved Hydrological Simulation Using SMAP Data: Relative Impacts of Model Calibration and Data Assimilation

Randal D. Koster[1], Qing Liu[1,2], Sarith P. P. Mahanama[1,2], and Rolf H. Reichle[1]

[1]Global Modeling and Assimilation Office, NASA/GSFC, Greenbelt, Maryland

[2]Science Systems and Applications, Inc., Lanham, Maryland

Address correspondence to:

Randal Koster

Global Modeling and Assimilation Office

Code 610.1, NASA/GSFC

Greenbelt, MD 20771

301-614-5781; randal.d.koster@nasa.gov

Submitted to: J. Hydrometeorology

Abstract

The assimilation of remotely sensed soil moisture information into a land surface model has been shown in past studies to contribute accuracy to the simulated hydrological variables. Remotely sensed data, however, can also be used to improve the model itself through the calibration of the model's parameters, and this can also increase the accuracy of model products. Here, data provided by the Soil Moisture Active/Passive (SMAP) satellite mission are applied to the land surface component of the NASA GEOS Earth system model using both data assimilation and model calibration in order to quantify the relative degrees to which each strategy improves the estimation of near-surface soil moisture and streamflow. The two approaches show significant complementarity in their ability to extract useful information from the SMAP data record. Data assimilation reduces the ubRMSE (the RMSE after removing the long-term bias) of soil moisture estimates and improves the timing of streamflow variations, whereas model calibration reduces the model biases in both soil moisture and streamflow. While both approaches lead to an improved timing of simulated soil moisture, these contributions are largely independent; joint use of both approaches provides the highest soil moisture simulation accuracy.

1. Introduction

One of the flagship science data products of the National Aeronautics and Space Administration's (NASA's) Soil Moisture Active Passive (SMAP) satellite mission (Entekhabi et al. 2010a) is an extensive set of estimates (retrievals) of moisture content in the top several centimeters of soil. These retrievals, provided on a global ~36km grid with a repeat time of 3 days or less, are derived from L-band measurements taken with a passive radiometer and feature significantly reduced errors from the radiofrequency interference that can plague such datasets (Piepmeier et al, 2014, 2017; Kerr et al. 2016). Overall SMAP soil moisture retrieval accuracy has been shown to be quite high (Chan et al. 2016a, 2018).

A unique feature of the core SMAP mission is the publication of an enhanced Level-4 Soil Moisture (L4_SM) product through the assimilation of the observed brightness temperatures into a land surface model (LSM). Through the assimilation process (Reichle et al. 2017a), the LSM combines the SMAP brightness temperatures with observations-based meteorological forcing to produce a soil moisture product that is superior to an "open loop" LSM-based product, i.e., the LSM product generated without the assimilation of SMAP data. The development of the SMAP L4_SM product is based on an extensive body of research into soil moisture data assimilation (Reichle et al. 2002a, Reichle 2008, Kumar et al. 2008, Drusch et al. 2009, Draper et al. 2012, de Rosnay et al. 2013, Carrera et al. 2015, De Lannoy and Reichle 2016ab,). The L4_SM product has already been evaluated successfully against a host of in-situ soil moisture observations (Reichle et al. 2017a) and in the context of key assimilation diagnostics (Reichle et al. 2017b).

SMAP data, however, can potentially interface with an LSM in another useful way: through the calibration of model parameters. Calibration in this context involves identifying the

values of targeted parameters (typically, parameters that cannot be easily quantified with direct measurements) that lead, in a model simulation, to the most accurate reproduction of the satellite-measured variable. Using similar L-band soil moisture retrievals from the Soil Moisture and Ocean Salinity mission (Kerr et al. 2010), Shellito et al. (2016) calibrated an LSM's soil hydraulic properties, improving its simulation of soil moisture; their study calls to mind earlier calibrations of LSM soil and vegetation parameters using satellite-based surface temperatures (e.g., Crow et al. 2003, Gutmann and Small 2010, Corbari and Mancini 2014).

L-band soil moisture retrievals indeed reveal important timescales of near-surface soil moisture dynamics (McColl et al. 2017) – timescales that could serve as targets for a model calibration exercise. To demonstrate the potential of calibration more clearly, we present here an example taken from the analysis performed later in this paper. Consider the time series plots shown in Figure 1, obtained for a grid cell in the Little Washita watershed of southwestern Oklahoma (O'Neill et al. 2016) during the period May – September 2016. Figure 1a shows the gauge-based precipitation rates recorded there, and Figure 1b shows the contemporaneous SMAP Level-2 (non-assimilated) soil moisture retrievals. The retrieved soil moisture increases as expected during precipitation events (e.g., on day 164), and it subsequently dries down with a time scale of a few days. Now consider the soil moisture time series in Figure 1c, which was produced at the site by an LSM without the benefit of data assimilation but with the precipitation information contained in Figure 1a. The LSM used here is the Catchment LSM of the NASA Global Earth Observing System (GEOS) – the LSM underlying the L4_SM product, as discussed in Section 2b. The modeled soil moisture also increases as desired during precipitation events, but the time scale of drydown is noticeably longer – the drydown occurs over a span of 1-2 weeks. In this context the model does not behave like nature, at least nature as represented by

SMAP. This particular facet of Catchment LSM behavior was in fact heretofore never carefully examined.

Now consider the soil moisture time series in Figure 1d, which was produced by the same LSM after calibrating a particular parameter. (Details are provided below in Section 2c.) While the model results still differ somewhat from the observations in terms of absolute magnitude, the timescale of the drydown is more in line with that captured by the SMAP retrievals. We thus might expect the calibrated model results in Figure 1d to be more realistic than the uncalibrated results in Figure 1c – they might agree better with independent in-situ soil moisture observations.

Data assimilation and model calibration are in fact expected to improve soil moisture estimation in different ways. Model calibration specifically addresses deficiencies in the model's representations of physical processes, improvements that can manifest themselves at every simulation time step. Data assimilation corrects for such deficiencies "after the fact" and only at selected times and locations, depending on the availability of the satellite data; however, unlike calibration, data assimilation also corrects for potentially important deficiencies in the meteorological forcing. To some extent the contributions of the two approaches to improved soil moisture estimation are complementary. They may indeed build on each other, so that applying both approaches together may lead to soil moisture estimates of unprecedented accuracy.

We explore this potential complementarity in the present paper. We use the Catchment LSM to produce four sets of soil moisture estimates: (i) open-loop estimates with the default version of the LSM, (ii) estimates obtained through the assimilation of SMAP data into the default LSM, (iii) open-loop estimates obtained after the LSM has been calibrated with SMAP data, and (iv) estimates obtained through the assimilation of SMAP data into the calibrated LSM. By evaluating the relative accuracies of the four sets of estimates against independent in-situ

data, we can isolate the contributions of data assimilation and model calibration to hydrological estimation as well as quantify their joint impact.

Section 2 below describes data used in the analysis as well as the LSM, the calibration procedure, and the data assimilation system. Section 3 presents the results, focusing on the accuracy of the simulated near-surface soil moisture and streamflow. Section 4 provides a summary and discussion.

2. Data and Models Used

a. SMAP Soil Moisture and Brightness Temperature Data

Different components of the SMAP data suite are used in this study. For the calibration exercise (Section 2c), we use Version 4 of the SMAP Level-2 soil moisture retrievals (O'Neill et al. 2016), a set of retrievals derived from L-band radiometer measurements that represent volumetric soil moisture in roughly the top 5 cm of soil. We use the data associated with the descending overpasses, which correspond to a 6AM local collection time. The data are provided on the 36-km Equal Area Scalable Earth (EASE) grid (Brodzik et al. 2012), with retrieval values provided at a given grid cell at least once every three days. We achieve extensive spatial and temporal coverage of soil moisture data for our analysis by utilizing the retrievals flagged as having "uncertain quality" along with those flagged as having "recommended quality".

For the data assimilation exercise (Section 2d), we use Version 3 of the 36-km resolution SMAP Level-1C brightness temperature observations (Chan et al. 2016b). The assimilated SMAP observations include horizontal-polarization and vertical-polarization brightness

temperatures from ascending and descending half-orbits (after first averaging over fore- and aft-looking data).

b. Land Surface Modeling System

The LSM used for all simulations is the Catchment LSM (Koster et al. 2000, Ducharne et al. 2000), the LSM underlying the MERRA-2 reanalysis (Gelaro et al. 2017, Reichle et al. 2017c) and the SMAP L4_SM product (Reichle et al. 2017a). It solves the land surface energy and water balance at every simulation time step, partitioning precipitation inputs into runoff, evapotranspiration, and changes in water storage, and partitioning radiative energy inputs into latent heat, sensible heat, and changes in energy storage. A key feature of the LSM is its explicit treatment of spatial soil moisture heterogeneity (as determined from topographic conditions) and its effects on the surface water fluxes – evapotranspiration and runoff generation, for example, both occur more efficiently in the (dynamically varying) sub-catchment areas characterized by wetter conditions.

The Catchment LSM follows a prognostic soil water variable representing the top 5 cm of soil, and the average soil moisture in the top 5 cm is a standard simulation output. This depth is consistent with the ostensible sensing depth of the SMAP radiometer (section 2a). Also standardly produced are surface runoff and baseflow fluxes, the sum of which are averaged here in space and time for comparison against streamflow measurements.

c. Land Surface Model Calibration Strategy

While the Catchment LSM's performance has been evaluated in numerous venues (e.g., Bowling et al. 2003, Boone et al. 2004, Reichle et al. 2011), its treatment of near-surface

moisture and how it relates to the root zone has never been properly calibrated. Indeed, one study (Kumar et al. 2009) suggests that the connection between the near-surface and deeper soil moisture in the model may be too strong – it is, in any case, stronger than that seen in some other models. This particular aspect of the LSM can thus be considered ripe for calibration.

The Catchment LSM's formulation of near-surface soil moisture dynamics uses two independent processes to replenish drying soil via recharge from below (see Koster et al. [2000] for details). First, replenishment occurs through changes in the equilibrium moisture state of the catchment; as this equilibrium water increases, some of the increase is deposited in the near-surface soil. Second, the upward flow of soil moisture between the root zone and the near-surface soil in non-equilibrium situations is determined through parameterized fits of detailed Richards equation calculations – in effect, as the near-surface soil dries, the increasing vertical gradient in matric potential overcomes gravity, allowing upward moisture flow.

Of relevance to the present study is the recent inclusion of a time-invariant parameter, α, into the formulation of the second process. In the new version of the formulation, any upward moisture flux in non-equilibrium situations is multiplied by α, where α lies between 0 and 1. The imposed reduction in upward flow can be considered a reflection of the fact that near-surface soils in nature are more heterogeneous than those tested in laboratories, making upward flow more difficult than laboratory-established soil parameters would suggest. This interpretation, however, is rather loose, since the equilibrium profile of soil moisture is not similarly adjusted; α is thus perhaps best considered here to be a simple tuning parameter. The value of α turns out to have a first order impact on the character of the simulated soil moisture, as illustrated earlier in Figure 1d. At this grid cell, replacing α's default value of 1 with a value of 0.01 produced a

185 better match (in terms of temporal variability and the speed of drydown) of simulated soil
186 moisture with the SMAP data.

187 In a calibration exercise, a number of open-loop (land model only) simulations, each
188 utilizing a different value for α, generated soil moisture time series for 2015-2016 across the
189 continental US and portions of Canada and Mexico. This region was chosen for study because it
190 offers two key advantages: (i) high quality precipitation measurements based on a dense rain
191 gauge network, and (ii) a broad range of climates, with wetter conditions in the east and much
192 drier conditions in the west. The forcing data applied in these simulations were derived from
193 atmospheric analysis, with the analysis-generated precipitation corrected by gauge observations;
194 the forcing data essentially match those used in the generation of the Version 2 L4_SM product
195 for 2015-2017 and those used in the production of the corresponding 2000-2014 model-only
196 simulation for the SMAP project (Nature Run v4; Reichle et al. 2017a).

197 At a given SMAP grid cell, and for a given simulation, we evaluated the agreement
198 (using the temporal correlation coefficient, R) between the local time series of SMAP retrievals
199 and the model-simulated surface soil moisture time series. We then repeated the process with
200 data at that grid cell from each of the other simulations. By comparing the different R values, we
201 were able to determine for the grid cell the single value of α that allows the best reproduction of
202 the behavior of the SMAP retrievals. The spatial distribution of these optimized α values is
203 shown in Figure 2. The optimal values clearly vary in space, with smaller values in the east.
204 Notice that the default value of 1 works best at only a handful of locations.

205

206 **d. Data Assimilation System**

The data assimilation system used in this study is essentially the same as that used to generate the SMAP L4_SM product (Reichle et al. 2017a,b). It uses an ensemble Kalman filter (Evensen, 2003) and assimilates horizontally and vertically polarized SMAP brightness temperature observations from the Version 3 SMAP L1C_TB product (Chan et al. 2016b) from both ascending and descending half-orbits. The observed brightness temperatures are differenced with corresponding brightness temperatures generated from the Catchment model's soil moisture and temperature estimates, which are calculated using a zero-order "tau-omega" radiative transfer model (De Lannoy et al. 2013). The brightness temperature differences are then inverted into corrections of the model forecast soil moistures and surface temperatures based on the modeled error covariances, which are diagnosed from the ensemble. The analysis is bias-corrected by a rescaling – prior to assimilation – of the SMAP brightness temperature observations into the spatially and seasonally varying climatology of the modeled brightness temperature. Reichle et al. (2017a, see their section 2d) provide a detailed description of the different facets of the data assimilation system, including the model and observation error parameters.

e. Simulations Performed

Results from four offline simulations with the Catchment LSM were evaluated for this study:

(i) A model-only "baseline" simulation (BL) with the default version of the Catchment LSM, i.e., the version used to produce the Version 2 L4_SM product (Reichle et al. 2017a). Note that in the default model, α is set to 1 everywhere.

(ii) A data assimilation simulation (BL_DA) with the default version of the Catchment LSM that includes the assimilation of SMAP brightness temperatures, as outlined in Section 2d above.

(iii) A model-only simulation (OPT, for "optimized parameters") with a version of the Catchment LSM that uses the spatial distribution of optimized parameters illustrated in Figure 2.

(iv) A data assimilation simulation (OPT_DA) with both the use of the optimized parameters from Figure 2 and the assimilation of SMAP brightness temperatures.

All four simulations covered the period April 2015 – March 2017 and were run across the conterminous United States (CONUS) on the SMAP 36-km EASE grid (section 2a). The BL and OPT simulations were spun up independently for 30 years. The BL_DA and OPT_DA simulations used the same perturbation and radiative transfer model parameters as used for the SMAP L4_SM product (Reichle et al. 2017 a,b). The SMAP brightness temperatures were assimilated after removing their seasonally-varying bias relative to the model forecast brightness temperatures. The rescaling parameters were constructed separately for the BL and OPT simulations using the (version 6) brightness temperature from the Soil Moisture Ocean Salinity mission (Kerr et al. 2016) for the period July 2010 to June 2016 and the underlying Catchment LSM. Output diagnostics produced by each simulation include 3-hourly near-surface (top 5 cm) soil moistures and total runoff fluxes.

f. Validation Data

For validating simulated near-surface soil moisture, we utilize a number of in-situ soil moisture measurement sites encompassed by the USDA Natural Resources Conservation Service Soil Climate Analysis Network (SCAN; Schaefer et al. 2007) and the US Climate Reference Network (USCRN; Diamond et al. 2013, Bell et al. 2013). Quality control was applied to the hourly in-situ measurements at these sites; we eliminated measurements indicating volumetric soil moisture below 0 m^3m^{-3} or greater than 0.6 m^3m^{-3} as well as measurements taken when the contemporaneous soil temperature was below 4^{o}C. We also filtered out obviously unphysical measurements associated with spikes, sudden dry-downs, or high-frequency oscillations. The quality-controlled hourly data were averaged into 3-hourly time series, on which we base the soil moisture-related evaluations in section 3a.

Some caveats regarding the use of such in-situ observations for validation are worth noting. First, the site measurements are highly localized, whereas the soil moisture estimates produced by the simulations represent a large-scale spatial average. This leads to a potentially important spatial representativeness error – the local measurement may not properly represent the large-scale average. Second, the LSM's near-surface soil moisture variable represents an average over the top 5 cm of soil, whereas the in-situ measurements do not represent such a depth average – for the sites examined here, the surface soil moisture measurements instead represent conditions at a depth of about 5 cm. Time variability of soil moisture at a 5 cm depth can differ from that of the depth-averaged soil moisture above it (Shellito et al. 2016); presumably, time scales for the depth-averaged moisture will be shorter than those at 5 cm depth. The spatial representativeness error and the vertical mismatch between the in-situ measurements and the modeled soil moisture variable will influence (and will presumably inflate inappropriately, though perhaps to only a small degree) the error metrics we compute for the

simulations. We make the assumption here that these issues affect to some extent all four of our simulations, so that the relative magnitudes of the skill metrics across the simulations are still telling. In addition, we emphasize the key advantage of the SCAN and USCRN networks: they encompass the continental US and thereby cover a broad range of soil textures and background climates (Reichle et al. 2017c).

For the validation of simulated streamflow, we examine a subset of the 573 unregulated CONUS hydrological basins analyzed by Kumar et al. (2014): the 240 basins that lie within the intermediate size range (2,000—10,000 km^2). Daily streamflow data for the 240 basins were obtained from the U.S. Geological Survey (USGS; http://nwis.waterdata.usgs.gov/nwis) for the period 1980 through September 2017. For each basin, observed river discharges were normalized by basin area to convert the discharges to units of mm d^{-1}. Conversions between the irregular basin areas and the SMAP EASE grid cells were facilitated by EASE grid arrays containing the fractions of each basin held within each grid element.

g. Skill Metrics

We evaluate model simulation skill against the in-situ observations using three metrics: the unbiased root mean square error (ubRMSE), the bias, and the temporal correlation (R). The bias, of course, is simply the long-term mean of the simulation variable minus the long-term mean of the corresponding set of observations. The ubRMSE at a given site can be computed from the bias and the traditional root mean square error (RMSE) via

$$\mathrm{ubRMSE}^2 = \mathrm{RMSE}^2 - \mathrm{bias}^2. \quad (1)$$

In essence, (1) recognizes the fact that the error characterized by the traditional RMSE stems both from a mean bias and from residual, time-varying errors that remain after the mean bias is removed from the time series. We represent the latter error contribution with the ubRMSE, using (1) (see Entekhabi et al. 2010b).

The final metric, the temporal correlation R (the traditional Pearson correlation coefficient), concerns itself with the timing and relative magnitudes of anomalies in a time series. For the near-surface soil moisture comparisons below, both R and ubRMSE are computed from time series of 3-hourly data. For the streamflow comparisons, R is computed from the smoothed time series obtained by applying a 10-day moving average window to the daily observed and simulated totals during warm season months (April through September of 2015 and 2016). We limit the streamflow R calculation to warm season months in order to avoid, for most of CONUS, intense periods of snowmelt-dominated runoff generation. We use the 10-day moving window to minimize errors associated with streamflow routing times (the time it takes for locally-generated runoff to reach a stream gauge observations site), given that a routing model is not employed in this analysis.

For both near-surface soil moisture and streamflow, when averaging the skill metrics across CONUS, we account for spatial correlations in the site values, so that clusters of nearby similar measurements do not contribute excessively to the computed averages (Reichle et al. 2017a). For soil moisture, we also compute 95% confidence intervals for the averages using the approach in Reichle et al. (2017a), indicating these uncertainty estimates with lines (or "error bars") in the histograms showing average skill across CONUS. Note that the estimation of these uncertainty estimates is far from perfect. The estimation approach assumes that the errors in the simulation products are completely random, which is not the case for a model that, for example,

317 always produces an overly extended drydown period after a storm event (as the Catchment LSM
318 appears to do in Figure 1). Because presumably not all of the simulation errors examined here
319 are random, the estimated uncertainty levels provided here are probably overestimated, perhaps
320 significantly so. As for streamflow, we avoid assigning uncertainty estimates due to the small
321 sample size of the temporal data relative to that for soil moisture – again, for the streamflow R
322 metric, we are considering 10-day averages during the warm season of two years. We instead
323 qualitatively address the relevance of any streamflow simulation improvements in terms of "field
324 significance", i.e., the preponderance of sites that show improvement over those that show
325 degradation.

326

327 **3. Results**

328

329 **a. Near-Surface Soil Moisture**

330 1) Time Series Comparison at a Representative Site

331 A representative comparison of the soil moisture time series produced in the four
332 simulations is shown in Figure 3. The Prairie View site (30.08°N, 95.98°W), a SCAN site
333 located west of Houston, Texas, provides sub-diurnal soil moisture measurements at multiple
334 depths throughout the SMAP period. The site measurements at the 5-cm depth, averaged over
335 each day during February-July 2016, are plotted in both panels of the figure as a heavy black
336 curve. The daily-averaged simulated soil moistures at the corresponding grid cell are also
337 plotted, with BL and BL_DA included in Figure 3a and OPT and OPT_DA in Figure 3b. For
338 ease of considering the ubRMSE and R metrics, the mean bias of each simulated time series
339 computed over the February-July 2016 period was removed. These biases at this location and for

this time period amount to 0.18 m^3/m^3 for BL, 0.072 m^3/m^3 for OPT, 0.15 m^3/m^3 for BL_DA, and 0.066 m^3/m^3 for OPT_DA.

The baseline simulation (BL) is seen to follow the ups and downs of the observations fairly well, indicating that the applied rainfall forcing for the period – the timing of the storms and interstorm periods – is reasonably accurate. As already suggested in Figure 1, however, the timescale of post-storm drydowns is excessive in the model; the simulated soil moisture in BL (blue curve in Figure 3a) takes about a month to dry following the storm occurring just prior to mid-March, whereas significant drydown for the in-situ measurements occurs within a week. The assimilation of SMAP brightness temperatures into the baseline model (BL_DA) leads to more realistic amplitudes of soil moisture variation (particularly in June and July) and a somewhat more accurate drydown timescale, with a faster drydown, for example, in late June and early July.

Using the calibrated α parameter in OPT clearly leads to faster, and thus generally more realistic, drydowns (Figure 3b). For example, unlike BL and BL_DA, both OPT and OPT_DA produce reasonable drydowns in late March. While the large June drydown is simulated better in OPT than in either BL or BL_DA, OPT_DA performs better still. Also, note that while the plots appear to suggest that OPT and OPT_DA are wetter than BL and BL_DA following rainfall events in late February and mid-March, the soil moisture maxima achieved then by the four simulations are in fact roughly the same – a significantly larger bias (by about 0.1 m^3/m^3) had been subtracted from the BL and BL_DA results prior to plotting (see above). The soil moisture minima for the four simulations are accordingly very different, with significantly lower minima obtained for OPT and OPT_DA. The precipitation event in late February had a larger impact on

soil moisture in OPT than in BL presumably because the pre-storm soil in the former was much drier.

The ubRMSE for BL, BL_DA, OPT, and OPT_DA over the particular time period shown in Figure 3 are, respectively, 0.066, 0.053, 0.069, and 0.052. Thus, according to the ubRMSE metric, and for this particular site and time period, only data assimilation produces an improved simulation of soil moisture – the excessive amplitude of variation (relative to observations) produced in the OPT simulation apparently counteracts the effects of the improved drydown timescale. The temporal correlation metric R, which focuses less on such amplitudes, tells a different story – R for BL, BL_DA, OPT, and OPT_DA is 0.73, 0.82, 0.85, and 0.85, respectively. According to the R metric, both model calibration and data assimilation contribute accuracy to simulated soil moisture at this site.

Regarding the somewhat excessive amplitude seen in the OPT results and its apparent effect on the ubRMSE metric, it is worth remembering (section 2f) that the depth of the in-situ measurements is inconsistent with the depth represented by the land model and of the SMAP signal used in the calibration, which ostensibly corresponds to average soil moisture conditions within the top 5 cm. A SMAP-calibrated model might indeed be expected to produce such higher amplitudes – because soil moisture variations tend to decrease with increasing depth into the soil, the variations at 5 cm depth (where the in-situ measurements are taken) should show a reduced amplitude relative to those in the soil above. This may artificially increase the estimated ubRMSE.

2) Results across CONUS

384 The results in Figure 3 for the Prairie View site are in fact typical, as indicated in Figure
385 4. Figure 4a first shows the distribution of ubRMSE across the USCRN and SCAN sites within
386 CONUS over the April 2015 – March 2017 period. Some large errors appear in the Southeast
387 and up through the Mississippi Valley. The smallest errors are seen in the Southwest, perhaps
388 reflecting the drier soils there and the associated lower temporal variability.

389 Figures 4b, 4c, and 4d then show, respectively, the changes in ubRMSE obtained in the
390 BL_DA, OPT, and OPT_DA experiments with respect to the BL simulation. Warm colors
391 (yellow to red) in the latter three panels indicate an increase in ubRMSE and thus a degradation
392 in simulation skill compared to BL, whereas blue shading indicates a reduction in ubRMSE and a
393 more accurate simulation. Each map shows a mix of improvements and degradations. Note that
394 we can expect some degradations even for the OPT_DA simulation simply because the in situ
395 validation data are themselves imperfect; the in situ data are subject, for example, to spatial and
396 vertical representativeness error (section 2f). As noted above, vertical representativeness issues
397 are particularly relevant to the ubRMSE calculation and can complicate skill comparisons
398 between the simulations. Spot checks of ubRMSE degradations seen for OPT_DA, for example,
399 show that the amplitudes of soil moisture variations in OPT_DA can be greater than those for
400 BL, which look more like those of the in situ measurements. Again, this does not necessarily
401 imply a reduction in true skill relative to BL, given that the model data effectively (and properly)
402 represent soil moisture at a depth shallower than 5 cm, which should indeed vary somewhat more
403 than the in situ values. Some of the degradations seen in the maps, however, are suggestive for
404 other reasons. The higher ubRMSE for the OPT experiment along the Mississippi-Arkansas
405 border (Figure 4c), for example, may reflect difficulties in calibrating the model in such regions

406 of extensive irrigation (Kumar et al. 2015), given that irrigation is not explicitly treated in the
407 model.

408 Such issues aside, the maps show that overall, reductions outweigh increases in ubRMSE,
409 particularly for BL_DA and OPT_DA. This implies that SMAP data indeed contribute to
410 improved soil moisture estimation through both data assimilation and (to a lesser extent) the
411 optimization of the model parameters.

412 This result is summarized in Figure 4e, which shows the average ubRMSE across the
413 CONUS validation points for each of the four simulations. While all three experiment
414 simulations (BL_DA, OPT, and OPT_DA) on average perform better than the baseline run, the
415 improvement seen with OPT is very small; the improvements are clearly larger when data
416 assimilation is employed. Note that the improvements obtained with data assimilation are likely
417 to be more significant than suggested by the overlapping 95% uncertainty ranges, given that
418 these ranges are themselves likely to be overestimated (section 2g).

419 The temporal correlation metric R is examined in Figure 5. Curiously, despite the
420 baseline simulation's relatively poor performance in the Southeast according to the ubRMSE
421 metric (Figure 4a), the R values produced there are high (Figure 5a). We also see that for the R
422 metric, all three experiment simulations show a general improvement (Figures 5b,c,d) over the
423 baseline simulation, with the increases in R overwhelming the handful of decreases.

424 The CONUS-wide averages of R shown in Figure 5e are especially telling. First, the
425 average R values for all three experiment simulations (BL_DA, OPT, and OPT_DA) lie
426 significantly above that for BL (as indicated by the non-overlapping 95% confidence intervals,
427 which, again, are likely to be overestimated anyway). Second, the contributions of data

assimilation and parameter calibration to the average R values in Figure 5e appear largely complementary – the increase in R from data assimilation (BL_DA minus BL) added to that obtained from parameter estimation (OPT minus BL) roughly equals the increase obtained when data assimilation and parameter calibration are employed together (OPT_DA minus BL). Such complementarity speaks to the value of considering multiple facets of SMAP data together when attempting to maximize the data's usefulness.

Results for bias are shown in Figure 6. Data assimilation is seen to have little impact on the bias (Figure 6b). In contrast, model calibration has a large impact, sometimes increasing the absolute value of the bias and sometimes decreasing it (Figure 6c). Generally, though, the calibration leads to an improvement, as indicated by the averages in Figure 6e.

The relative impacts of data assimilation and model calibration on model bias in Figure 6 are not unexpected. By design, our data assimilation procedure ingests SMAP data after transforming the data to be consistent with the host model's climatology, so that data assimilation by itself should have minimal impact on bias. The calibration of the model parameters, on the other hand, has a first order impact on the model's physics and thus on any biases generated. The bias reductions found for OPT and OPT_DA indicate that this overall impact is, on average, positive.

b. Streamflow in Small, Unregulated Basins

The above analysis shows that SMAP soil moisture retrievals and associated brightness temperatures have a positive impact on soil moisture simulation, with complementary contributions from data assimilation and model calibration. To what extent, though, do the

450 different strategies for using SMAP data lead to improvements in overall hydrological simulation
451 – in the partitioning, for example, of incident precipitation into streamflow, evapotranspiration,
452 and changes in storage? In this section we focus specifically on the simulation of streamflow
453 (given the wealth of available streamflow data relative to large-scale evapotranspiration data);
454 we compare the abilities of the baseline simulation and the three experiment simulations to
455 reproduce streamflow characteristics observed across CONUS.

456 Figure 7a shows, for the baseline simulation, the error in the runoff ratio at each of the
457 unregulated, medium-sized basins described in section 2f. For both the simulation and the
458 observations, we divide a given basin's total streamflow, Q, for September 2015 through August
459 2017 by the total precipitation, P, in that basin over the same period (computed directly from the
460 gridded precipitation data used in the four simulations). We then plot in Figure 7a the difference
461 between the modeled and observed ratios, with the dots positioned on the centroid of the basin.
462 Because the Catchment LSM tends to underestimate streamflow, the raw errors in Figure 7a tend
463 to be negative. (This problem is discussed in more detail, along with a potential solution, by
464 Koster and Mahanama (2012)). The runoff ratio errors for BL are especially large in parts of the
465 Northwest, in the upper Midwest, and in Appalachia.

466 Figures 7b, 7c, and 7d show, respectively, the changes in the absolute runoff ratio error
467 (compared to BL) for simulations BL_DA, OPT, and OPT_DA, with the color of the dots
468 indicating an improvement (blue shading) or degradation (yellow to red shading) in the
469 simulated runoff ratio. Averages (section 2g) across the basins for the different simulations are
470 provided in Figure 7e. The averages indicate an improvement in runoff ratio estimation
471 stemming from the use of the calibrated model parameters – an improvement that appears
472 significant, given the preponderance of blue dots in Figures 7c and (to a lesser extent) 7d. Data

assimilation by itself is seen to have little impact on Q/P accuracy and even seems, for this metric, to reduce slightly the ability of model calibration to have a positive impact, as seen by the higher average error for OPT_DA relative to OPT.

Figure 8 shows results for an alternative measure of runoff simulation skill: the temporal correlation, R, between observed and simulated runoff totals. For the warm period (April through September) of both 2015 and 2016, time series of observed 10-day basin streamflows were correlated against the corresponding streamflows simulated in BL. The resulting R value for each basin is plotted at the basin's centroid in Figure 8a; correlations are seen to be reasonably high, particularly in the east (except in Florida and Maine) and the Pacific Northwest.

Figures 8b, 8c, and 8d show, respectively, the change in R obtained in simulations BL_DA, OPT, and OPT_DA. Improvements strongly overwhelm degradations for BL_DA (Figure 8b). The average of R for all four simulations is provided in Figure 8e. For this metric, model calibration has little impact, whereas the impact of data assimilation is relatively strong. This is presumably because only data assimilation corrects for errors in the timing of precipitation, which necessarily has a first order impact on the timing of streamflow volumes.

Note that using both data assimilation and model calibration together (simulation OPT_DA) leads to an increase in R relative to the BL and OPT simulations, but not to the extent seen in simulation BL_DA. For this particular evaluation, the effects of the two data utilization strategies do not appear independent and additive. Potential "destructive interference" of the two approaches in the generation of higher level fields such as runoff may be a fundamental characteristic of their joint application.

Alternatively, we can speculate that the apparent non-additivity in Figure 8e reflects an insufficient tuning of the data assimilation system underlying the OPT_DA simulation. Unlike the system underlying BL_DA, which underwent extensive development and testing for the generation of the SMAP L4_SM product, the system underlying OPT_DA has only been exercised in the present study. The system underlying OPT_DA lacks, for example, a proper tuning of model and observation error settings. Moreover, recalibration of the parameters underlying the radiative transfer model would bring the modeled brightness temperatures closer to the observations in a climatological sense and lessen the work left to the rescaling process, which might further improve the assimilation estimates. Further investigation of these potential improvements is left for future work.

4. Summary and Discussion

In the present study, two different approaches are used to integrate SMAP data into a land surface model's representation of near-surface soil moisture and hydrological fluxes. These approaches are distinct and largely complementary. In a standard open-loop land modeling exercise, a land model driven with observations-based meteorological forcing (rainfall, air temperature, etc.) produces, as a matter of course, estimates for hydrological states (e.g., soil moisture) and fluxes (e.g., evapotranspiration and runoff). Data assimilation, the first approach toward integrating SMAP data into these estimates, operates as a "course correction" to the evolving states, intermittently adjusting the states toward measured values and thereby correcting for errors that stem from inadequate model parameterizations and uncertainty in the meteorological forcing. The second approach, model calibration, utilizes the SMAP data to improve the performance of the land model itself by increasing the realism of its underlying parameterizations.

518 Our four simulations quantify the skill of reproducing both observed near-surface soil
519 moisture and observed streamflow when: (i) neither approach is used (simulation BL); (ii)
520 SMAP data are assimilated into the system (simulation BL_DA); (iii) SMAP data are used to
521 calibrate a particularly relevant model parameter (simulation OPT); and (iv) SMAP data are used
522 for both model calibration and assimilation (simulation OPT_DA). The results indeed
523 demonstrate some complementarity in the contributions of the two approaches to simulation
524 accuracy. For near-surface soil moisture, data assimilation produces the largest reductions in
525 ubRMSE (Figure 4), but model calibration produces the greatest reduction in bias (Figure 6).
526 Both data assimilation and model calibration produce significant improvements in the temporal
527 correlation R, and these improvements appear independent; the sum of these improvements, as
528 obtained from simulations BL_DA and OPT, roughly equals the improvement obtained in
529 OPT_DA, the simulation that combines the two approaches (Figure 5). The two approaches
530 appear particularly complementary in their contributions to the simulation of streamflow. Model
531 calibration with SMAP data leads to improvements in the simulation of the long-term runoff
532 ratio (Figure 7) but has little impact on the timing of streamflow (Figure 8). In contrast, data
533 assimilation has little impact on simulated runoff ratio but a positive impact on streamflow
534 timing.

535 We emphasize again that this complementarity is not surprising. The data assimilation
536 strategy employed here (which in fact underlies the generation of the SMAP L4_SM product)
537 transforms the SMAP brightness temperatures into values consistent with the climatology of the
538 model before ingesting them into the model. Assimilating the SMAP data will thus have
539 relatively small impacts on the climatology of the model products, as represented here by biases
540 in both the near-surface soil moisture and the runoff ratio. The assimilation will, however, have

important impacts on the timing of the variables produced, since it corrects for errors in the meteorological forcing data that drive the model. Correcting forcing-related errors should improve both ubRMSE and R.

In contrast, model calibration directly affects the climatology of the model, and thus associated improvements can be seen in the model biases. Model calibration also improves the R metric for near-surface soil moisture (computed from 3-hourly values), presumably through its improvement of drydown behavior (Figure 3). This, however, does not translate here into an improvement in R for 10-day streamflow totals (Figure 8c); the 10-day runoff averaging period, necessitated by the comparisons against the stream gauge measurements, may have precluded this benefit. In any case, unlike data assimilation, model calibration cannot correct for errors in the precipitation forcing.

It is important to note that the data assimilation and model calibration exercises performed and compared here utilized different subsets of the SMAP data product, subsets specific to the needs of the given procedure. For example, as noted above, the model calibration exercise utilized soil moisture retrievals flagged as having "uncertain quality" as well as "recommended quality". This was necessary to extend spatially the areas in which calibration could be performed; by allowing the additional data, calibrations could be performed for this study, for better or worse, across CONUS. (Note that in Figures 5-7, the OPT experiment does show improvements over the baseline experiment even in the far eastern part of CONUS, where almost all of the data are flagged as "uncertain".) The data assimilation procedure used here, however, follows that of the SMAP Level 4 system and therefore only assimilates brightness temperatures flagged as having "recommended quality". While the different flag criteria across the experiments may appear inconsistent, note that ensuring flag consistency is somewhat

564 inappropriate given that the flags for soil moisture retrievals and brightness temperatures are
565 themselves different – SMAP brightness temperature measurements are often flagged as
566 "recommended" even when the soil moisture retrievals themselves are flagged as "uncertain".
567 Given the varying degrees to which even "recommended quality" brightness temperatures are
568 allowed to affect soil moisture in a data assimilation system, perfect consistency in this regard is
569 presumably unattainable.

570 Another SMAP data subsetting difference involves the use of only descending (6AM)
571 passes for the model calibration experiments versus the use of both descending and ascending
572 (6AM and 6PM) passes in the data assimilation experiments. Here again this difference reflects
573 the specific needs of the two procedures; we used what we felt to be optimal for each. The 6AM
574 data on their own have an acceptable revisit interval (typically less than three days) to address
575 the timescale calibration problem, and these data have slightly better error characteristics (Chan
576 et al. 2018) than the 6PM data, making them desirable for capturing second order properties such
577 as drydown timescales. Data assimilation, on the other hand, does not focus on such second
578 order properties, considering both the 6AM and 6PM data as appropriate inputs to guide the
579 model states.

580 Also worth mentioning here are a number of limitations associated with the two
581 approaches, particularly when applied on the global scale. It is not possible to extract soil
582 moisture information from the SMAP data where the observed brightness temperatures are not
583 sensitive to soil moisture or have limited quality, such as in regions with dense vegetation or
584 strong radio-frequency interference (e.g., Japan). Model calibration, as employed in this study,
585 further requires suitably accurate meteorological forcing data (particularly precipitation
586 information) during the calibration period, and because the quality of meteorological information

587 is poor in many regions of the globe, calibration in these areas may prove difficult or even
588 impossible. The veracity of the dry-down time scales implied by the SMAP retrievals (Figure 1),
589 upon which the model calibration relies, may also be impacted by errors in the radiative transfer
590 model underlying the retrieval algorithm. More work is indeed needed to determine the impact
591 of errors in the SMAP retrievals on the effectiveness of model calibration.

592 Also worth pointing out is the seemingly small ubRMSE improvement in soil moisture
593 estimation obtained here with the two approaches – the largest improvement in Figure 4e is
594 associated with data assimilation, but this improvement amounts to only about 0.003 m^3m^{-3}.
595 Presumably this reflects to some degree the nature of the in-situ data networks examined. The
596 SMAP core validation sites (Reichle et al. 2017a), for example, have the relative advantage of
597 providing careful, spatially-distributed measurements that are more relevant to the spatial scales
598 of SMAP data. Reichle et al. (2017a) show that when the impacts of data assimilation (using the
599 same systems as used here) are quantified at the SMAP core validation sites rather than at sparse
600 network sites, the improvements in ubRMSE increase by a factor of about 2; furthermore,
601 quantified improvements in temporal correlation R are about twice as large for the core sites as
602 they are for sparse network sites. The core validation sites, however, have the distinct
603 disadvantage of being far fewer in number, leading to greater noise in the multi-site skill metrics.
604 Applying our analyses to the core sites instead (not shown) produces similar results that are
605 nevertheless affected by the smaller sample size. Here, for our joint analysis of data assimilation
606 and model calibration, we take advantage of the broader range of conditions covered by the
607 much more numerous SCAN and USCRN sites, accepting the disadvantage of the point-scale
608 nature of the measurements at these sites.

609 The results of the present study – in particular the demonstrated complementarity of the
610 data assimilation and model calibration approaches – have important implications for
611 maximizing the effective utilization of SMAP data in hydrological simulation. For maximum
612 simulation accuracy, use of both approaches should be considered, since they each effectively
613 access independent information contained within the SMAP data. Indeed, the two approaches
614 together underline the wealth of hydrological information inherent in these data. The full
615 hydrological information content of the SMAP data record – accessed through these two
616 approaches or through other approaches not described here – will undoubtedly be easier to
617 ascertain as the data record grows with time.

618

619 *Acknowledgments*.

620 Funding for this work was provided by the NASA SMAP mission. Computational resources
621 were provided by the NASA High-End Computing program through the NASA Center for
622 Climate Simulation. We are grateful for the datasets and data archiving centers that supported
623 this work and appreciate those who make the generation and dissemination of the SMAP data
624 products and various validation datasets possible, including SMAP team members at JPL, GSFC,
625 and NSIDC and staff at NOAA CPC, NOAA NCEI, USDA NRCS, and USGS.

626

627 **References**

628

629 Bell, J., and Coauthors, 2013: U.S. Climate Reference Network soil moisture and temperature
630 observations. J. Hydrometeor., 14, 977–988, doi:10.1175/JHM-D-12-0146.1.

631 Boone, A., F. Habets, J. Noilhan, and coauthors 2004: The Rhone-Aggregation Land Surface
632 Scheme Intercomparison Project: An Overview, *J. Climate*, **17**, 187-208.

633 Bowling L. C., D. P. Lettenmaier, B. Nijssen and coauthors 2003: Simulation of high latitude
634 hydrological processes in the Torne-Kalix basin: PILPS Phase 2e 1: Experiment
635 description and summary intercomparison, *Global and Planetary Change*, **38**, 1-30.

636 Brodzik, M. J., B. Billingsley, T. Haran, B. Raun, and M. H. Savoie, 2012: EASE-Grid 2.0:
637 Incremental but Significant Improvements for Earth-Gridded Data Sets. ISPRS
638 International Journal of Geo-Information, 1(1): 32-45.

639 Carrera, M. L., S. Bélair, and B. Bilodeau, 2015: The Canadian Land Data Assimilation System
640 (CaLDAS): Description and Synthetic Evaluation Study. J. Hydrometeor., 16, 1293–
641 1314, doi:10.1175/JHM-D-14-0089.1.

642 Chan, S. K., and Coauthors, 2016a: Assessment of the SMAP Passive Soil Moisture Product.
643 IEEE Transactions on Geoscience and Remote Sensing, 54, 4994-5007,
644 doi:10.1109/TGRS.2016.2561938.

645 Chan, S., E. G. Njoku, and A. Colliander, 2016b: SMAP L1C Radiometer Half-Orbit 36 km
646 EASE-Grid Brightness Temperatures, Version 3. Boulder, Colorado USA. NASA
647 National Snow and Ice Data Center Distributed Active Archive Center.
648 doi:10.5067/E51BSP6V3KP7. Last accessed 5 April 2017.

649 Chan, S. K., and Co-authors (2018), Development and assessment of the SMAP enhanced
650 passive soil moisture product, Remote Sens. Env., 204, 931-941.

651 Crow, W. T., E. F. Wood, and M. Pan, 2003: Multiobjective calibration of land surface model
652 evapotranspiration predictions using streamflow observations and spaceborne surface
653 radiometric temperature retrievals. J. Geophys. Res., 108(D23), 4725,
654 doi:10.1029/2002JD003292.

655 De Lannoy, G. J. M., R. H. Reichle, and V. R. N. Pauwels, 2013: Global calibration of the
656 GEOS-5 L-band microwave radiative transfer model over nonfrozen land using SMOS
657 observations. J. Hydrometeor., 14, 765–785, doi: 10.1175/JHM-D-12-092.1.

658 De Lannoy, G. J. M., and R. H. Reichle, 2016a: Global Assimilation of Multiangle and
659 Multipolarization SMOS brightness temperature Observations into the GEOS-5
660 Catchment Land Surface Model for Soil Moisture Estimation. Journal of
661 Hydrometeorology, 17, 669-691, doi:10.1175/JHM-D-15-0037.1.

662 De Lannoy, G. J. M., and R. H. Reichle, 2016b: Assimilation of SMOS brightness temperatures
663 or Soil Moisture Retrievals into a Land Surface Model. Hydrology and Earth System
664 Sciences, 20, 4895-4911, doi:10.5194/hess-20-4895-2016.

665 de Rosnay, P., M. Drusch, D. Vasiljevic, G. Balsamo, C. Albergel, and L. Isaksen, 2013: A
666 simplified Extended Kalman Filter for the global operational soil moisture analysis at
667 ECMWF. Q. J. R. Meteorol. Soc., 139, 1199-1213, doi:10.1002/qj.2023.

668 Diamond, H., and Coauthors, 2013: U.S. Climate Reference Network after one decade of
669 operations: Status and assessment. Bull. Amer. Meteor. Soc., 94, 485–498,
670 doi:10.1175/BAMS-D-12-00170.1.

671 Draper, C. S., R. H. Reichle, G. J. M. De Lannoy, and Q. Liu (2012), Assimilation of passive and
672 active microwave soil moisture retrievals, Geophysical Research Letters, 39, L04401,
673 doi:10.1029/2011GL050655.

674 Drusch, M., and Coauthors, 2009: Towards a Kalman Filter based soil moisture analysis system
675 for the operational ECMWF Integrated Forecast System. Geophys. Res. Lett., 36,
676 L10401, doi:10.1029/2009GL037716.

677 Ducharne A, R. D. Koster, M. J. Suarez, A. M. Stieglitz, and P. Kumar 2000: A catchment-based
678 approach to modeling land surface processes in a general circulation model. (2)
679 Parameter estimation and model demonstration. J Geophys Res 2000;105(D20):24823–
680 38.

681 Entekhabi, D., and Co-authors, 2010a: The Soil Moisture Active Passive (SMAP) mission. Proc.
682 IEEE, 98, 704-716.

683 Entekhabi, D., R. H. Reichle, R. D. Koster, and W. T. Crow, 2010b: Performance metrics for soil
684 moisture retrievals and application requirements. J. Hydromet., 11, 832-840.

685 Evensen, G. 2003. TheEnsemble Kalman Filter: Theoretical Formulation and Practical
686 Implementation. Ocean Dynamics 53 (4): 343-367. doi:10.1007/s10236-003-0036-9.

687 Gelaro, R., and Co-authors, 2017: The Modern-Era Retrospective Analysis for Research and
688 Applications, Version 2 (MERRA-2). J. Climate, 30, 5419-5454.

689 Gutmann, E. D., and E. E. Small, 2010: A method for the determination of the hydraulic
690 properties of soil from MODIS surface temperature for use in land-surface models.
691 Water Resour. Res., 46, W06520, doi:10.1029/2009WR00820.

692 Kerr, Y. H., and Co-authors, 2010: The SMOS mission: New tool for monitoring key elements
693 of the global water cycle. *Proc. IEEE*, 98, doi:10.1109/JPROC.2010.2043032.

694 Kerr, Y. H., and Coauthors, 2016: Overview of SMOS performance in terms of global soil
695 moisture monitoring after six years in operation. Remote Sensing of Environment, 180,
696 40-63, doi:10.1016/j.rse.2016.02.042.

697 Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2000: A catchment-based
698 approach to modeling land surface processes in a general circulation model: 1. Model
699 structure, J. Geophys. Res., 105(20), 24,809– 24,822.

700 Koster, R. D., and S. P. P. Mahanama, 2012: Land surface controls on hydroclimatic means and
701 variability. J. Hydromet., 13, 1604-1620.

702 Kumar, S. V., R. H. Reichle, C. D. Peters-Lidard, R. D. Koster, X. Zhan, W. T. Crow, J. B.
703 Eylander, and P. R. Houser, 2008: A Land Surface Data Assimilation Framework using
704 the Land Information System: Description and Applications. Advances in Water
705 Resources, 31, 1419-1432, doi:10.1016/j.advwatres.2008.01.013.

706 Kumar, S. V., R. H. Reichle, R. D. Koster, W. T. Crow, and C. D. Peters-Lidard, 2009: Role of
707 subsurface physics in the assimilation of surface soil moisture observations. J.
708 Hydromet., 10, 1534-1547.

709 Kumar, S., Christa D. Peters-Lidard, David Mocko, Rolf Reichle, Yuqiong Liu, Kristi R.
710 Arsenault, Youlong Xia, Michael Ek, George Riggs, Ben Livneh, and Michael Cosh,
711 2014: Assimilation of remotely sensed soil moisture and snow depth retrievals for
712 drought estimation. J. Hydrometeor., 15, 2446–2469, doi:https://doi.org/10.1175/JHM-D-
713 13-0132.1.

714 Kumar, S., C. D. Peters-Lidard, J. A. Santanello, R. H. Reichle, C. S. Draper, R. D. Koster, G.
715 Nearing, and M. F. Jasinski, 2015: Evaluating the utility of satellite soil moisture
716 retrievals over irrigated areas and the ability of land data assimilation methods to correct
717 for unmodeled processes. Hydrol. Earth Syst. Sci., 19, 4463–4478, doi:10.5194/hess-19-
718 4463-2015.

719 McColl, K. A., W. Wang, B. Peng, R. Akbar, D. J. Gianotti, H. Lu, M. Pan, and D. Entekhabi,
720 2017: Global characterization of surface soil moisture drydowns. Geophys. Res. Lett.,
721 44, 3682-3690, doi:10.1002/2017GL072819.

722 O'Neill, P. E., S. Chan, E. G. Njoku, T. Jackson, and R. Bindlish. 2016: SMAP L2 Radiometer
723 Half-Orbit 36 km EASE-Grid Soil Moisture, Version 4. Boulder, Colorado USA. NASA
724 National Snow and Ice Data Center Distributed Active Archive Center. doi:
725 10.5067/XPJTJT812XFY. Accessed October 2016.

726 Piepmeier, J. R., and Coauthors, 2014: Radio-Frequency Interference Mitigation for the Soil
727 Moisture Active Passive Microwave Radiometer. IEEE Transactions on Geoscience and
728 Remote Sensing, 52, 761-775, doi:10.1109/TGRS.2013.2281266.

729 Piepmeier, J. R., and Coauthors, 2017: SMAP L-Band Microwave Radiometer: Instrument
730 Design and First Year on Orbit. IEEE Transactions on Geoscience and Remote Sensing,
731 55, 1954-1966, doi:10.1109/TGRS.2016.2631978.

732 Reichle, R. H., D. B. McLaughlin, and D. Entekhabi, 2002a: Hydrologic data assimilation with
733 the Ensemble Kalman filter. Monthly Weather Review, 130, 103-114, doi:10.1175/1520-
734 0493(2002)130<0103:HDAWTE>2.0.CO;2.

735 Reichle, R. H. (2008), Data Assimilation Methods in the Earth Sciences, Advances in Water
736 Resources, 31, 1411-1418, doi:10.1016/j.advwatres.2008.01.001.

737 Reichle, R. H., R. D. Koster, G. J. M. De Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama, and
738 A. Toure, et al., 2011: Assessment and enhancement of MERRA land surface hydrology
739 estimates. *J. Climate*, **24**, 6322-6338.

740 Reichle, R. H., and Co-authors, 2017a. Assessment of the SMAP Level-4 Surface and Root-Zone
741 Soil Moisture Product Using In Situ Measurements. J. Hydrometeorl, 18, 2621-2645,
742 doi:10.1175/JHM-D-17-0063.1.

743 Reichle, R. H., and Co-authors, 2017b: Global assessment of the SMAP Level-4 surface and root
744 zone soil moisture product using assimilation diagnostics. J. Hydromet., in press.

745 Reichle, R. H., C. S. Draper, Q. Liu, M. Girotto, S. P. Mahanama, R. D. Koster, and G. J. De
746 Lannoy, 2017c: Assessment of MERRA-2 Land Surface Hydrology Estimates. J.
747 Climate, 30, 2937–2960, doi:10.1175/JCLI-D-16-0720.1

748 Schaefer, G. L., M. H. Cosh, and T. J. Jackson, 2007: The USDA Natural Resources
749 Conservation Service Soil Climate Analysis Network (SCAN). J. Atmos. Oceanic
750 Technol., 24, 2073– 2077.

751 Shellito, P. J., and Co-authors, 2016: SMAP soil moisture drying more rapid than observed in
752 situ following rainfall events. Geophys. Res. Lett., 43, 8068-8075,
753 doi:10.1002/2016GL069946.

754

Figure Captions

Figure 1. Time series of precipitation (a), SMAP Level-2 soil moisture retrievals (b), and model-simulated near-surface soil moisture (c,d) at a grid cell centered on the Little Washita watershed in southwestern Oklahoma. The model simulated time series in (c) uses a value of 1 for the calibrated parameter α (the default value in GEOS systems), and the time series in (d) uses an α value of 0.01. For ease of comparison, model values are plotted only on the dates of SMAP retrievals.

Figure 2. Optimal values of the studied model parameter α (dimensionless), as determined by optimization against SMAP retrieval time series. These are the values used in the OPT and OPT_DA simulations; in the BL and BL_DA simulations, α is set to 1 everywhere.

Figure 3. Time series (February – July 2016) of daily-averaged soil moisture (in m^3m^{-3}) at the EASE grid cell containing the Prairie View SCAN measurement site located in Texas, USA. The observations are shown as the heavy black curves in both panels. The top panel shows the simulated time series from BL (blue) and BL_DA (red), while the bottom panel shows the simulated time series from OPT (blue) and OPT_DA (red). For ease of comparison, the mean simulated bias February – July 2016 was removed from all simulation results prior to plotting. The BL and BL_DA simulations used the default value of 1 for α, whereas the OPT and OPT_DA simulations used an α of 0.001, the optimized value for this location.

Figure 4. (a) Spatial distribution of the ubRMSE of surface soil moisture estimation for the baseline (BL) simulation. Circles refer to SCAN sites and triangles refer to USCRN sites. (b) Differences in ubRMSE: BL_DA minus BL. Blue shades indicate improved

778 near-surface soil moisture estimation. (c) Same, but for OPT minus BL. (d) Same, but
779 for OPT_DA minus BL. (e) Average ubRMSE across the CONUS sites for each
780 simulation (see text for details).

781 Figure 5. Same as Figure 4, but for the temporal correlation metric R. As in Figure 4, blue
782 shading in (b)-(d) indicates improvement, though here the blue shading indicates a
783 positive difference.

784 Figure 6. Same as Figure 4, but for bias (a) and for differences in the absolute values of the
785 biases (b-d). As in Figure 4, blue shading in (b)-(d) indicates improvement.

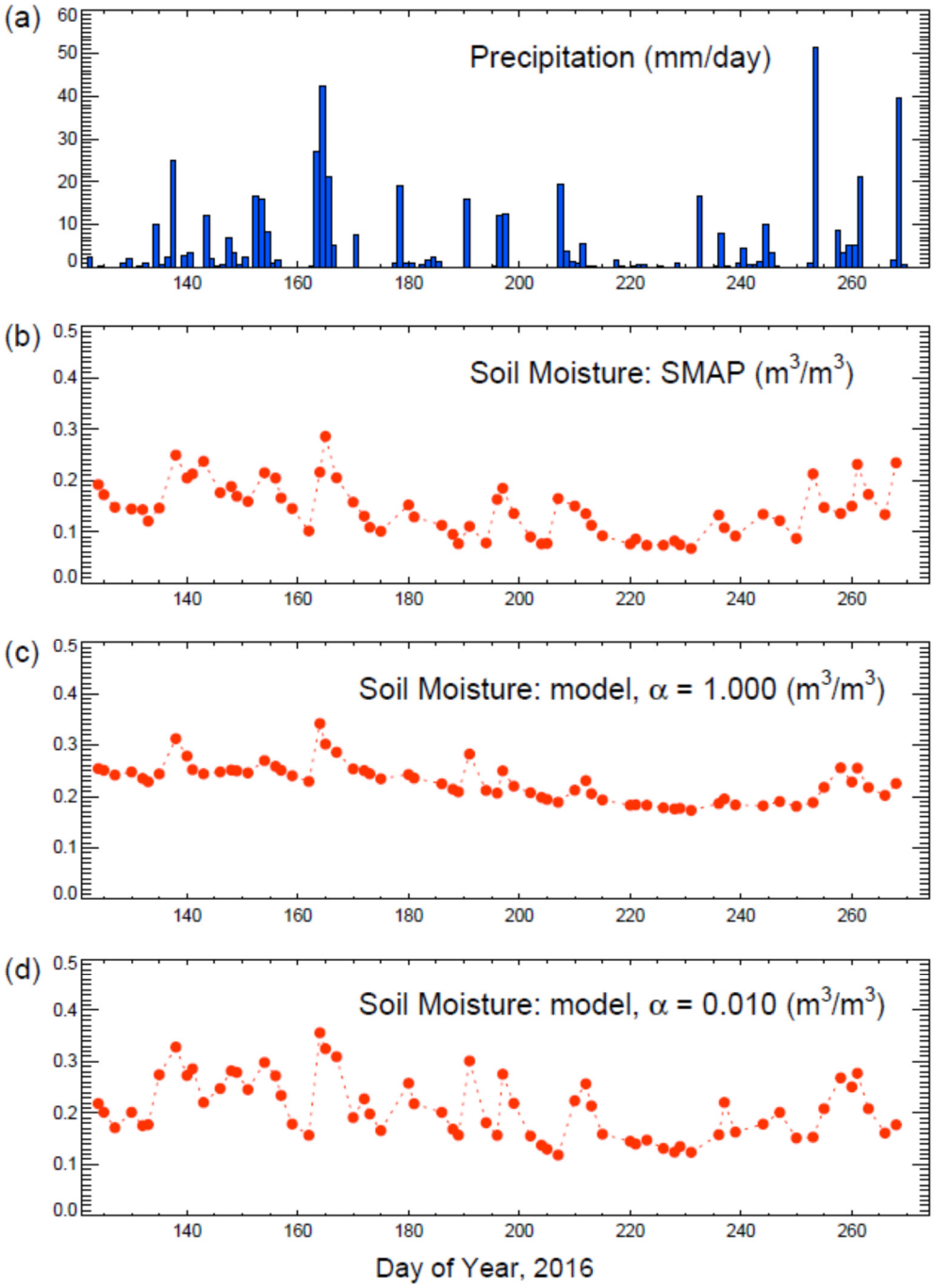
786 Figure 7. (a) The baseline (BL) simulation's bias in long-term average runoff ratio (ratio of total
787 2-year streamflow to total 2-year precipitation) in multiple unregulated basins. Values
788 are plotted at the centroids of the basins. (b) Differences in the absolute value of the bias:
789 BL_DA minus BL. Blue shades indicate an improved estimation of runoff ratio. (c)
790 Same as (b), but for OPT minus BL. (d) Same as (b), but for OPT_DA minus BL. (e)
791 Average absolute bias across the unregulated basins (see text for details).

792 Figure 8. Same as Figure 7, but for the temporal correlation R between observed and simulated
793 10-day streamflow totals in the warm season (April-September of 2015-2016) in multiple
794 unregulated basins. R values and R differences are plotted at the centroids of the basins.
795 As in Figure 4, blue shading in (b)-(d) indicates improvement.

796

797

798



799

800

801 Figure 1. Time series of precipitation (a), SMAP Level-2 soil moisture retrievals (b), and model-
802 simulated near-surface soil moisture (c,d) at a grid cell centered on the Little Washita watershed
803 in southwestern Oklahoma. The model simulated time series in (c) uses a value of 1 for the
804 calibrated parameter α (the default value in GEOS systems), and the time series in (d) uses an α
805 value of 0.01. For ease of comparison, model values are plotted only on the dates of SMAP
806 retrievals.

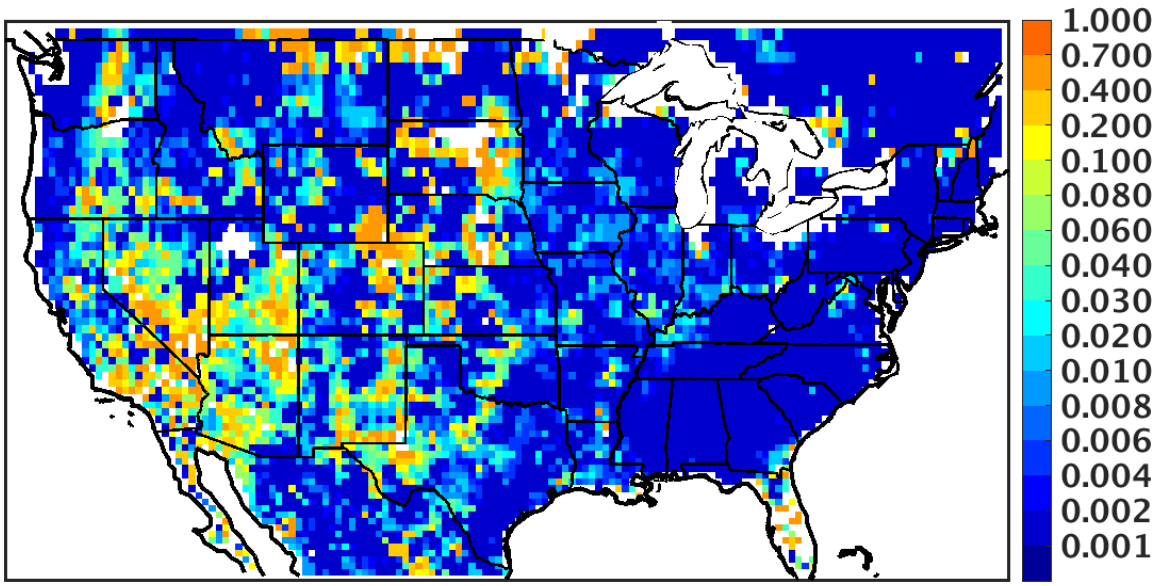


Figure 2. Optimal values of the studied model parameter α (dimensionless), as determined by optimization against SMAP retrieval time series. These are the values used in the OPT and OPT_DA simulations; in the BL and BL_DA simulations, α is set to 1 everywhere.

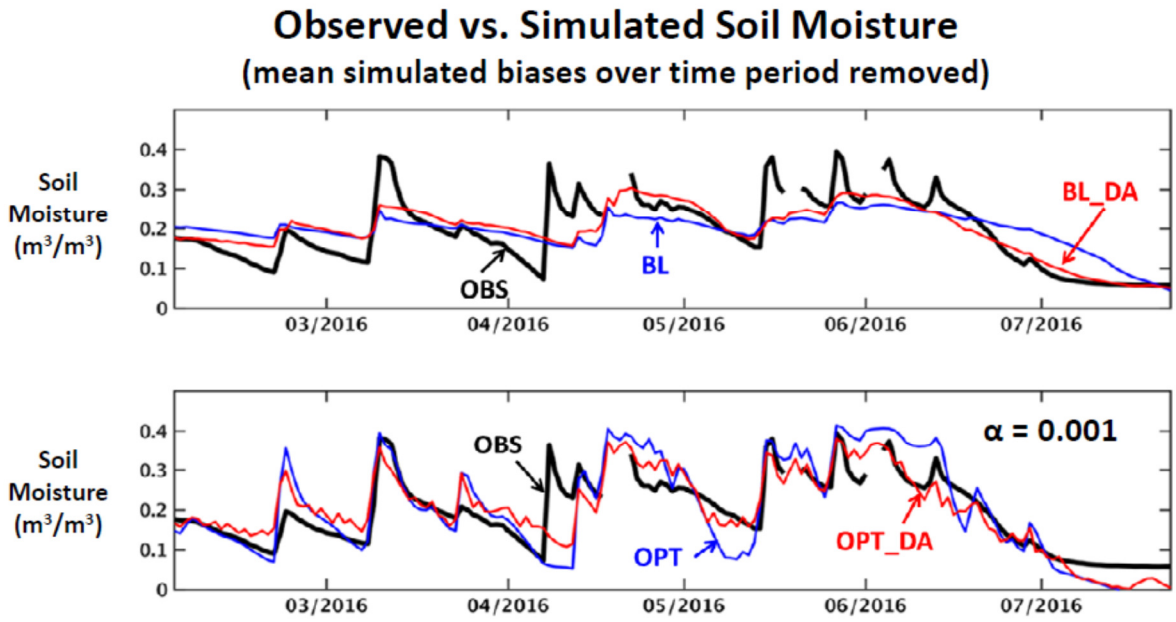


Figure 3. Time series (February – July 2016) of daily-averaged soil moisture (in m^3m^{-3}) at the EASE grid cell containing the Prairie View SCAN measurement site located in Texas, USA. The observations are shown as the heavy black curves in both panels. The top panel shows the simulated time series from BL (blue) and BL_DA (red), while the bottom panel shows the simulated time series from OPT (blue) and OPT_DA (red). For ease of comparison, the mean simulated bias for February – July 2016 was removed from all simulation results prior to plotting. The BL and BL_DA simulations used the default value of 1 for α, whereas the OPT and OPT_DA simulations used an α of 0.001, the optimized value for this location.

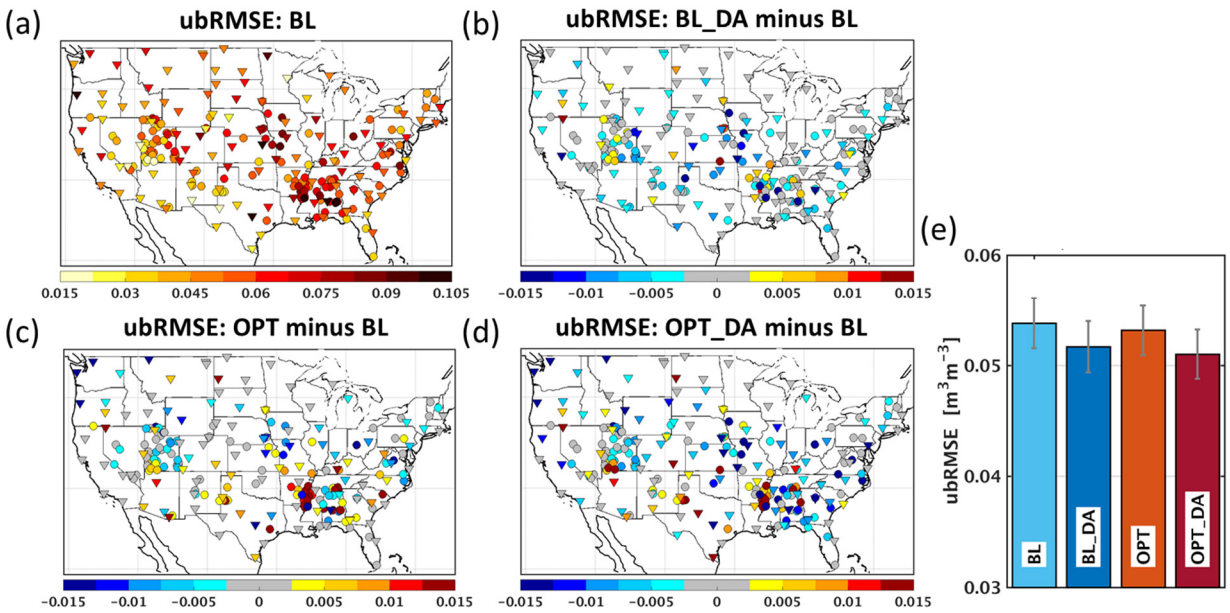


Figure 4. (a) Spatial distribution of the ubRMSE of surface soil moisture estimation for the baseline (BL) simulation. Circles refer to SCAN sites and triangles refer to USCRN sites. (b) Differences in ubRMSE: BL_DA minus BL. Blue shades indicate improved near-surface soil moisture estimation. (c) Same as (b), but for OPT minus BL. (d) Same as (b), but for OPT_DA minus BL. (e) Average ubRMSE across the CONUS sites for each simulation (see text for details).

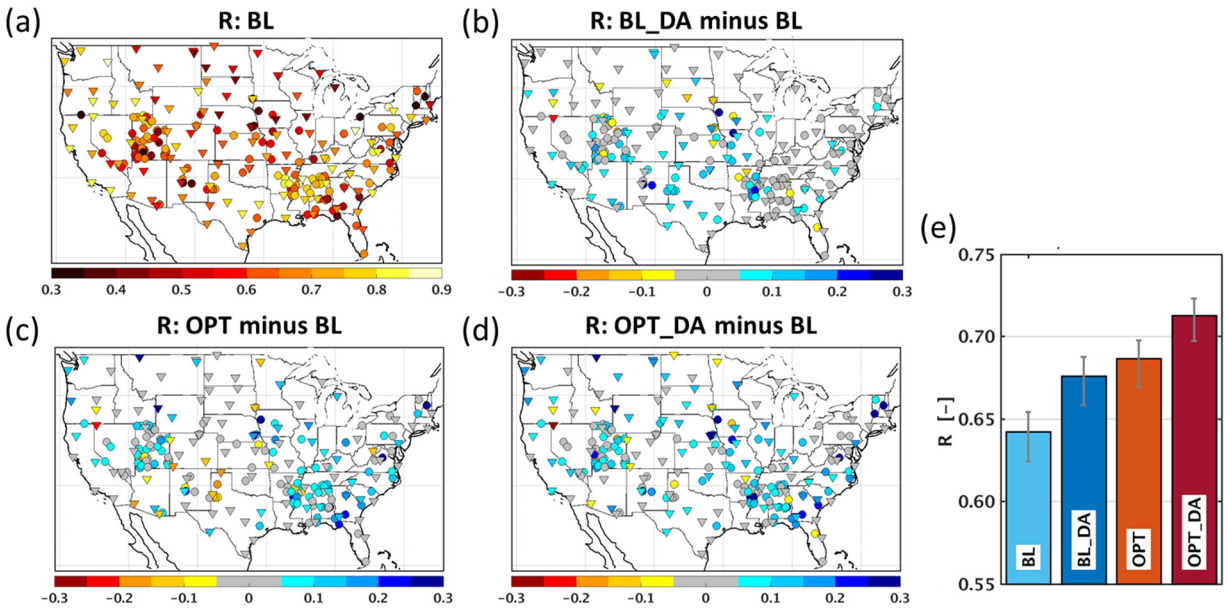


Figure 5. Same as Figure 4, but for the temporal correlation metric R. As in Figure 4, blue shading in (b)-(d) indicates improvement, though here the blue shading indicates a positive difference.

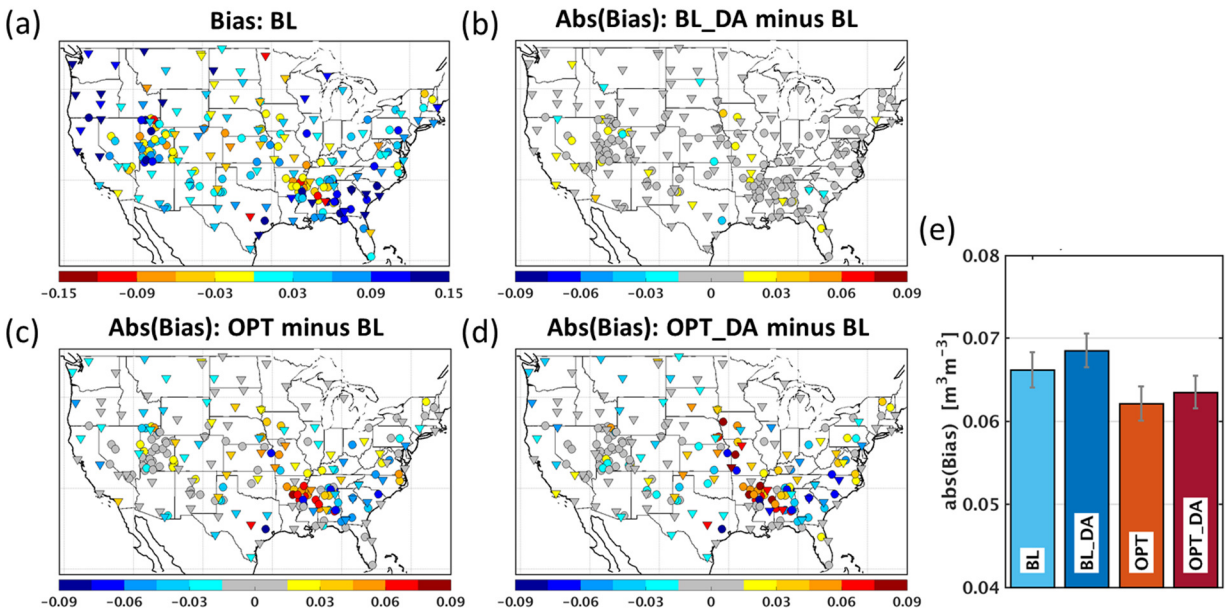


Figure 6. Same as Figure 4, but for bias (a) and for differences in the absolute values of the biases (b-d). As in Figure 4, blue shading in (b)-(d) indicates improvement.

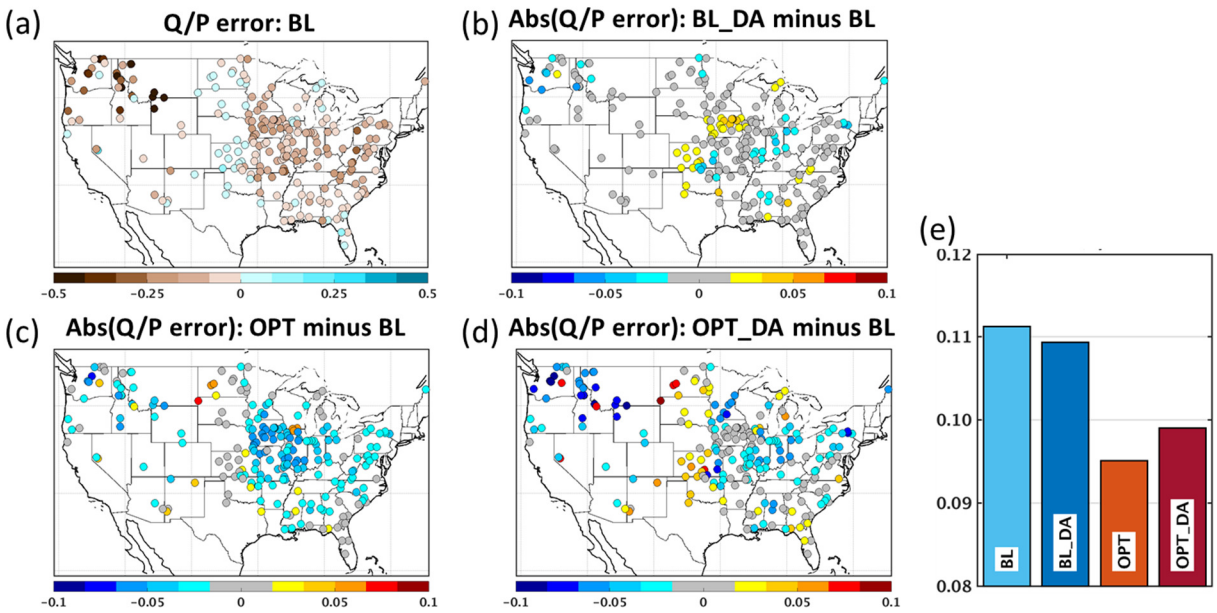


Figure 7. (a) The baseline (BL) simulation's bias in long-term average runoff ratio (ratio of total 2-year streamflow to total 2-year precipitation) in multiple unregulated basins. Values are plotted at the centroids of the basins. (b) Differences in the absolute value of the bias: BL_DA minus BL. Blue shades indicate an improved estimation of runoff ratio. (c) Same as (b), but for OPT minus BL. (d) Same as (b), but for OPT_DA minus BL. (e) Average absolute bias across the unregulated basins (see text for details).

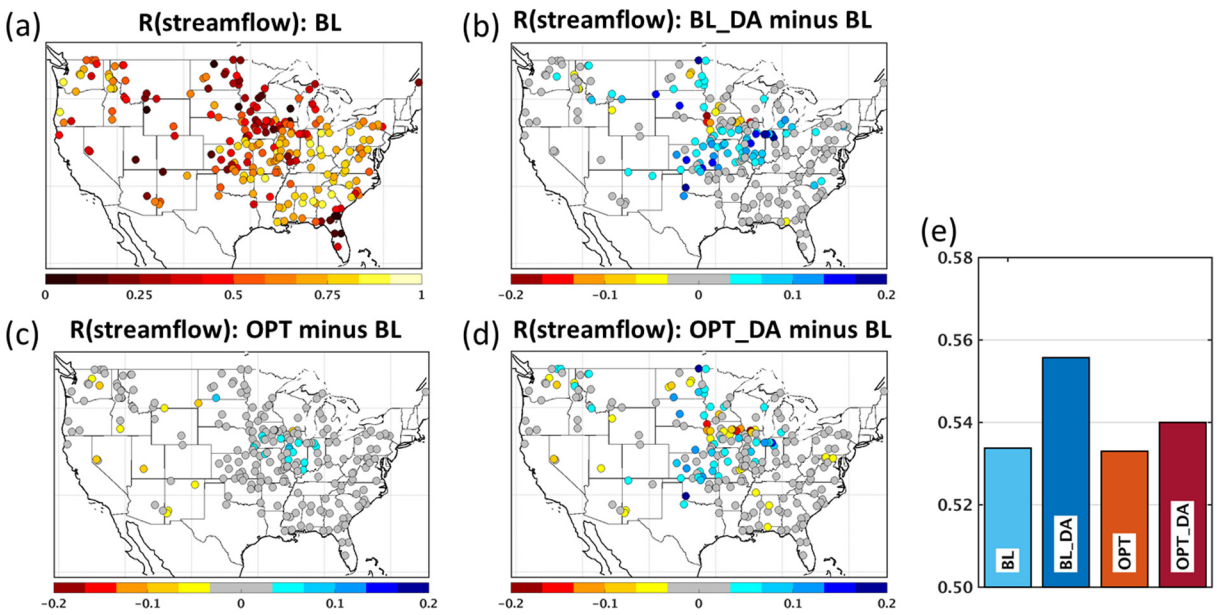


Figure 8. Same as Figure 7, but for the temporal correlation R between observed and simulated 10-day streamflow totals in the warm season (April-September of 2015-2016) in multiple unregulated basins. R values and R differences are plotted at the centroids of the basins. As in Figure 4, blue shading in (b)-(d) indicates improvement.