Water Resources Research

RESEARCH ARTICLE
10.1029/2019WR025245

# Diagnosing Bias in Modeled Soil Moisture/Runoff Coefficient Correlation Using the SMAP Level 4 Soil Moisture Product

W. T. Crow[1], F. Chen[1,2], R. H. Reichle[3], and Y. Xia[4]

[1]USDA Hydrology and Remote Sensing Laboratory, Beltsville, MD, USA, [2]SSAI Inc., Greenbelt, MD, USA, [3]NASA GSFC Global Modeling and Assimilation Office, Greenbelt, MD, USA, [4]I.M. Systems Group at NCEP EMC, College Park, MD, USA

**Key Points:**
- Correlation between prestorm surface soil moisture and event-scale runoff coefficient is an important diagnostic for land surface models
- Improved model representation of this correlation is associated with more skillful estimation of storm-to-storm variations in runoff
- Satellite-based soil moisture analysis products can be applied to detect bias in model-based correlation and correct for its impact

**Correspondence to:**
W. T. Crow,
wade.crow@ars.usda.gov

**Citation:**
Crow, W. T., Chen, F., Reichle, R. H., & Xia, Y. (2019). Diagnosing bias in modeled soil moisture/runoff coefficient correlation using the SMAP Level 4 soil moisture product. *Water Resources Research*, *55*, 7010–7026. https://doi.org/10.1029/2019WR025245

Received 28 MAR 2019
Accepted 29 JUL 2019
Accepted article online 04 AUG 2019
Published online 20 AUG 2019



**Abstract** The physical parameterization of key processes in land surface models (LSMs) remains uncertain, and new techniques are required to evaluate LSMs accuracy over large spatial scales. Given the role of soil moisture in the partitioning of surface water fluxes (between infiltration, runoff, and evapotranspiration), surface soil moisture (SSM) estimates represent an important observational benchmark for such evaluations. Here, we apply SSM estimates from the NASA Soil Moisture Active Passive Level-4 product (SMAP_L4) to diagnose bias in the correlation between SSM and surface runoff for multiple Noah-Multiple Physics (Noah-MP) LSM parameterization cases. Results demonstrate that Noah-MP surface runoff parameterizations often underestimate the correlation between prestorm SSM and the event-scale runoff coefficient (RC; defined as the ratio between event-scale streamflow and precipitation volumes). This bias can be quantified against an observational benchmark calculated using streamflow observations and SMAP_L4 SSM and applied to explain a substantial fraction of the observed basin-to-basin (and case-to-case) variability in the skill of event-scale RC estimates from Noah-MP. Most notably, a low bias in LSM-predicted SSM/RC correlation squanders RC information contained in prestorm SSM and reduces LSM RC estimation skill. Based on this concept, a novel case selection strategy for ungauged basins is introduced and demonstrated to successfully identify poorly performing Noah-MP parameterization cases.

**Plain Language Summary** Land surface models are commonly tasked with determining what fraction of incoming rainfall infiltrates into the soil versus runs off into stream channels. The key factor determining this partitioning is the amount of water in the soil column prior to a storm event (e.g., more prestorm soil moisture is generally associated with decreased amounts of infiltration and increased surface runoff). However, due to a lack of soil moisture observations available at large scales, it has generally been difficult to assess whether existing models are accurately capturing the true relationship between prestorm soil moisture and runoff. Using newly available data from the NASA Soil Moisture Active/Passive (SMAP) mission, this paper demonstrates that land surface models often misrepresent the impact of prestorm surface soil moisture on runoff generation. This misrepresentation is shown to have a strong negative impact on the ability of models to accurately estimate runoff. A new calibration technique, based on the SMAP Level 4 soil moisture product, is introduced for eliminating this bias. Overall, results demonstrate how remotely sensed soil moisture can potentially play an important role in enhancing the operational forecasting of streamflow.

## 1. Introduction

Serious questions remain concerning the characterization of key processes in land surface models (LSMs). For example, the representation of surface runoff generation in LSMs is often highly parameterized and does not generally reflect our best understanding of physical processes responsible for the generation of runoff (Clark et al., 2015; Lohmann et al., 1998; Yang et al., 2011; Zheng et al., 2017). Shortcomings in model physics, parameter selection, and forcing data manifest themselves in the relatively limited ability of LSMs to predict streamflow at short time scales (Xia, Mitchell, Ek, Cosgrove, et al., 2012). Despite these concerns, LSMs are increasingly being utilized for operational hydrologic forecasting. For example, the National Water Model (NWM), currently in development as the next-generation hydrological forecasting system in the United States, relies on the Noah-Multiple Physics (Noah-MP) LSM to estimate surface and subsurface runoff fluxes used as the basis for its streamflow forecasts (Salas et al., 2018).

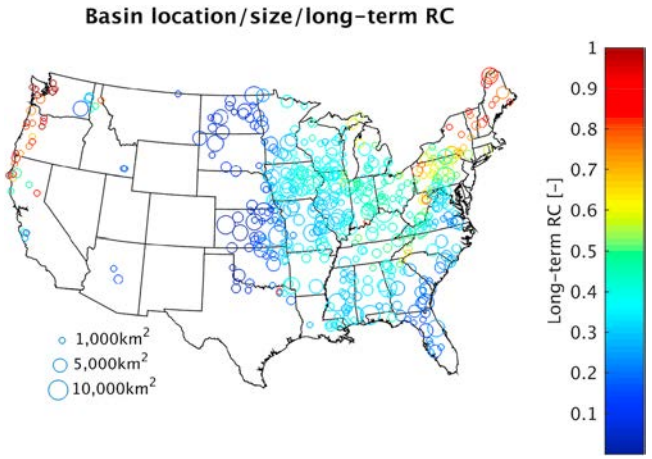


**Figure 1.** Basin centroid location, basin area [km$^2$] and long-term RC [-] (i.e., annual mean streamflow divided by annual mean precipitation) for all 522 medium-scale (500 to 10,000 km$^2$) basins. Long-term RC values are based on USGS streamflow and NLDAS-2 precipitation observations. RC = runoff coefficient; USGS = United States Geological Survey; NLDAS-2 = North American Land Data Assimilation System project phase 2.

The relationship between soil moisture and runoff generation lies at the heart of LSM water balance parameterizations (Koster & Milly, 1997; Meyles et al., 2003; Penna et al., 2011). From a hydrologic forecasting perspective, the key issue is the degree to which variations in prestorm surface soil moisture (SSM) impact the evolution of runoff during a storm event and subsequent values of event-scale runoff coefficients (RC; i.e., the fraction of total precipitation volume during a storm event converted into streamflow). To date, the direct evaluation of this partitioning has been difficult due to shortcomings in the availability and accuracy of large-scale SSM products. However, despite their shallow vertical support (generally less than 5 cm), recent advances in remote sensing have yielded new SSM products with significant information content for hydrologic forecasting (Crow et al., 2017; Koster et al., 2018). This, in turn, has enhanced our ability to critique the representation of SSM/RC coupling in LSMs.

Crow et al. (2017) and Koster et al. (2018) reflect a notable trend toward utilizing remotely sensed SSM products to evaluate the parameterization of hydrologic processes in LSMs. This goes beyond the typical use of satellite-based SSM to improve model estimates through the minimization of random errors, generally associated with uncertain external forcing data, via data assimilation (e.g., De Lannoy & Reichle, 2016). Using SSM from the NASA Soil Moisture Active Passive Level-4 product (SMAP_L4), Crow et al. (2018) demonstrated that LSMs tend to underestimate the correlation between prestorm SSM and event-scale RC. This implies that runoff parameterizations within LSMs place excessive weight on factors unrelated to prestorm SSM (e.g., fine-scale variations in precipitation intensity during a storm event) and neglect a portion of the predictive skill afforded by skillful estimates of prestorm SSM.

However, the analysis in Crow et al. (2018) was limited to 16 basins located in the south central contiguous United States (CONUS). In addition, its use of multiple LSMs, each with much different surface runoff parameterizations, soil layer structures, and forcing data, made it challenging to obtain reliable insights into the impact of model structure on the skill of modeled RC. For these reasons, Crow et al. (2018) stopped short of demonstrating a clear linkage between biases in LSM correlation and the reduction of skill in LSM surface runoff predictions. Here, we conduct a series of more systematic experiments involving a single LSM (Noah-MP) using multiple surface runoff parameterizations. The present analysis is based on a larger set of 522 medium-scale (500 to 10,000 km$^2$) basins covering the entire CONUS region. In addition, we directly trace the impact of bias in SSM/RC correlation on the ability of Noah-MP to accurately characterize event-scale RC variations. Therefore, our dual objectives are to (i) systematically assess the impact of LSM SSM/RC correlation bias on the accuracy of their event-scale RC estimates and (ii) evaluate the use of SMAP_L4 prestorm SSM for identifying Noah-MP parameterization cases that best represent the dependence of RC on SSM.

## 2. Approach

The geographic domain for the analysis consists of the 522 medium-scale (500–10,000 km$^2$) basins shown in Figure 1. Using criteria established during the Model Parameter Estimation Experiment [MOPEX; Duan et al., 2006; Schaake et al., 2006], basins were selected for (i) acceptable levels of anthropogenic impoundment or diversion, (ii) adequate rain gauge coverage (i.e., greater than $0.6A^{0.3}$ gauges per basin where $A$ [km$^2$] is basin area; Schaake et al., 2006) and (iii) the availability of near-continuous United States Geologic Survey (USGS) streamflow observations at their outlet. Note that selected basins exhibit a wide range of long-term RC (i.e., mean annual steam flow divided by mean annual precipitation) and thus span a variety of hydroclimatic conditions (Figure 1). Our experimental period was 31 March 2015 to 30 March 2018.

### 2.1. Storm Event Definition

Within the experimental period, storm events were separated out using the approach of Crow et al. (2017). In summary, each storm event started on a day with a precipitation accumulation exceeding 10 mm. The impact of other threshold values is discussed below. Storm events were assumed to last for an $N$-day period defined by rounding the saturation time scale for each basin, calculated using the empirical expression given by Linsley et al. (1982), up to the nearest positive integer value:

$$N\ [\text{days}] = \text{CEIL}\left[(A{*}2.59)^{0.2}\right] \quad (1)$$

where $A$ is the area [km$^2$] of each basin. CEIL is the upward integer rounding function. In this way, $N$ is defined to be larger than each basin's saturation time scale. Derived values of $N$ range from 5 days for smaller (~500 km$^2$) basins to 8 days for the largest (~10,000 km$^2$) basins considered here. "Prestorm" SSM refers to the lowest end-of-day (i.e., 24 UTC or the closest available alternative), basin-averaged SSM for the 2-day period preceding a given storm event. Storm events interrupted by a new storm (i.e., another daily precipitation accumulation exceeding 10 mm day$^{-1}$) were masked, and a new event was assumed to begin coincident with this second rain event.

### 2.2. Definition of Event-Scale RC and Correlation Statistics

During storm events, precipitation and streamflow accumulation volumes were summed and divided (streamflow over precipitation) to calculate the event-scale RC. All daily precipitation was derived from the North American Land Data Assimilation System project phase 2 (NLDAS-2) forcing data set (Xia, Mitchell, Ek, Cosgrove, et al., 2012). Streamflow was acquired from both LSMs (i.e., Noah-MP and the NWM) and actual USGS streamflow observations at the outlet of each basin in Figure 1. See below for detailed descriptions of Noah-MP, the NWM, and the SMAP_L4 SSM data.

Storm events were also masked for the presence of snow cover and/or frozen soil (determined via Noah-MP simulations). Since the event-scale aggregation periods applied here ($N$) were, by construction, slightly larger than the basin saturation time scales, no runoff routing was performed. Therefore, unless noted otherwise, RC values were obtained directly from streamflow measurements or total (i.e., surface runoff plus baseflow) runoff estimates acquired from an LSM. Sensitivity results (described below) were also generated following the application of the USGS Hydrologic Separation [HYSEP; Sloto & Crouse, 1996] algorithm to isolate surface runoff in individual storm events. At least 15 qualifying storm events were available for each of the 522 basins during our 3-year analysis period between 31 March 2015 and 30 March 2018.

The analysis was based on sampling the temporal correlation between LSM estimates of prestorm SSM and event-scale RC ($R[\text{SSM}_{\text{LSM}}, \text{RC}_{\text{LSM}}]$)—reflecting internal LSM estimates of RC dependence on prestorm SMM, and between SMAP_L4 prestorm SSM estimates and USGS-observed RC ($R[\text{SSM}_{\text{SMAPL4}}, \text{RC}_{\text{USGS}}]$)—reflecting an external assessment of SSM/RC correlation between the two independently acquired estimates. Unless otherwise noted, all correlations were Spearman-type, where correlation was sampled after the transformation of each time series variables into ranks (i.e., relative position labels within each individual time series). This ensured that correlations were insensitive to the potential presence of monotonic nonlinearity in the relationship between event-scale RC and prestorm SSM (Crow et al., 2017). More robust, full mutual information calculations were not applied due to bias issues associated with calculating mutual information for small sample sizes (which, for the 3-year period considered here, were as small as 15 events per basin). It should be noted that sampled Spearman correlations measure two-way SSM/RC coupling strength since SSM impacts RC (via runoff physics) and RC feeds back onto SSM (via soil water balance considerations). However, our implicit assumption (verified below) is that the temporal correlation between SSM and RC is dominated by processes controlling the dependence of RC on SSM.

Likewise, our evaluation of LSM RC skill was restricted to the ability to capture relative event-to-event variations in RC—as summarized by the Spearman rank correlation coefficient between LSM-estimated and USGS-observed RC ($R[\text{RC}_{\text{LSM}}, \text{RC}_{\text{USGS}}]$). Such skill is a necessary *but not sufficient* requirement for a robust hydrologic forecasting system. It does not, for example, guarantee good peak-storm flow estimation nor is it sensitive to bias in LSM RC estimates. Nevertheless, it is an important metric for streamflow estimation. If biases are a concern, then the percentile rank of an RC estimate can be transformed into an absolute estimate using long-term histograms of event-scale RC for a given basin.

### 2.3. Noah-MP LIS Modeling

Within each basin, off-line, retrospective Noah-MP v3.6 simulations (Xia et al., 2017) were conducted using Version 7.2 of the NASA Land Information System (LIS; Kumar et al., 2006) and NLDAS-2 meteorological forcing data. NLDAS-2 forcing data consists of North American Regional Reanalysis variables for all fields except precipitation, which is instead derived via the use of hourly precipitation radar data to downscale daily totals derived from the real-time Climate Prediction Center rain gauge network (Xia, Mitchell, Ek, Sheffield, et al., 2012). Noah-MP simulations utilized a 15-min time step and the 0.125° NLDAS-2 spatial grid (Xia, Mitchell, Ek, Sheffield, et al., 2012). All presented results are for a 3-year period stretching from the start of SMAP observations on 31 March 2015. Noah-MP SSM values reflect the average soil moisture in the top 10-cm model layer of the vertical soil column. End-of-day (24 UTC) Noah-MP surface (0- to 10-cm) soil moisture and daily (0 to 24 UTC) total runoff (surface runoff + baseflow) were extracted and spatially resampled to generated daily, basin-scale time series for each study basin. Noah-MP RC estimates were derived as the event-scale ratio of accumulated streamflow volume normalized by accumulated precipitation volume. See above for details on how individual events were defined and isolated from continuous precipitation and streamflow time series data.

Separate Noah-MP simulations were generated for all four Noah-MP surface runoff parameterizations described in Niu et al. (2011). These parameterizations include: (i) a simplified groundwater (SIM GW) case, (ii) a simplified TOPMODEL (SIM TOP) case, (iii) a free-drainage (FD) lower-boundary assumption, and (iv) a simplified surface runoff parameterization taken from the Biosphere Atmosphere Transfer Scheme (BATS).

The SIM GW approach is based on the simplified groundwater modeling approach presented in Niu et al. (2007), where vertical recharge to an unconfined aquifer is estimated via a parameterization of Darcy's Law. Groundwater storage calculations are then used to derive the pixel-scale water table depth ($Z_{WT}$ [m]) that is converted into saturated surface fractional area ($F_{sat}$). Neglecting frozen surfaces (which are screened out of our analysis), this conversion is

$$F_{sat} = F_{sat,\,max}\ \exp[-0.50\, f\ (Z_{WT} - 2.0\ [\mathrm{m}])] \tag{2}$$

where $F_{sat,max}$ is the maximum fraction of surface saturation (hard coded to a unitless fixed value of 0.38), and $f\,[\mathrm{m}^{-1}]$ describes the exponential decay of soil hydraulic conductivity with depth. Surface runoff is estimated, via a saturation excess runoff mechanism, as $F_{sat}$ multiplied by precipitation (or dew or snowmelt) incident on the top of the soil column. The SIM GW approach also imposes a minimum water table depth of 1.5 m. Baseflow ($Q$) is assumed proportional to $\exp[-f\,(Z_{WT} - 2.0\ [\mathrm{m}])]$.

SIM TOP runoff calculations are analogous to the SIM GW case except for two differences. First, $Z_{WT}$ is calculated using an equilibrium water table calculation instead of a dynamic groundwater water balance (Niu et al., 2005, 2011). Second, no minimum value is enforced for water table depths. Hence, $F_{sat}$ and $Q$ calculations are assumed to be directly proportional to $\exp[-0.50\, f\,(Z_{WT})]$ and $\exp[-f(Z_{WT} - 2.0\ [\mathrm{m}])$, respectively.

In contrast, both the FD and BATS parameterizations do not capture water table dynamics and instead employ a gravitational free-drainage baseflow approach as a bottom boundary condition. The two approaches differ in their conceptualization of surface runoff. The FD approach uses the infiltration excess surface runoff approach described by Schaake et al. (1996) based on an adaptation of the Soil Conservation Service curve number approach. The amount of surface runoff predicted by this approach is known to be sensitive to the specification of the Noah-MP parameter *REFKDT* (Niu & Yang, 2011). This parameter effectively controls the amount of weight applied to prestorm SSM conditions and is linearly related to the $K_{dt}$ parameter described by Schaake et al. (1996). The BATS physics package follows Yang and Dickinson (1996) and parameterizes the fraction of incident precipitation converted into runoff simply as the $q$th power of the degree of saturation in the top 2 m of the soil column.

For each surface runoff parameterization case described above, we selected (and subsequently varied) one key parameter. For the SIM GW and SIM TOP cases, it was the TOPMODEL $f$ parameter constraining the decay of saturated conductivity with depth. For the FD case, it was the *REFKDT* parameter modulating the impact of prestorm soil moisture conditions on surface runoff. For the BATS case, it was the exponential parameter ($q$) linking 2-m soil moisture and surface runoff. We then ran additional Noah-MP cases spanning the physically feasible range for each parameter. In this way, we generated 16 Noah-MP parameterization

**Table 1**
*Details for the 16 Noah-MP LIS Parameterization Cases Considered Here*

| Case number | Noah-MP runoff parameterization | Key parameter value |
|---|---|---|
| 1 | SIM GW | $f = 1.0$ [$m^{-1}$] |
| **2** | **SIM GW** | ***f* = 2.0 [$m^{-1}$]** |
| 3 | SIM GW | $f = 4.0$ [$m^{-1}$] |
| 4 | SIM GW | $f = 8.0$ [$m^{-1}$] |
| 5 | SIM TOP | $f = 3.0$ [$m^{-1}$] |
| **6** | **SIM TOP** | ***f* = 6.0 [$m^{-1}$]** |
| 7 | SIM TOP | $f = 12.0$ [$m^{-1}$] |
| 8 | SIM TOP | $f = 24.0$ [$m^{-1}$] |
| 9 | SIM TOP | $f = 48.0$ [$m^{-1}$] |
| 10 | FD | *REFKDT* = 0.5 [-] |
| **11** | **FD** | ***REFKDT* = 1.0 [-]** |
| 12 | FD | *REFKDT* = 3.0 [-] |
| 13 | FD | *REFKDT* =5.0 [-] |
| 14 | BATS | $q = 2$ [-] |
| **15** | **BATS** | ***q* = 4 [-]** |
| 16 | BATS | $q = 8$ [-] |

*Note*. Default cases (reflecting the use of LIS Noah-MP v3.6 recommended parameter values) for each parameterization are indicated with bold type.

cases (4 for SIM GW, 5 for SIM TOP, 4 for FD, and 3 for BATS). Table 1 numbers these cases and lists their associated parameter values. For the *f* parameter, values were modified by applying fixed multipliers (e.g., 0.5, 2.0, 4.0, and 8.0 [-]) to the recommended (i.e., hard-coded) *f* value for the Noah-MP SIM GW and SIM TOP cases. Note that the hard-coded *f* value for the SIM GW case (2 [$m^{-1}$]) is lower than that of the SIM TOP case (6 [$m^{-1}$]), which resulted in different *f* ranges applied for both cases (Table 1).

All Noah-MP LIS cases were spun up from a cold start on 1 January 2003 until the start of SMAP data availability, and our 3-year experiment period, on 31 March 2015. Spin-up adequacy was evaluated by dividing the last 12 years of this (combined spin-up plus the analysis) period into two distinct 6-year periods (i.e., 31 March 2006 to 31 March 2012 and 1 April 2012 to 30 March 2018) and examining differences in time-averaged Noah-MP SSM values sampled within both periods. For all parameterization cases listed in Table 1, the sign of these long-term SSM differences was the same as the sign of precipitation accumulation differences (sampled over the same period) for (at least) 79% of the study basins. This high percentage suggests that the relative impact of long-term, SSM transients (associated with the misspecification of initial SSM and water table conditions) is relatively small—since long-term SSM variations are adequately explained by coincident long-term differences in accumulated precipitation.

Due to the need to initialize a dynamic groundwater model, achieving adequate model spin-up was expected to be especially challenging for the Noah-MP SIM GW case. Therefore, we conducted an additional spin-up analysis where the SIM GW parameterization case was spun up with 27 years of data (instead of our baseline 12-year spin-up period). This (more than) doubling of the SIM GW spin-up period resulted in no noticeable impact on results—suggesting that, for this application, a 12-year spin-up period is acceptable.

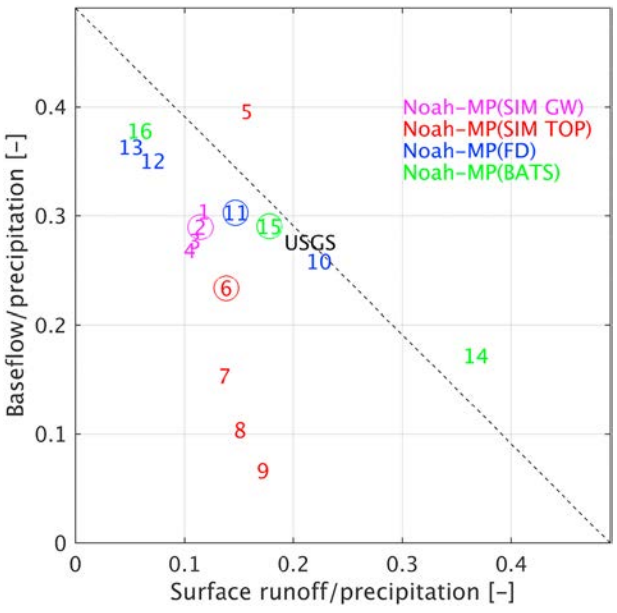


**Figure 2.** Long-term mean surface runoff versus baseflow partitioning (averaged across all 522 basins in Figure 1) for all Noah-MP surface runoff parameterization cases (see Table 1 for details). "USGS" results are based on the processing of USGS streamflow measurements within each basin through the USGS HYSEP baseflow separation package. Noah-MP cases using default parameter values are circled. The dashed line captures the magnitude of total runoff (baseflow + surface runoff) acquired from USGS observations. Noah-MP = Noah-Multiple Physics; USGS = United States Geological Survey; HYSEP = Hydrologic Separation; SIM GW = simplified groundwater; SIM TOP = simplified TOPMODEL; FD = free-drainage.

Figure 2 plots mean annual surface runoff and baseflow (both normalized by mean annual precipitation) averaged across all 522 basins for each of the 16 Noah-MP LIS cases summarized in Table 1. For comparison, observation-based surface runoff and baseflow fractions derived from USGS streamflow and the application of the USGS HYSEP baseflow separation procedure are also plotted. While there is a slight tendency to underestimate total runoff, all four baseline Noah-MP parameterizations (indicated by bold type in Table 1 and circled in Figure 2) produce reasonable bulk estimates of total surface runoff and baseflow. However, the variation of key parameters has a large impact on the relative partitioning of total runoff into surface and baseflow contributions and a smaller, but still considerable, impact on the magnitude of total (i.e., surface plus baseflow) runoff. While our decision to focus only on a single parameter for each Noah-MP surface runoff case is somewhat arbitrary, it should be noted that the resulting ensemble of Noah-MP runoff parameterization cases spans the physically realistic range of surface runoff/baseflow partitioning (Figure 2) and SSM/RC correlation (see discussion of Figure 3 below).

### 2.4. NWM Modeling

NWM soil moisture and runoff results were extracted from Version 1.2 of the NWM retrospective run generated between 31 March 2015 and 31 December 2017 on a 1-km spatial grid. A custom version of Noah-MP model based on the FD surface runoff case coupled with custom groundwater and overland flow modules forms the modeling core of

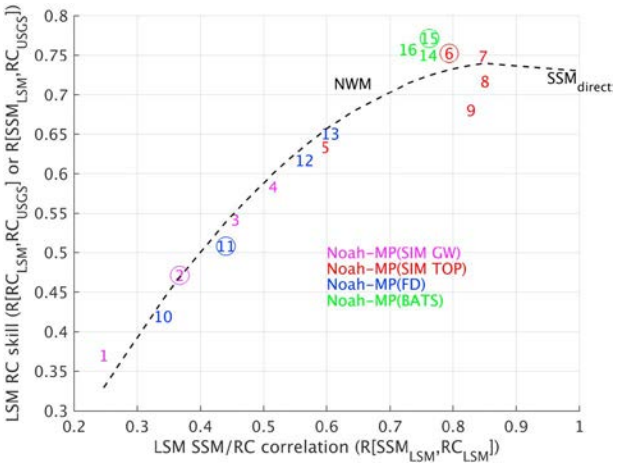


**Figure 3.** Skill of event-scale, LSM RC estimates ($R$[$RC_{LSM}$, $RC_{USGS}$] along the ordinate) as a function of LSM internal, event-scale SSM/RC correlation ($R$[$SSM_{LSM}$, $RC_{LSM}$] along the abscissa) for all 16 Noah-MP LIS parameterization cases in Table 1 and the NWM results (labeled as "NWM"). Also plotted are the results using Noah-MP SSM estimates as a proxy for RC (i.e., $R$[$SSM_{LSM}$, $RC_{USGS}$] along the ordinate; labeled as "$SSM_{direct}$"). The dashed black line indicates the LOcally WEighted Scatterplot Smoother (LOWESS) fit to all plotted results. 95% confidence sampling intervals for points are generally smaller than plotted point labels. Noah-MP LIS cases using default parameter values are circled. LSM = land surface model; RC = runoff coefficient; SSM = surface soil moisture; Noah-MP = Noah-Multiple Physics; NWM = National Water Model; LIS = Land Information System.

the NWM. Therefore, for the purposes of this analysis, the NWM can be considered functionally equivalent to the retrospective, off-line Noah-MP simulations described above. The only two differences being that NWM v1.2 soil moisture and runoff results were downloaded (http://edc.occ-data.org/nwm/getdata/), while Noah-MP results were custom-generated using LIS, and NWM simulations were generated at a finer spatial resolution (1-km versus 0.125° for Noah-MP LIS simulations).

As with the Noah-MP LIS cases, the off-line NWM retrospective simulation was forced using 0.125° NLDAS-2 observation forcing data (downscaled via linear interpolation onto a 1-km NWM grid cell). The retrospective NMW simulation was initialized on 1 January 1993; therefore, results from our period of interest (30 March 2015 and 31 December 2017) were derived after a 22-year model spin-up period. Unfortunately, the retrospective NWM simulation is not yet available for 2018, which forced us to utilize a slightly shorter 33-month analysis period (30 March 2015 to 31 December 2017) for NWM simulations relative to the full 36-month period (31 March 2015 to 30 March 2018) provided by the Noah-MP LIS simulations described above. In addition, since the NWM v1.2 retrospective run only archived total runoff results, it was not included in surface runoff/baseflow partitioning results shown above in Figure 2.

End-of-day (24 UTC) NWM surface (0- to 10-cm) soil moisture and daily total (0- to 24-UTC) runoff (surface runoff + baseflow) estimates were extracted and spatially resampled (from their native 1-km resolution) to generate a daily, basin-scale time series for each basin in Figure 1.

### 2.5. SMAP_L4 SSM

Basin-averaged, daily, 0- to 5-cm SSM values for the period 31 March 2015 to 30 March 2018 were extracted from the global, 3-hourly, 9-km-resolution SMAP_L4 Version 3 product (Reichle, de Lannoy, Koster, Crow, et al., 2017). Note that this definition of SSM is slightly shallower than the 0- to 10-cm depth applied for the Noah-MP LIS and NWM cases. The impact of this discrepancy will be discussed below. The SMAP_L4 product is based on the assimilation of SMAP brightness temperatures (Piepmeier et al., 2017) into the NASA Catchment LSM using an ensemble-based data assimilation system (Reichle, de Lannoy, Liu, Ardizzone, et al., 2017; Reichle, de Lannoy, Liu, Koster, et al., 2017). The Catchment LSM is driven with surface meteorological forcing data from the Goddard Earth Observing System Forward-Processing (GEOS-FP) product [https://gmao.gsfc.nasa.gov/GMAO_products/; Lucchesi, 2013]. Over CONUS, the GEOS-FP precipitation forcing is modified to match the precipitation from the NOAA Climate Prediction Center product at its 0.5°, daily scale.

Prior to the start of the SMAP data period in March 2015, the Catchment LSM was spun up from 1 January 1980 using Modern-Era Retrospective Analysis for Research and Applications, Version-2 (Gelaro et al., 2017). For further details on the SMAP_L4 product and validation results, see (Reichle, de Lannoy, Liu, Ardizzone, et al., 2017; Reichle, de Lannoy, Liu, Koster, et al., 2017). For consistency with the end-of-day (24-UTC) SSM output acquired from Noah-MP LIS and the NWM, the 3-hour time-average SSM centered at 22:30 UTC was used here as the daily SMAP_L4 SSM value. Note that SMAP_L4 surface runoff and baseflow estimates, while available, were not utilized.

As a data assimilation product, the seasonality of SMAP_L4 SSM data matches that of the Catchment LSM. Moreover, SMAP_L4 SSM is impacted by errors in the precipitation data used to force the Catchment LSM (which differ from the NLDAS-2 precipitation data used to force the Noah-MP and NWM simulations and to calculate observed RC results—see above). Nevertheless, Crow et al. (2017) demonstrated that SMAP_L4 SSM estimates contain significantly more skill for estimating event-scale RC variations than SSM estimates derived from either SMAP Level 2 SSM data or LSM simulations generated without data assimilation. In

addition, due to the independence of the SMAP_L4 SSM data from the USGS streamflow observations used to construct observed RC, the superior skill of the SMAP_L4 data can be confidently ascribed to the improved precision of its SSM estimates (as opposed to spurious error cross correlation; Crow et al., 2018).

## 3. Results

### 3.1. Variation and Impact of SSM/RC Correlation

Figure 3 plots the skill in LSM estimates of event-scale RC ($R[RC_{LSM}, RC_{USGS}]$) as a function of the LSM-predicted correlation between prestorm SSM and RC ($R[SSM_{LSM}, RC_{LSM}]$) calculated for each Noah-MP LIS case (Table 1) and the NWM. Plotted values are obtained by averaging sampled correlations across all 522 study basins in Figure 1. Due to this extensive spatial averaging, 95% confidence sampling intervals for all plotted correlations are typically smaller than plotted symbols. Therefore, any visible gaps between point labels in Figure 3 can generally be considered statistically significant. The positive sign of $R[SSM_{LSM}, RC_{USGS}]$ values reflects the known tendency for wetter prestorm SSM conditions to produce larger event-scale RC (Crow et al., 2017).

The variation of $R[SSM_{LSM}, RC_{LSM}]$ between Noah-MP (and NWM) cases is generally consistent with expectations given the physical assumptions underlying each surface runoff parameterization. In both the SIM TOP and BATS cases, surface runoff, and thus event-scale RC, is modeled as a direct function of a quantity closely related to SSM (i.e., $F_{sat}$ and 2-m soil moisture, respectively). Therefore, this type of "saturation excess" runoff formulation tends to predict strong SSM/RC correlation (i.e., large $R[SSM_{LSM}, RC_{LSM}]$). The single exception being SIM TOP case #5, where the use of a small $f$ value ($3\ m^{-1}$) leads to relatively lower $R[SSM_{LSM}, RC_{LSM}]$ but also excessively small values of mean surface runoff (see Figure 2).

In contrast to the SIM TOP and BATS cases, the FD cases assign only secondary importance to prestorm SSM values and instead bases RC estimates primarily on rainfall rate considerations via an "infiltration excess" surface runoff formulation. As a result, $R[SSM_{LSM}, RC_{LSM}]$ is relatively low for all FD cases. Applied values of *REFKDT* in cases #10–#13 (0.5 to 5 [-]) span the entire physically realistic range suggested by the Noah-MP user guide (Niu et al., 2011). As expected, correlation increases with increasing *REFKDT* (see Figure 3 and Table 1). However, even the largest applied *REFKDT* value (5.0 [-] in case #13) yields relatively low correlation (Figure 3), and values of *REFKDT* above 1.0 [-] (i.e., FD cases #12 and #13) are associated with unrealistically low levels of mean surface runoff (Figure 2). Therefore, it does not appear that the low SSM/RC correlation bias in the FD cases can be realistically resolved solely through the recalibration of *REFKDT*.

The low $R[SSM_{LSM}, RC_{LSM}]$ correlation values obtained for the SIM GW cases (Figure 3) are unexpected given the close conceptual connection between SIM GW and the (well-correlated) SIM TOP cases. While larger $f$ values tend to produce correspondingly larger $R[SSM_{LSM}, RC_{LSM}]$ for the SIM GW cases (see Table 1 and Figure 3), SIM GW SSM/RC correlations remain less than SIM TOP *even when utilizing a larger f value*. For example, SIM GW case #3 with $f = 4\ m^{-1}$ produces significantly less correlation than SIM TOP case #5 with $f = 3\ m^{-1}$. This demonstrates that differences between the SIM GW and SIM TOP cases cannot be attributed solely to differences in $f$. An alternative possibility is that the difference is due to the enforced 1.5-m minimum for the SIM GW water table depth ($Z_{WT}$; see section 2.3 above). This modification effectively decouples SIM GW $F_{sat}$ estimates from fast variations in SSM since larger values of $Z_{WT}$ are associated with relatively slower water dynamics due to the assumed exponential decay of saturated hydrologic conductivity with depth in both the SIM GW and SIM TOP cases. The relatively higher value of $R[SSM_{LSM}, RC_{LSM}]$ obtained from the NWM suggests that the NWM water table representation avoids this pitfall.

Results in Figure 3 also illustrate how low internal SSM/RC correlation is strongly associated with low RC skill ($R[RC_{LSM}, RC_{USGS}]$) in corresponding LSM RC estimates. Therefore, low SSM/RC correlation has a profound influence on the ability of LSMs to adequately capture event-scale RC variations. For example, low values of $R[RC_{LSM}, RC_{USGS}]$ obtained from the SIM GW and FD cases in Figure 3 are associated with underrepresenting the impact of prestorm SSM on event-scale RC and, therefore, overstating the role of other factors (e.g., fine-scale variations in precipitation intensity). Conversely, all other things being equal, the much stronger correlation predicted by the SIM TOP and BATS parameterization cases enhances their ability to track storm-to-storm variations in RC.

Also included in Figure 3 is an extreme case (labeled "$SSM_{direct}$") derived by directly sampling the Spearman rank correlation between Noah-MP prestorm SSM estimates and observed RC ($R[SSM_{LSM}, RC_{USGS}]$). As such, it corresponds to a case of assuming perfect correlation between SSM and RC and, consequently, neglecting the impact of all other factors on modeled RC. Put another way, it summarizes the skill available from considering only prestorm Noah-MP SSM values and ignoring Noah-MP runoff physics (and/or any SSM modeling errors). The actual plotted value for the $SSM_{direct}$ case is the median of the basin-averaged $R[SSM_{LSM}, RC_{USGS}]$ sampled separately for all Noah-MP LIS cases in Table 1. However, since all 16 Noah-MP cases yield $R[SSM_{LSM}, RC_{USGS}]$ values within a very tight range ($\pm$0.02 [-]), the plotted $SSM_{direct}$ case can be considered a single fixed value across all cases. Note also that this lack of case-to-case variability in SSM supports our earlier assumption that overall SSM/RC coupling is dominated by RC dependence on SSM (as oppose to SSM dependence on RC). That is, large case-to-case variations in RC skill found in Figure 3 (reflecting the importance of RC dependence on SSM) do not feedback into comparably large variations in SSM.

While underestimating the dependence of RC on SSM appears to be a significant problem in Figure 3, there is less evidence that overestimating this dependence harms LSM RC skill. There is a slight decrease in RC skill ($R[RC_{LSM}, RC_{USGS}]$) observed at high SSM/RC correlation ($R[SSM_{LSM}, RC_{LSM}]$); however, this decrease is not statistically significant, and using SSM as a proxy for RC—that is, neglecting Noah-MP runoff estimates altogether, as captured via the $SSM_{direct}$ case of Figure 3—only slightly underperforms RC estimates obtained from the best Noah-MP surface runoff parameterization cases. In fact, the $SSM_{direct}$ case provides significantly more RC skill than RC estimates made by 10 out of the 17 total LSM cases considered in Figure 3. In these cases, LSM RC estimates are less skillful for describing event-scale variations in RC than their corresponding prestorm SSM estimates. This represents a physical deficiency given that *SSM inputs* into these runoff parameterizations are more skillful (for event-scale RC estimation) than the actual *RC outputs* provided by the LSM runoff parameterization.

### 3.2. Spearman Rank Versus Pearson Correlation

All correlations plotted Figure 3 are Spearman rank type and therefore reflect nonparametric skill in the RC estimates (or, for the $SSM_{direct}$ case, a soil moisture proxy for RC). In contrast, a Pearson correlation describes the ability of each estimate to *linearly* correlate with observed RC. Therefore, examining differences between Spearman rank and Pearson correlation results provides an assessment of linearity. To this end, we regenerated Figure 3 for the case of sampling Pearson correlation instead of Spearman rank correlation (not shown). For cases based on Noah-MP and NWM RC estimates (i.e., all labeled points in Figure 3 except for the $SSM_{direct}$ case) this transition had little impact—suggesting that the relationship between LSM-estimated RC and observed RC is relatively linear.

However, the use of Pearson correlation led to a reduction in skill for the $SSM_{direct}$ case in Figure 3—which captures the linear correlation between LSM-predicted prestorm SSM and observed RC ($R[SSM_{LSM}, RC_{USGS}]$). Consistent with earlier results in Crow et al. (2017), this reduction suggests the presence of modest monotonic nonlinearity in the relationship between prestorm SSM estimates and event-scale RC. Therefore, while LSM runoff parameterizations (used to convert prestorm SSM into RC) do not appear to generally improve upon the nonparametric skill of prestorm SSM estimates (see discussion above and in section 3.2), they may still provide a superior *linear* proxy for true RC. While the rest of our analysis is restricted to the use of Spearman rank correlation, this subtlety should be kept in mind.

### 3.3. Observed SSM/RC Correlation

Figure 3 suggests that establishing adequate RC dependence on SSM is critical for obtaining skillful RC estimates from LSMs. An external source of observed correlation estimates can be obtained by sampling the Spearman rank correlation between prestorm, basin-averaged SMAP_L4 SSM values (described above in section 2.3) and event-scale USGS-observed streamflow volumes (normalized by event-scale NLDAS-2 precipitation volumes to produce an observation-based, event-scale RC value). Figure 4 plots a map of benchmark SSM/RC correlation estimates ($R[SSM_{SMAPL4}, RC_{USGS}]$) obtained in this way.

It is difficult to discern coherent regional patterns and/or clear correlation with known land surface characteristics within Figure 4. There appears to be a tendency for higher values of $R[SSM_{SMAPL4}, RC_{USGS}]$ along

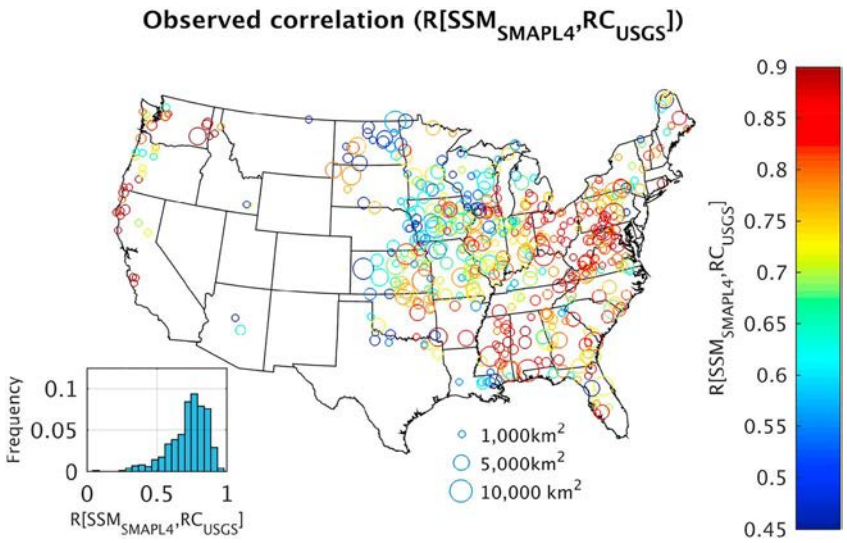


**Figure 4.** Observed correlation between SMAP_L4 prestorm SSM and USGS-based RC ($R$[RC$_{LSM}$, RC$_{USGS}$]) for all 522 basins in Figure 1. Circle size scales with basin size. Inset contains the normalized histogram of $R$[RC$_{LSM}$, RC$_{USGS}$] results for all basins. SMAP_L4 = Soil Moisture Active Passive Level-4 product; SSM = surface soil moisture; USGS = United States Geological Survey; LSM = land surface model; RC = runoff coefficient.

the Appalachian Mountains in eastern CONUS and lower values in flatter (and generally drier) areas of central and north central CONUS. This may reflect the greater role of saturation excess surface runoff mechanisms (and thus higher SSM/RC correlation) in wetter areas with more topographic relief. However, this interpretation is speculative, and it should be noted that some patterns in Figure 4 are likely the spurious reflection of spatially varying errors in SMAP_L4 SSM estimates.

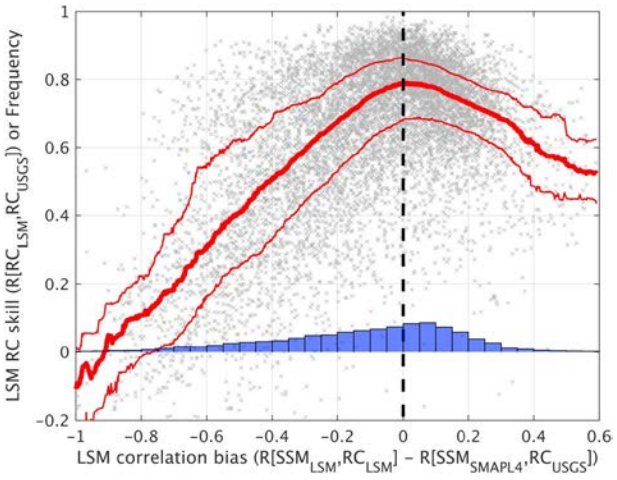


**Figure 5.** Event-scale LSM RC skill ($R$[RC$_{LSM}$, RC$_{USGS}$]) versus event-scale correlation bias ($R$[RC$_{LSM}$, RC$_{LSM}$] − $R$[RC$_{SMAPL4}$, RC$_{USGS}$]). The point cloud consists of 17 × 522 = 8,874 symbols generated from all 16 Noah-MP LIS simulations listed in Table 1 (plus the single NWM retrospective simulation) for each of the 522 basins shown in Figure 1. The red line plots 25th, 50th (median), and 75th quartiles of $R$[RC$_{LSM}$, RC$_{USGS}$] values sampled within a moving window along the abscissa. The normalized histogram describes the relative frequency of LSM correlation biases. SMAP_L4 = Soil Moisture Active Passive Level-4 product; SSM = surface soil moisture; USGS = United States Geological Survey; LSM = land surface model; RC = runoff coefficient; NWM = National Water Model.

A more certain inference is that $R$[SSM$_{SMAPL4}$, RC$_{USGS}$] values in Figure 4 are uniformly positive and relatively large (see inset histogram). This is consistent with the suggestion in Figure 3 that relatively strong SSM/RC correlation is required for Noah-MP (or NWM) cases to accurately capture event-to-event variations in RC. In addition, it implies that observed $R$[SSM$_{SMAPL4}$, RC$_{USGS}$] in Figure 4 can describe basin-to-basin (and case-to-case) variations in RC skill by diagnosing bias in LSM estimates of SSM/RC correlation.

To this end, Figure 5 plots the skill of LSM RC estimates (i.e., $R$[RC$_{LSM}$, RC$_{USGS}$]) as a function of bias in LSM SSM/RC correlation estimates relative to observed correlation plotted in Figure 4 (i.e., $R$[SSM$_{LSM}$, RC$_{LSM}$] - $R$[SSM$_{SMAPL4}$, RC$_{USGS}$]). Results are plotted for each basin individually (pooled across all 16 Noah-MP parameterization cases summarized in Table 1 plus the NWM simulation). In the resulting point cloud, there is a clear tendency for the best RC estimates to be associated with LSMs that best match observed SSM/RC correlation levels. Since all LSM cases utilize the same NLDAS-2 forcing data and apply the same simple routing scheme, RC variations are the only source of case-to-case variability in streamflow results. Therefore, patterns in Figure 5 do not qualitatively change when streamflow skill is considered in place of RC skill.

As previously seen in Figure 3, the largest identifiable LSM problem in Figure 5 is the underrepresentation of RC dependence on SSM. Note, for example, the large quantity of negative abscissa points in Figure 5 (indicating $R$[SSM$_{LSM}$, RC$_{LSM}$] < $R$[SSM$_{SMAPL4}$, RC$_{USGS}$]) and their association with low LSM RC skill ($R$[RC$_{LSM}$, RC$_{USGS}$]). While less pronounced, there is also evidence in Figure 5 that the overestimation of

RC dependence on SSM can degrade RC skill (see the slight decline in LSM RC skill associated with large, positive $R[SSM_{LSM}, RC_{LSM}]$ minus $R[SSM_{SMAPL4}, RC_{USGS}]$ values). Note that this tendency was not significant earlier in Figure 3 and emerges only after evaluating correlation bias separately (on a case-by-case and basin-by-basin) relative to the observed datum ($R[SSM_{SMAPL4}, RC_{USGS}]$). Despite the slight discrepancy in the vertical support of SMAP_L4 (0–5 cm) and Noah-MP (0–10 cm) SSM estimates, the LSM cases that exhibit the lowest absolute bias in SSM/RC correlation. That is, the lowest values of $abs(R[SSM_{LSM}, RC_{LSM}] - R[SSM_{SMAPL4}, RC_{USGS}])$ are generally associated with the largest LSM RC skill ($R[RC_{LSM}, RC_{USGS}]$). Therefore, when used in conjunction with USGS RC observations, SMAP_L4 SSM estimates can robustly detect both the overdependence and underdependence of LSM RC estimates on SSM.

As discussed above in section 2.1, several preprocessing decisions underpin our definition of a storm event and thus all results presented in Figures 3–5. Nevertheless, key qualitative results appear relatively insensitive to these choices. For example, modifying the storm event threshold within the range 5 to 15 mm/day has only a small impact on results in Figure 3–5. Likewise, applying the HYSEP baseline separation algorithm to isolate fast stormflow in USGS streamflow measurements, and utilizing only surface runoff estimates to calculate RC, does not qualitatively impact results. Therefore, results are not affected by the presence (or absence) of baseflow. Finally, utilizing Pearson (as opposed to Spearman rank) correlation to summarize LSM RC skill (not shown) does not substantially alter Figure 5. These conclusions are broadly consistent with Crow et al. (2017) who found that SSM/RC correlation results based on the ranked correlation between prestorm SSM and event-scale RC are relatively insensitive to the details of the approach used to define and characterize storm events.

### 3.4. Individual Noah-MP Surface Runoff Parameterizations

As mentioned above, the point cloud in Figure 5 is sampled across both space (i.e., all 522 basins in Figure 1) and all 17 LSM parameterization cases (i.e., the 16 Noah-MP cases listed in Table 1 plus the NWM simulation). It is also instructive to subset results by individual Noah-MP parameterization cases. To this end, Figures 6–9 show example results for individual Noah-MP runoff parameterization cases SIM GW, SIM TOP, and FD and the NWM case. For brevity, only Noah-MP LIS cases #2, #6, and #11 (corresponding to the default SIM GW, SIM TOP, and FD parameterizations cases listed in Table 1, respectively) are discussed, and the BATS case is omitted entirely.

Figure 6a is analogous to Figure 5 in that it plots bias in LSM SSM/RC correlation estimates ($R[SSM_{LSM}, RC_{LSM}]$ - $R[SSM_{SMAPL4}, RC_{USGS}]$) versus Noah-MP RC skill ($R[RC_{LSM}, RC_{USGS}]$)—except now only for a single SIM GW case (i.e., Noah-MP LIS case #2 in Table 1). Here, nearly all basins demonstrate a deficit of RC dependence on SSM (i.e., negative abscissa values in Figure 6a). As in Figure 5, this bias has an impact on the skill of Noah-MP RC estimates. In geographic terms, the low bias in SSM/RC correlation is strongest in central CONUS (Figure 6b) and appears to be the source of relatively low Noah-MP RC skill in this region (Figure 6c).

Relative to the SIM GW case in Figure 6, internal LSM estimates of SSM/RC correlation ($R[SSM_{LSM}, RC_{LSM}]$) are larger for the SIM TOP case (i.e., Noah-MP LIS case #6 in Table 1) in Figure 7. This increased correlation translates into improved LSM RC skill—particularly over central CONUS (cf. Figure 7c to Figure 6c). Relative to $R[SSM_{SMAPL4}, RC_{USGS}]$, the SIM TOP overestimates RC dependence on SSM across a large fraction of northern CONUS (see blue shading in Figure 7b). However, this does not translate into a degradation of Noah-MP RC estimates—which remain generally skillful (see red shading in Figure 7c). This is consistent with earlier results in Figures 3 and 5 and demonstrates that the negative impact of LSM SSM/RC overdependence, while not completely negligible, is less pronounced than that of underdependence. This can also be seen in Figure 7a where the overdependence of RC on SSM (i.e., positive $R[SSM_{LSM}, RC_{LSM}]$ minus $R[SSM_{SMAPL4}, RC_{USGS}]$ values) is associated with only a modest reduction in the skill of Noah-MP RC estimates. Results for the BATS surface runoff parameterization are generally consistent with those of the SIM TOP parameterization case and therefore omitted here for brevity.

Moving to a free-drainage ("FD") parameterization (i.e., Noah-MP LIS case #11 in Table 1) in Figure 8 returns to the case of underestimating RC dependence on SSM. Using an older version of the Noah model (with a FD parameterization), Crow et al. (2017) previously noted this underestimation within the south central United States. Figure 8 confirms this, while also indicating that this correlation bias extends throughout

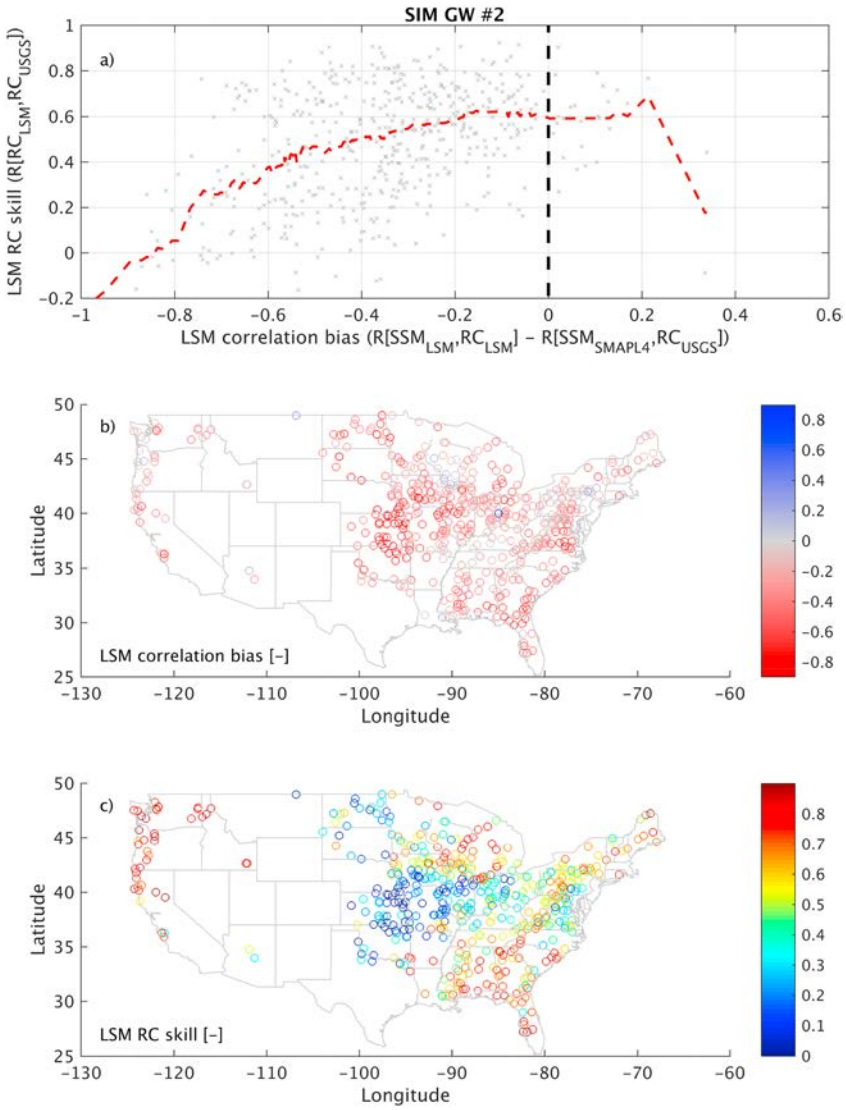


**Figure 6.** (a) Skill of Noah-MP RC estimates ($R$[$RC_{LSM}$, $RC_{USGS}$]) versus Noah-MP SSM/RC correlation bias ($R$[$RC_{LSM}$, $RC_{LSM}$] - $R$[$RC_{SMAPL4}$, $RC_{USGS}$]) for Noah-MP SIM GW case #2. Dashed red line in (a) describes a moving-window median filter. Corresponding spatial maps of (b) Noah-MP SSM/RC correlation bias and (c) skill of LSM RC estimates. Noah-MP = Noah-Multiple Physics; SMAP_L4 = Soil Moisture Active Passive Level-4 product; SSM = surface soil moisture; USGS = United States Geological Survey; LSM = land surface model; RC = runoff coefficient; SIM GW = simplified groundwater.

CONUS (see Figure 8b). The area of strongest correlation bias appears concentrated in central CONUS and extends into the northern Great Plains region (see red shading in Figure 8b). This region of low correlation is colocated with areas of relatively poor RC estimation skill (see blue shading in Figure 8c) and the poorly drained "prairie pothole" region of north central CONUS—an area where a FD boundary condition is expected to perform poorly.

All Noah-MP FD cases produce a net low bias in mean SSM/RC correlation (Figure 3) – particularly in north central CONUS. This bias can be reduced, but not removed, by increasing *REFKDT* within its physically realistic range (see Table 1 and Figure 3). This suggests the SSM/RC correlation biases within the Noah-MP FD parameterization case cannot be easily resolved via the calibration of parameter values.

There also appears to be some geographic variation in the relationship between SSM/RC correlation bias and RC skill. These quantities are tightly linked in, for example, central CONUS (see above and Figures 6–8). In contrast, they diverge somewhat in extreme southeastern CONUS. While the Noah-MP SIM GW (Figure 6b),

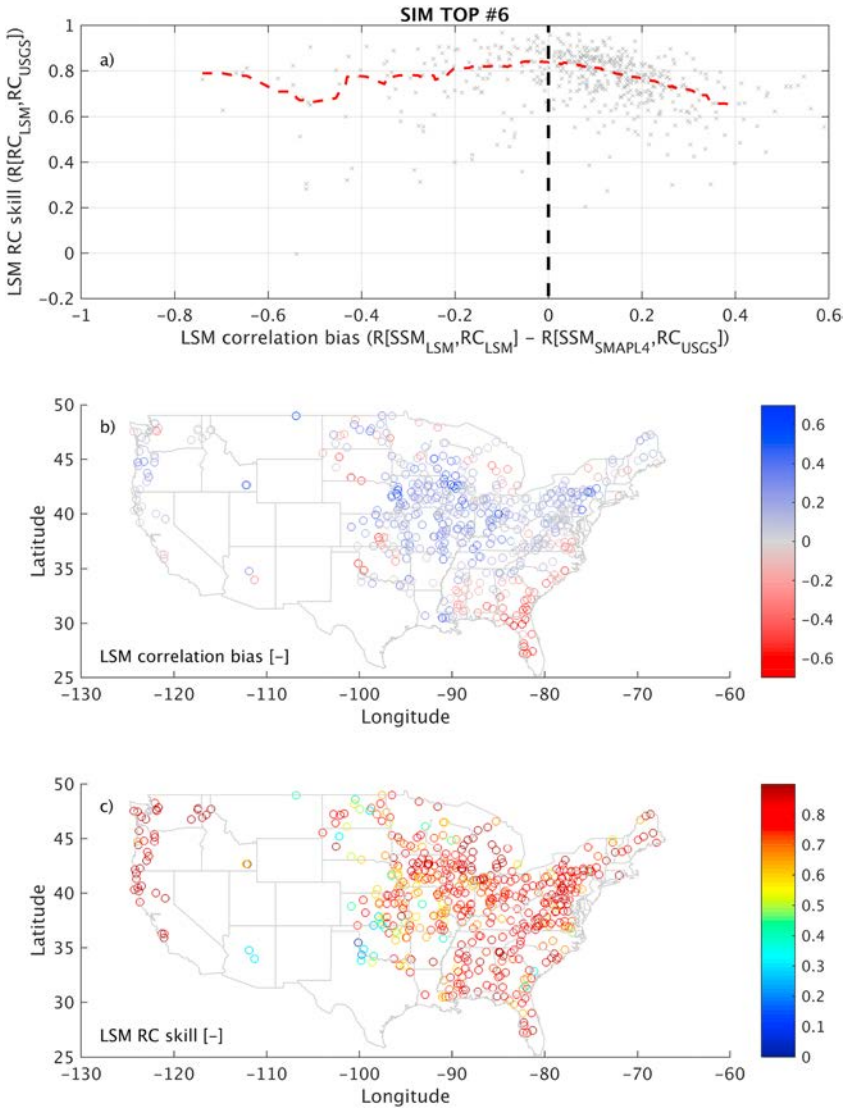


**Figure 7.** Same as Figure 6, except for Noah-MP SIM TOP case #6 (Table 1). SIM TOP = simplified TOPMODEL.

SIM TOP (Figure 7b), and FD (Figure 7c) cases all underestimate RC dependence on SSM in this region, they do not show a corresponding decrease in RC skill (see Figures 6c, 7c, and 8c). This region is well known for its flat topography and karst carbonate geology and, therefore, a tendency toward low long-term RC (see Figure 1). These unusual hydrologic characteristics may conspire to disrupt the relationship between SSM/RC correlation bias and RC skill seen elsewhere.

Finally, Figure 9 breaks out basin-by-basin results for the NWM. As noted above, the NWM performs as well as any Noah-MP LIS parameterization—with respect to both matching observed SSM/RC correlation (Figure 9b) and obtaining skillful RC estimates (Figure 9c). This success is somewhat surprising given that the NWM surface runoff formulation uses an infiltration-excess approach for surface runoff generation and has a close conceptual connection with the Noah-MP FD case, which performs relatively poorly. Indeed, there are clear similarities in the SSM/RC correlation bias patterns found in both cases (cf. Figure 8b to Figure 9b). However, the NWM generally increases SSM/RC correlation—thus correcting for the low bias found in the FD case (Figure 8c). This correction is particularly strong in central and east central CONUS where the NWM strongly outperforms the FD case (cf. Figure 8c to Figure 9c) and suggests that the inclusion of custom groundwater and overland flow modules in the NWM successfully increases both SSM/RC correlation and RC skill in these regions. Nevertheless, pronounced underestimation of prestorm SSM/RC

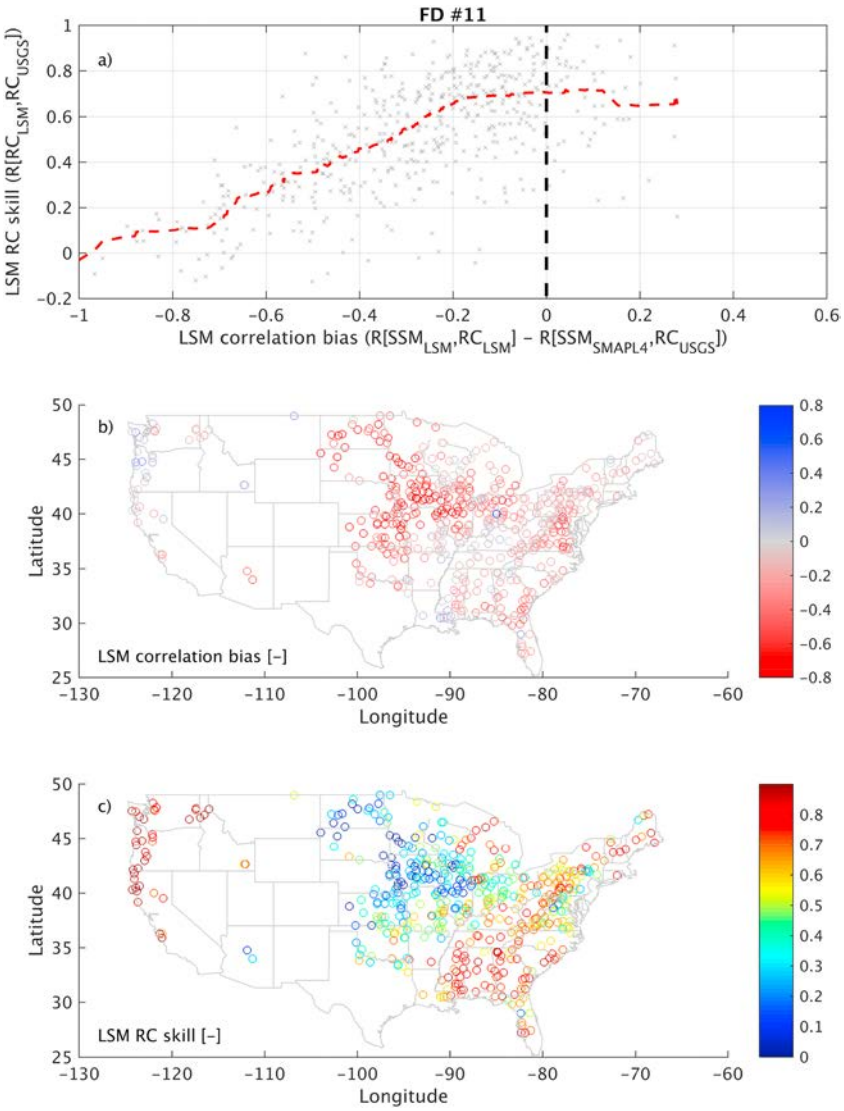


**Figure 8.** Same as Figure 6, except for Noah-MP FD case #11 (Table 1). MP = Multiple Physics; FD = free drainage.

correlation remains in the NWM for multiple basins in central and eastern CONUS (see red circles in Figure 9b), which is reflected (albeit weakly) in degraded RC skill for these regions (see Figure 9c).

### 3.5. Implications for Ungauged Basins

Across all four cases shown in Figures 6a, 7a, 8a, and 9a, the maximum value of the dashed red trend line is reached at, or near, a SSM/RC net correlation bias of zero. This result suggests that, regardless of which surface runoff parameterization case is applied, minimized absolute bias in LSM estimates of SSM/RC correlation (relative to SMAP Level 4 SSM and USGS RC estimates) is generally associated with maximized RC skill. Given the large explanatory power of SMAP_L4 SSM for RC estimation, and the relative lack of consequences associated with overestimating the dependence of RC on SSM (Figure 3), one possibility in ungauged basins is to use SMAP_L4 prestorm SSM values as a proxy for RC. In this approach, Noah-MP parameterization cases are ranked based on the positive Spearman rank correlation between their event-scale RC estimates and SMAP_L4 prestorm SSM ($R[SSM_{SMAPL4}, RC_{LSM}]$). For each individual basin, Noah-MP cases associated with the largest (positive) $R[SSM_{SMAPL4}, RC_{LSM}]$ value are selected with the expectation that they also provide the highest LSM RC skill ($R[RC_{LSM}, RC_{USGS}]$).

Figure 10 examines such a strategy by plotting $R[SSM_{SMAPL4}, RC_{LSM}]$ versus $R[RC_{LSM}, RC_{USGS}]$. Each of the 8,352 gray circles represents one of the 16 Noah-MP LIS parameterizations (the NWM case is omitted due to

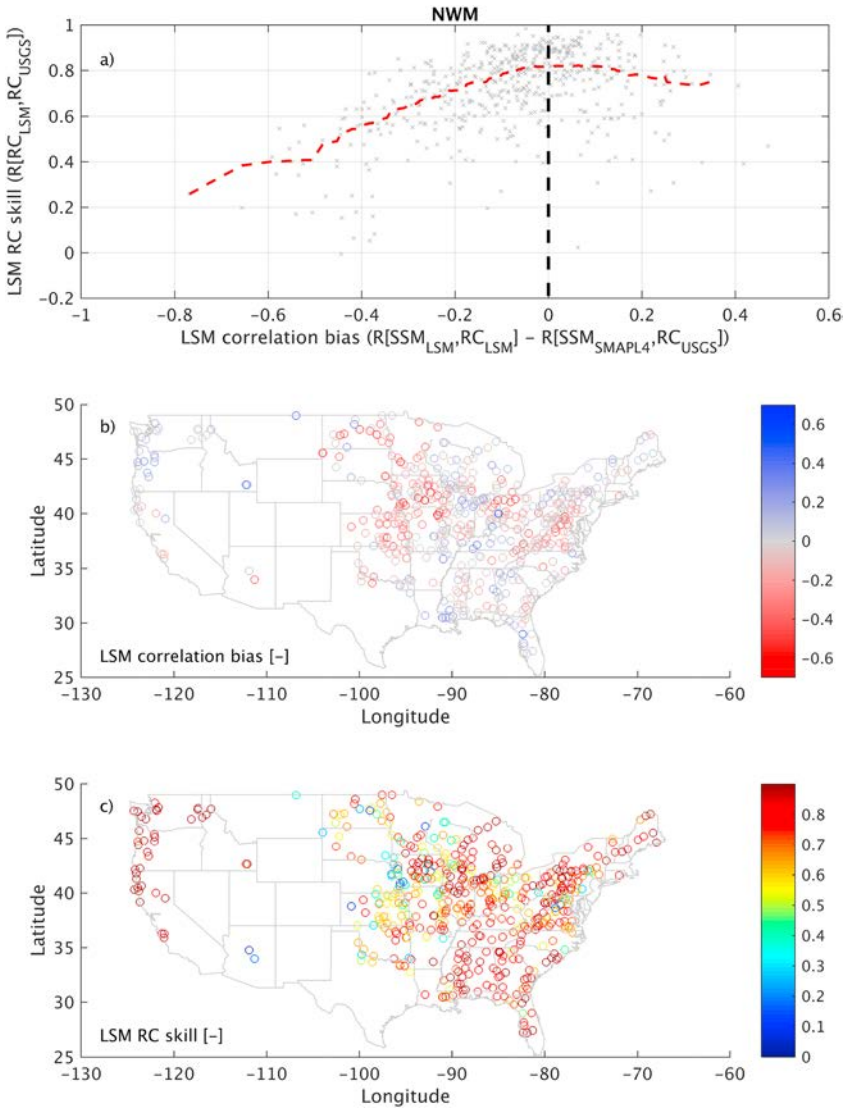


**Figure 9.** Same as Figure 6, except for the NWM simulation. NWM = National Water Model.

its slightly different historical period) within one of the 522 study basins in Figure 1. The resulting scatterplot suggests that substantial positive rank correlation between Noah-MP RC and SMAP_L4 prestorm SSM tends to identify LSM parameterization cases with the most skillful RC estimates. Given that abscissa values can be obtained in the absence of streamflow observations, this implies a potential case selection strategy for ungauged hydrologic basins. To this end, red crosses in Figure 10 reflect results for the single Noah-MP LIS parameterization case (selected individually for each basin) with the largest $R[SSM_{SMAPL4}, RC_{LSM}]$ values. Such classification does not require streamflow data—yet still successfully eliminates most Noah-MP LIS cases with poor RC skill (i.e., gray circles in Figure 10 with small positive or negative $R[RC_{LSM}, RC_{USGS}]$ values).

However, it should be noted that the relative success of this strategy depends on the baseline parameterization case considered. For each Noah-MP LIS case in Table 1, Figure 11 shows a boxplot of RC rank correlation skill across all basins. For comparison, the horizontal black lines in the figure show the same quantiles for the case of applying the case selection strategy described above (i.e., for each basin selecting the single Noah-MP parameterization case whose RC estimates maximize $R[SSM_{SMAPL4}, RC_{LSM}]$). Such basin-by-basin case selection clearly improves upon the least successful Noah-MP LIS parameterization cases (i.e., cases #1–5 and #8–13). However, it offers no advantage relative to the most successful Noah-MP LIS cases

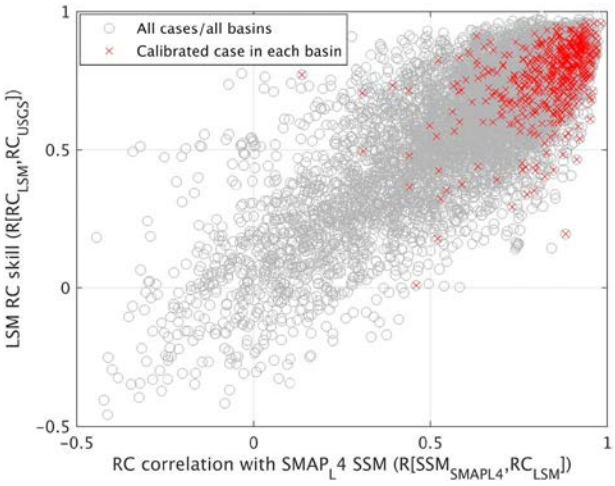


**Figure 10.** Skill of LSM RC estimates ($R$[$RC_{LSM}$, $RC_{USGS}$]) plotted against the rank correlation between prestorm SMAP_L4 SSM and LSM RC ($R$[$SSM_{SMAPL4}$, $RC_{LSM}$]). The 8,352 gray circles represent all 16 Noah-MP LIS cases (Table 1) across all 522 study basins. The 522 red crosses represent the single case for each basin with maximized $R$[$SSM_{SMAPL4}$, $RC_{LSM}$]. LSM = land surface model; RC = runoff coefficient; USGS = United States Geological Survey; SMAP_L4 = Soil Moisture Active Passive Level-4 product; Noah-MP = Noah-Multiple Physics; LIS = Land Information System; SSM = surface soil moisture.

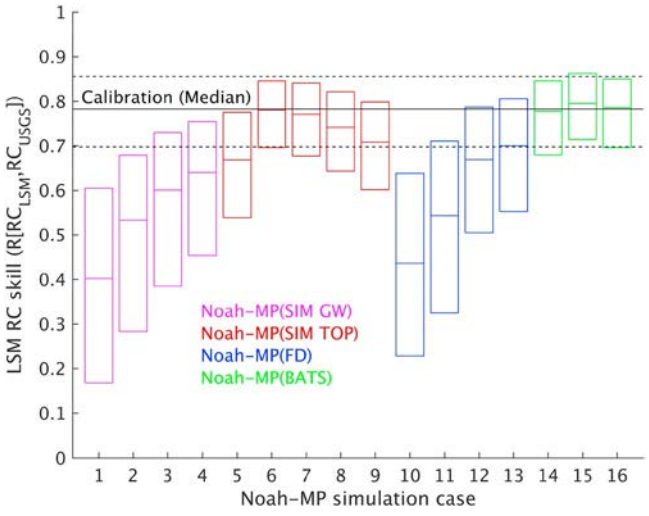


**Figure 11.** For each of the 16 Noah-MP LIS parameterization cases listed in Table 1, box plots (i.e., 25th quartile, median, and 75th quartile) of Noah-MP RC skill ($R$[$RC_{LSM}$, $RC_{USGS}$]) sampled across the 522 study basins. Horizontal lines show the same information for selecting the single Noah-MP case (on a basin-by-basin basis) with maximized $R$[$SSM_{SMAPL4}$, $RC_{LSM}$] (i.e., parameterization cases represented by red crosses in Figure 10). Noah-MP = Noah-Multiple Physics; LIS = Land Information System; LSM = land surface model; RC = runoff coefficient; USGS = United States Geological Survey; SMAP_L4 = Soil Moisture Active Passive Level-4 product.

(i.e., cases #6 and 7 and #14–16). The uniform application of these "successful" cases performs at least as well as selecting cases on a basin-by-basin basis using our proposed calibration strategy. This calls into question the true skill of the approach in robustly identifying spatially variable patterns in optimal Noah-MP cases. Instead, the approach seems best suited to broadly identifying Noah-MP parameterization cases that are clearly suboptimal. In addition, this approach should be compared to more direct calibration strategies that attempt to improve streamflow estimates by direct calibration against retrieved SSM (see, e.g., Koster et al., 2018).

## 4. Summary and Conclusions

The dependence of RC on SSM is a critical component of accurately modeling the surface water and energy balance and has direct relevance for the use of LSMs in hydrologic forecasting applications—including the proposed use of the Noah-MP LSM as the modeling core for the Unites States NWM. Results here reveal a striking consistency between SSM/RC correlation (captured by various Noah-MP runoff parameterization cases) and the skill of Noah-MP RC estimates (Figure 3). For the SIM GW and FD Noah-MP cases, failure to properly leverage skill contained in prestorm Noah-MP SSM estimates ensures that SSM estimates often correlate better with the observed RC than actual Noah-MP RC estimates (Figure 3). An analysis of variations in key model parameters suggests that performance issues in these Noah-MP surface runoff parameterizations cases are not easily resolved via model calibration (section 3.1).

In contrast, the SIM TOP and BATS Noah-MP cases demonstrate significantly higher levels of prestorm SSM/RC correlation and relatively higher RC skill. Somewhat surprisingly, the simple, power law relationship between 2-m soil moisture and saturated fraction employed by the BATS surface runoff case produces the highest RC skill (Figure 3) and very good estimates of long-term mean runoff (Figure 2). However, even in these best-case scenarios, the added RC skill provided by the LSM surface runoff parameterizations (versus the $SSM_{direct}$ case of simply using prestorm SSM as a RC proxy) remains somewhat marginal (Figure 3).

Two caveats should be noted regarding these results. First, nonparametric skill is quantified based on sampled estimates of Spearman rank correlation. If (linear-based) Pearson correlation is applied instead, larger marginal skill differences are found in Noah-MP and NWM RC estimates relative to the underlying prestorm SSM estimates. Second, the parameter sensitivity analysis presented here is admittedly cursory. More comprehensive studies (see, e.g., Nasonova et al., 2009) generally assign a higher level of importance to calibration for optimizing LSM runoff estimates.

While the theoretical benefits of basing LSM runoff parameterizations on saturation excess runoff concepts are well known, the recent availability of SSM products from L-band satellite missions (e.g., SMAP) affords a new opportunity to directly measure these advantages. For example, the strong relationship between SSM/RC correlation and LSM RC skill in Figure 3 suggests that bias in SSM/RC correlation can be used to diagnose and predict variations in the quality of LSM RC estimates. In fact, independent estimates of SSM/RC correlation derived from SMAP_L4 prestorm SSM, NLDAS-2 precipitation, and USGS streamflow observations can be spatially mapped (Figure 4) and used as the baseline for

correlation bias estimates that explain a large fraction of variability in the skill of event-scale LSM RC estimates (Figure 5). Across the entire range of land surface and climate conditions present in CONUS, and across the broad range of surface runoff parameterizations described in Table 1, a coherent relationship emerges between SSM/RC correlation bias and the event-scale RC skill. Most notably, the performance is maximized in the case of zero correlation bias. This suggests that bias in prestorm SSM/RC correlation estimates provides a unifying diagnostic for LSM-based RC estimates that describes RC skill variations encountered across both space and via the use of multiple runoff physical parameterizations. Therefore, once quantified using the SMAP_L4 SSM product, this diagnostic bias can be used to map variations in LSM performance and explain geographic patterns in the RC skill of individual LSM parameterization cases (see Figures 6–9). This represents an advancement in our ability to understand, and thus address, spatial and case-based variations in LSM performance.

Certain correlation biases (e.g., the negative biases in SSM/RC correlation captured by the FD parameterization case) are physically intuitive; however, the source of others is less apparent. For example, reasons behind higher prestorm SSM/RC correlation in the SIM TOP cases versus the SIM GW cases are unclear (since the two share relatively similar runoff physics). One possibility is that the difference is due to the specification of a minimum depth-to-water table in the Noah-MP v3.6 SIM GW parameterization. However, additional work is needed to confirm this hypothesis.

While the proposed use of the Noah-MP model within the NWM is a motivating factor here, it should be noted that the Noah-MP configuration currently implemented in the NWM does not correspond perfectly to any of the Noah-MP LIS cases listed in Table 1 and is subject to future change. Nevertheless, Version 1.2 of the publicly posted NWM retrospective run performs relatively well. It contains adequate SSM/RC correlation and exhibits relatively good rank correlation-based RC skill (see Figure 3). The inclusion of novel groundwater and overland flow modules in the NWM appears to enhance its ability to accurately reflect SSM/RC correlation versus the Noah-MP FD baseline case, which, in turn, enhances the NWM's RC skill (compare Figures 8 and 9). There is, however, room for improvement. For example, a slight degradation in NWM RC skill is noted for basins containing large negative SSM/RC correlation biases (see Figure 9a)—suggesting an ongoing role for the SMAP_L4 soil moisture product in refining future versions of the NWM.

Even though SSM/RC correlation biases (plotted along the abscissa of Figure 5) cannot be directly detected in ungauged basins, it is possible to calibrate LSM RC estimates by using event-scale SMAP_L4 prestorm SSM values as a proxy for (unavailable) RC observations and maximizing the Spearman rank correlation between SMAP_L4 and modeled RC. The relative success of the $SM_{direct}$ parameterization case in Figure 3 suggests that this approach has some merit. Indeed, selecting Noah-MP surface runoff parameterizations, on a basin-by-basin basis, that maximize the rank correlation of Noah-MP RC estimates with SMAP_L4 prestorm SSM enhances the overall performance of Noah-MP RC estimates (Figure 10). This is notable because such a strategy does not require streamflow observations and is therefore feasible in ungauged basins. However, our proposed approach fails to improve upon the alternative of broadly applying a single, well-performing Noah-MP LIS parameterization case (Figure 11). This calls into question its ability to provide robust, geographically varying information for Noah-MP case selection and/or calibration. Therefore, additional work is needed to develop more effective ungauged basin calibration strategies based on the relationship between correlation bias and LSM RC skill described here.

**Acknowledgments**
Funding was provided by the NASA SMAP mission (via Wade Crow and Rolf Reichle's membership on the SMAP Science Team) and the NASA Terrestrial Hydrology Program (Award 13-THP13-0022). All data sets utilized here are publicly available. Modeling runs were extracted directly from baseline Noah-MP (v3.6) simulations conducted within the NASA LIS framework (available at https://lis.gsfc.nasa.gov/) using hourly, NLDAS-2 forcing data (available at https://disc.gsfc.nasa.gov/). Version 3 of the SMAP_L4 product is available at https://nsidc.org/data/SPL4SMGP/versions/3. Version 1.2 of the retrospective NWM simulation is available at http://edc.occ-data.org/nwm/getdata/. All streamflow data were downloaded from the USGS at nwis.waterdata.usgs.gov. All web addresses are current as of 20 June 2019.

## References

Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., et al. (2015). Improving the representation of hydrologic processes in Earth System Models. *Water Resources Research*, *51*, 5929–5956. https://doi.org/10.1002/2015WR017096

Crow, W. T., Chen, F., Reichle, R. H., Xia, Y., & Liu, Q. (2018). Exploiting soil moisture, precipitation and streamflow observations to evaluate soil moisture/runoff coupling in land surface models. *Geophysical Research Letters*, *45*, 4869–4878. https://doi.org/10.1029/2018GL077193

Crow, W. T., Chen, R., Reichle, R. H., & Liu, Q. (2017). L band microwave remote sensing and land data assimilation improve the representation of prestorm soil moisture conditions for hydrologic forecasting. *Geophysical Research Letters*, *44*, 5495–5503. https://doi.org/10.1002/2017GL073642

De Lannoy, G. J. M., & Reichle, R. H. (2016). Assimilation of SMOS brightness temperatures or soil moisture retrievals into a land surface model. *Hydrology and Earth System Sciences*, *20*, 4895–4911. https://doi.org/10.5194/hess-20-4895-2016

Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H. V., et al. (2006). Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *Journal of Hydrology*, *320*(1-2), 3–17. https://doi.org/10.1016/j.jhydrol.2005.07.031

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version-2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Koster, R. D., Liu, Q., Mahanama, S. P., & Reichle, R. H. (2018). Improved hydrological simulation using SMAP data: Relative impacts of model calibration and data assimilation. *Journal of Hydrometeorology*, *19*(4), 727–741. https://doi.org/10.1175/JHM-D-17-0228.1

Koster, R. D., & Milly, P. C. D. (1997). The interplay between transpiration and runoff formulations in land surface schemes used with atmospheric models. *Journal of Climate*, *10*(7), 1578–1591. https://doi.org/10.1175/1520-0442(1997)010<1578:TIBTAR>2.0.CO;2

Kumar, S. V., Peters-Lidard, C. D., Tian, Y., Houser, P. R., Geiger, J., Olden, S., et al. (2006). Land Information System—An interoperable framework for high resolution land surface modeling. *Environmental Modelling & Software*, *21*(10), 1402–1415. https://doi.org/10.1016/j.envsoft.2005.07.004

Linsley, R. K., Kohler, M. A., & Paulhus, J. L. H. (1982). *Hydrology for engineers*, (3rd ed.p. 508). New York: McGraw-Hill.

Lohmann, D., Lettenmaier, D. P., Liang, X., Wood, E. F., Boone, A., Chang, S., et al. (1998). The Project for Intercomparison Of Land-Surface Parameterization Schemes PILPS phase 2 c Red–Arkansas River basin experiment: 3. Spatial and temporal analysis of water fluxes. *Global and Planetary Change*, *19*(1-4), 161–179. https://doi.org/10.1016/S0921-8181(98)00046-0

Lucchesi, R. (2013), File specification for GEOS-5 FP. GMAO Office Note No. 4 (Version 1.0), 63 pp, available from http://gmao.gsfc.nasa.gov/pubs/office_notes.

Meyles, E., Williams, A., Ternan, L., & Dowd, J. (2003). Runoff generation in relation to soil moisture patterns in a small Dartmoor catchment, Southwest England. *Hydrological Processes*, *17*(2), 251–264. https://doi.org/10.1002/hyp.1122

Nasonova, O. N., Gusev, Y. M., & Kovalev, Y. E. (2009). Investigating the ability of a land surface model to simulate streamflow with the accuracy of hydrological models: A case study using MOPEX materials. *Journal of Hydrometeorology*, *10*(5), 1128–1150. https://doi.org/10.1175/2009JHM1083.1

Niu, G.-Y. and Z.-L Yang (2011), The community Noah land-surface model (LSM) with multi-physics options: User guided, http://www.jsg.utexas.edu/noah-mp/files/Users_Guide_v0.pdf.

Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., & Gulden, L. E. (2005). A simple TOPMODEL-based runoff parameterization (SIMTOP) for use in global climate models. *Journal of Geophysical Research*, *110*, D21106. https://doi.org/10.1029/2005JD006111

Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., Gulden, L. E., & Su, H. (2007). Development of a simple groundwater model for use in climate models and evaluation with gravity recovery and climate experiment data. *Journal of Geophysical Research*, *112*, D07103. https://doi.org/10.1029/2006JD007522

Niu, G.-Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research*, *116*, D12109. https://doi.org/10.1029/2010JD015139

Penna, D., Tromp-van Meerveld, H. J., Gobbi, A., Borga, M., & Dalla Fontana, G. (2011). The influence of soil moisture on threshold runoff generation processes in an alpine headwater catchment. *Hydrology and Earth System Sciences*, *15*(3), 689–702. https://doi.org/10.5194/hess-15-689-2011

Piepmeier, J. R., Focardi, P., Horgan, K. A., Knuble, J., Ehsan, N., Lucey, J., et al. (2017). SMAP L-band microwave radiometer: instrument design and first year on orbit. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(4), 1954–1966. https://doi.org/10.1109/tgrs.2016.2631978

Reichle, R., de Lannoy, G., Koster, R. D., Crow, W. T., Kimball, J. S., & Liu, Q. (2017). *SMAP L4 Global 3-hourly 9 km EASE-grid surface and root zone soil moisture geophysical data, version 3*. Boulder, Colorado USA: NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/B59DT1D5UMB4 Last accessed July 2018

Reichle, R. H., de Lannoy, G. J. M., Liu, Q., Ardizzone, J. V., Colliander, A., Conaty, A., et al. (2017). Assessment of the SMAP Level-4 surface and root-zone soil moisture product using in situ measurements. *Journal of Hydrometeorology*, *18*(10), 2621–2645. https://doi.org/10.1175/JHM-D-17-0063.1

Reichle, R. H., de Lannoy, G. J. M., Liu, Q., Koster, R. D., Kimball, J. S., Crow, W. T., et al. (2017). Global assessment of the SMAP Level-4 surface and root-zone soil moisture product using assimilation diagnostics. *Journal of Hydrometeorology*, *18*(12), 3217–3237. https://doi.org/10.1175/JHM-D-17-0130.1

Salas, F. R., Somos-Valenzuela, M. A., Dugger, A., Maidment, D. R., Gochis, D. J., David, C. H., et al. (2018). Towards real-time continental scale streamflow simulation in continuous and discrete space. *Journal of the American Water Resources Association (JAWRA)*, *54*(1), 7–27. https://doi.org/10.1111/1752-1688.12586

Schaake, J., Cong, S., & Duan, Q. (2006). *The US MOPEX data set* (Vol. 307, pp. 9–28). Livermore, CA: IAHS Publication Series.

Schaake, J. C., Koren, V. I., Duan, Q.-Y., Mitchell, K., & Chen, F. (1996). Simple water balance model for estimating runoff at different spatial and temporal scales. *Journal of Geophysical Research*, *101*(D3), 7461–7475. https://doi.org/10.1029/95JD02892

Sloto, R.A., and M.Y. Crouse (1996), HYSEP: A computer program for streamflow hydrograph separation and analysis, U.S. Geological Survey Water-Resources Investigations Report 1996–4040, 46 p., https://pubs.er.usgs.gov/publication/wri964040.

Xia, Y., Mitchell, K., Ek, M., Cosgrove, B., Sheffield, J., Luo, L., et al. (2012). Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *Journal of Geophysical Research*, *117*, D03110. https://doi.org/10.1029/2011JD016051

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., et al. (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. *Journal of Geophysical Research*, *117*, D03109. https://doi.org/10.1029/2011JD016048

Xia, Y., Mocko, D., Huang, M., Li, B., Rodell, M., Mitchell, K. E., et al. (2017). Comparison and assessment of three advanced land surface models in simulating terrestrial water storage components over the United States. *Journal of Hydrometeorology*, *18*(3), 625–649. https://doi.org/10.1175/JHM-D-16-0112.1

Yang, Z.-L., & Dickinson, R. E. (1996). Description of the Biosphere-Atmosphere Transfer Scheme (BATS) for the soil moisture workshop and evaluation of its performance. *Global and Planetary Change*, *13*(1-4), 117–134. https://doi.org/10.1016/0921-8181(95)00041-0

Yang, Z.-L., Niu, G.-Y., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins. *Journal of Geophysical Research*, *116*, D12110. https://doi.org/10.1029/2010JD015140

Zheng, D., Van Der Velde, R., Su, Z., Wen, J., & Wang, X. (2017). Assessment of Noah land surface model with various runoff parameterizations over a Tibetan River. *Journal of Geophysical Research: Atmospheres*, *122*, 1488–1504. https://doi.org/10.1002/2016JD025572