Exploiting soil moisture, precipitation and streamflow observations to evaluate soil moisture/runoff coupling in land surface models

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

[1]USDA Hydrology and Remote Sensing Laboratory, Beltsville, MD

[2]SSAI Inc., Greenbelt, MD

[3]NASA GSFC Global Modeling and Assimilation Office, Greenbelt, MD

[4]I.M. Systems Group at NCEP EMC, College Park, MD

- The NASA SMAP mission provides a unique opportunity to evaluate land surface models using high-quality soil moisture retrievals.

- Land surface models (LSMs) tend to underestimate the strength of the relationship between soil moisture and storm event runoff coefficients.

- The underestimation is largest for LSMs employing an infiltration-excess approach to stormflow generation.

Abstract

Accurate partitioning of precipitation into infiltration and runoff is a fundamental objective of land surface models tasked with characterizing the surface water and energy balance. Temporal variability in this partitioning is due, in part, to changes in pre-storm soil moisture, which determine soil infiltration capacity and unsaturated storage. Utilizing the NASA Soil Moisture Active Passive Level-4 soil moisture product in combination with streamflow and precipitation observations, we demonstrate that land surface models (LSMs) generally underestimate the strength of the positive rank correlation between pre-storm soil moisture and event runoff coefficients (i.e., the fraction of rainfall accumulation depth converted into stormflow runoff during a storm event). Underestimation is largest for LSMs employing an infiltration-excess approach for stormflow runoff generation. More accurate coupling strength is found in LSMs that explicitly represent sub-surface stormflow or saturation-excess runoff generation processes.

1. Introduction

A first-order priority for land surface models (LSMs) is accurately capturing the degree to which pre-storm soil moisture levels constrain event runoff coefficients [Koster & Milly, 1997] (i.e., the fraction of rainfall accumulation depth converted into stormflow during a storm event). The relationship between pre-storm soil moisture and hydrologic basin response has received considerable attention in small-scale field studies [e.g., Western & Grayson, 1998] and the development of hydro-geomorphologic models capable of capturing the coupled relationship between stormflow, erosion and sediment transport [e.g., Kim et al., 2016]. Such work has contributed to an improved understanding of the complex role soil moisture plays in various runoff generation processes [e.g., Mirus & Loague, 2013]. Nevertheless, runoff parameterizations in LSMs still do not reflect best hydrologic process understanding [Clark et al., 2015], and LSMs demonstrate only modest skill in estimating daily streamflow within medium-scale (10^3 to 10^4 km^2) hydrologic basins [Xia et al., 2012a].

Satellite-based soil moisture products offer a potentially useful diagnostic for examining the relationship between mean soil moisture and basin runoff response in LSMs. However, diagnostic efforts involving these products have been hampered by the low-quality of historically available, satellite-based soil moisture products [Crow et al., 2017]. In this regard, the January 2015 launch of the National Aeronautics and Space Administration (NASA) Soil Moisture Active Passive (SMAP) mission [Entekhabi et al., 2010] affords a new opportunity to examine the relationship between pre-storm soil moisture and event runoff coefficients in LSMs. The SMAP mission produces a Level-4 Surface and Root-zone Soil Moisture (SMAP_L4) product based on the assimilation of SMAP brightness temperature observations into an LSM [Reichle et al., 2016; Reichle et al., 2017]. Crow et al. [2017] demonstrates that the improved accuracy, complete spatio-temporal coverage, and sub-daily frequency of the SMAP_L4 product make it uniquely suited for characterizing the relationship between pre-storm soil moisture and storm-scale runoff response.

Given that past studies have already focused on comparing streamflow estimates from multiple LSMs to stream gauge observations (see, e.g., Xia et al. [2012b; 2012c]), our emphasis here is on using SMAP_L4 soil moisture estimates (in concert with streamflow and precipitation accumulation observations) to verify the statistical strength of internal LSM coupling between pre-storm soil moisture and event runoff coefficients.

2. Basins and data sets

Our geographic domain consists of 16 medium-scale basins located in the south-central United States (see Figure 1 and Table 1). Due to their limited topographic complexity, relatively low levels of forest cover and low frequency of snow cover, these basins are well-suited to satellite retrieval of surface soil moisture. In addition, the region has experienced an extraordinarily large number of extreme precipitation events during the past few years (Figure 1) and therefore provides an unusually large sample of significant storm events during the SMAP data period. Specific basins are selected based on a screening analysis performed by the Model Parameterization Experiment [Duan et al., 2006] which eliminates those lacking adequate rain gauge density or containing significant anthropogenic modification to their river flow system. Both mean annual precipitation and mean annual runoff efficiency (i.e., mean annual streamflow divided by mean annual precipitation) increase when moving from west to east across the region (Table 1). Rangeland, grassland and winter wheat land cover is common in basins #1-#7. Higher-numbered basins towards the east (i.e., basins #8-#16) contain relatively more vegetation biomass including significant amounts of upland forest cover and summer agriculture in low-lying areas.

2.1. Daily streamflow and rainfall observations

Daily (0 to 24 UTC) basin-averaged rainfall accumulations for each basin in Figure 1 are estimated from the spatial and temporal aggregation of hourly, 0.125° rainfall accumulation estimates produced by phase 2 of the North American Land Data Assimilation System (NLDAS-2). These estimates are, in turn, based on the merger of hourly rainfall radar data with a daily rain gauge analysis [Cosgrove et al., 2003]. Daily (0 to 24 LST, UTC-6 hours) streamflow values are obtained from United States Geological Survey (USGS) stream gauge stations [USGS, 2016] located at each basin outlet (Figure 1). The 6-hour offset between daily averages of precipitation and streamflow is meant to approximate the natural travel time lag between precipitation and the subsequent streamflow response at basin outlets. The impact of this simplified routing representation is discussed in the supporting materials.

Daily total streamflow observations [L^3/T] are divided by basin area to produce daily flux [L/T] estimates. The fast stormflow runoff component of the total streamflow time series was isolated using the USGS HYdrograph SEparation Program [HYSEP; Sloto et al., 1996].

2.2. SMAP L4 surface and root-zone soil moisture estimates

The SMAP_L4 product is generated using an ensemble-based data assimilation system that integrates SMAP brightness temperature data into the NASA Goddard Earth Observing System (GEOS) Catchment land surface model [CLSM; Koster et al., 2000]. Surface meteorological forcing data for CLSM are derived from the GEOS atmospheric assimilation system with a correction for precipitation accumulation derived from rain gauge observations [Reichle et al., 2017]. The assimilation system interpolates and extrapolates information from the SMAP brightness temperature observations in time and in space based on the relative uncertainties of the model estimates and the observations to produce a 3-hourly surface (0-5 cm) and root-zone (0-100 cm) volumetric soil moisture analysis on the 9-km EASEv2 grid [Reichle et al., 2016; 2017]. The CLSM component of the SMAP_L4 system was initialized on 1 January 2014 using model states derived by looping twice through 2000-2013 forcing data. Here, SMAP_L4 (version Vv2030) 3-hourly, 9-km resolution estimates are averaged in both space and time to produce a single daily-averaged (0 to 24 UTC) soil moisture analysis for each basin. The SMAP_L4 product is wholly independent of USGS streamflow observations (used here as a point of comparison) and provides a better representation of pre-storm soil moisture conditions than SMAP Level-3 soil moisture retrieval products (Crow et al., 2017). The SMAP_L4 system also produces runoff estimates; however, these estimates are not considered here. See the supporting materials for additional discussion of the implications associated with our use of a soil moisture analysis rather than a direct remote sensing retrieval.

2.3. Land surface models

The NLDAS-2 project generates continuous, hourly, 0.125° output from four different LSMs: the Mosaic model [Koster & Suarez, 1994; 1996], version 2.8 of the Noah model [Xia et al., 2012a], the Sacramento (SAC) model [Koren et al., 2000; 2003], and version 4.0.3 of the Variable Infiltration Capacity (VIC) model [Liang et al., 1994; 1996]. Mosaic and Noah were developed for atmospheric general circulation models and emphasize water and energy interactions between the land surface and atmosphere [Koster and Suarez, 1996; Ek et al., 2003]. In contrast, SAC and VIC were developed as off-line (land-only) hydrological models with a focus on streamflow prediction [Burnash et al., 1973; Liang et al., 1994]. All four models are driven using NLDAS-2 forcing data and parameterizations previously described in Xia et al. [2012b] and run continuously from a January 1979 initialization based on climatological state values. In addition to these four NLDAS-2 LSMs, we also assess output from an open loop (i.e., no data assimilation) simulation with CLSM using surface meteorological forcing and a spin-up identical to that of the SMAP_L4 system (see Section 2.2; [Reichle et al. 2017]).

The representation of stormflow runoff processes in each LSM varies significantly. Noah v2.8 utilizes an infiltration-excess representation based on an adaptation of the Soil Conservation Service Curve Number approach [Schaake et al., 1996]. In contrast, CLSM utilizes a saturation-excess runoff parameterization based on calculating the fraction of the land surface saturated from below by a dynamic water table. VIC and Mosaic use a similar approach, except that sub-grid saturation fractions are based on the grid-scale mean soil moisture values (as opposed to an explicitly calculated water table depth as in CLSM). In addition, Mosaic utilizes a simple linear model for the relationship between soil moisture and saturated fraction [Koster and Suarez, 1996] while VIC employs a non-linear variable infiltration curve [Liang et al., 1994]. The SAC model calculates both the free and tension soil water state. Tension water is used to calculate so-called "direct runoff," while "surface runoff" and "sub-surface interflow" are based on free water calculations [Koren et al., 2000; 2003]. Note that the term "stormflow" is used here to refer to "surface runoff" results obtained from Noah, VIC, CLSM and Mosaic as well as the sum of the SAC "surface", "direct" and "sub-surface interflow" runoff components.

For each model, daily-averaged (0 to 24 UTC) top-layer volumetric soil moisture (0-5 cm for CLSM, and 0-10 cm otherwise), root-zone volumetric soil moisture (generally 0-100 cm), stormflow, and baseflow estimates are extracted and spatially-averaged within each basin. The conversion between SAC free/tension soil water estimates and multi-layer, volumetric soil moisture products is described in Xia et al. [2014]. Note that there is some variation, both within and between LSMs, with regards to the defined depth of root-zone soil moisture estimates. For example, Noah uses a 1-m depth for grasslands and shrubs and a 2-m depth for forests. Mosaic uses a 40-cm depth for all vegetation. VIC and SAC use a 1-m depth as a default but also apply a shallower rooting depth for certain land cover types.

3. Approach

3.1. Storm event definition and rank correlation metric

Our analysis is based on the separation of the daily time series into discrete 6-day storm event periods. The first day of each event period contains a daily precipitation amount that exceeds a pre-set accumulation threshold level. To avoid the confounding impact of multiple events within a single storm event period, we discard any 6-day period containing two or more days exceeding this threshold. New storm event periods must also be preceded by at least a single day with a daily precipitation amount below this threshold. To mask snow-dominated events, the first day of any event period must have a daily mean air temperature greater than 2° [C] (based on NDLAS-2 air temperature estimates). The observed event runoff coefficient is the ratio of accumulated streamflow to accumulated rainfall after both have been temporally summed over a given storm event period. All daily soil moisture products are 0 to 24 UTC averages. Pre-storm antecedent soil moisture is defined as the minimum daily soil moisture for the two-day interval preceding a storm event. Since all basins are small enough such that their (HYSEP-predicted) basin saturation times (i.e., the interval of time after a storm event at which stormflow is no longer observed at the basin outlet) are less than our 6-day storm event period, no routing is applied to LSM-derived runoff values.

Following Crow et al. [2004; 2017], the Spearman rank correlation (R) between antecedent soil moisture and the event runoff coefficient is sampled across all available storm event periods within each basin between 31 March 2015 and 31 May 2017. Rank correlation is applied to minimize the confounding effect of potential nonlinearity in the relationship between antecedent soil moisture and event runoff coefficient. Due to the relatively short SMAP data record, which precludes the sampling of accurate R for any single basin, results presented here are based on the spatial average of R values sampled across all 16 basins ($\bar{R}$). Sensitivity analyses summarized in Crow et al. [2017] demonstrate that $\bar{R}$ is relatively insensitive to the details of our storm-event identification approach (e.g., the use of a 6-day storm event period to define storm lengths and a 2-day interval to define pre-storm soil moisture).

Our analysis is based on comparing $\bar{R}$ values obtained from internal LSM estimates of soil moisture, runoff and LSM precipitation forcing ($\overline{R_{LSM}}$; based on the five different LSMs introduced in Section 2.3) to values computed from the SMAP_L4 soil moisture analysis and *external* observations of USGS streamflow and NLDAS-2 precipitation ($\overline{R_{obs}}$). Note that SMAP_L4 soil moisture does not utilize NLDAS-2 rainfall forcing and is independent of USGS streamflow observations. In addition, we never use SMAP_L4 runoff estimates. Therefore, the critical distinction between $\overline{R_{LSM}}$ and $\overline{R_{obs}}$ is that $\overline{R_{LSM}}$ reflects only internal LSM model physics, while $\overline{R_{obs}}$ provides an objective point of reference based on mutually independent soil moisture estimates and observed event runoff coefficients. This distinction has consequences for the impact of random error. The computation of $\overline{R_{LSM}}$ relies on internal model-based estimates of soil moisture and runoff that are derived from the LSM precipitation forcing. Therefore, model estimates are never confronted with independent external streamflow information. This ensures that $\overline{R_{LSM}}$ is insensitive to random errors in the LSM precipitation. In contrast, the presence of independent random errors (in either SMAP_L4 soil moisture, NLDAS-2 rainfall or USGS streamflow) will tend to bias $\overline{R_{obs}}$ low [Findell et al., 2015]. See the supporting materials for additional discussion regarding the interpretation of $\overline{R_{LSM}}$ and $\overline{R_{obs}}$.

3.2. Uncertainty description

Sampling error bars for *R* in individual basins are estimated using a 5000-member boot-strapping approach (where individual storm events are randomly sampled with replacement to preserve the underlying storm event sample size) and then combined to estimate uncertainty in $\bar{R}$. Based on the auto-correlation analysis in Crow et al. [2017], the 16 basins in Figure 1 are assumed to contain 7.4 spatially-independent samples. This adjusted sample size is used to calculate the expected reduction in sampling uncertainty associated with averaging across all basins.

4. Results

For all five LSMs (VIC, Noah, Mosaic, SAC and CLSM), the left-hand-side of Figure 2 plots $\overline{R_{LSM}}$ values sampled between 31 March 2015 and 31 May 2017 for the 16 basins in Figure 1. The right-hand-side of Figure 2 shows analogous $\overline{R_{obs}}$ values obtained from SMAP L4 soil moisture, USGS streamflow and NLDAS-2 precipitation observations. Results are shown for the cases of utilizing both surface and root-zone soil moisture to represent pre-storm soil moisture and a storm precipitation intensity threshold of 25 mm d^{-1} (which yields 333 individual storms events across all basins during the study period).

Figure 2a shows results for total streamflow (i.e., stormflow plus baseflow). As expected, $\overline{R_{obs}}$ values are significantly positive (~0.8 [-] for both root-zone and surface-zone soil moisture from SMAP_L4) and reflect a tendency for higher antecedent soil moisture to be associated with larger event runoff coefficients (and vice versa). For the surface soil moisture case, $\overline{R_{obs}}$ values are higher than corresponding $\overline{R_{LSM}}$ sampled from internal LSM predictions. However, these differences are significant (with 95% confidence) only for VIC and generally insignificant when using root-zone soil moisture to characterize pre-storm conditions. The sharp increase in VIC and Noah $\overline{R_{LSM}}$ for the root-zone soil moisture case (versus the surface-zone case) in Figure 2a is likely due to the dominance of baseflow as a runoff generation process in VIC and Noah (see relative stormflow percentages for LSMs in Figure 2b) and the close functional relationship between root-zone soil moisture and baseflow. For VIC, it may also reflect known issues with the estimation of surface-zone soil moisture [Xia et al. 2014; 2015a].

Since our focus here is on the storm event runoff response, it is useful to filter out the impact of baseflow and isolate stormflow runoff. After removing the effects of baseflow (see Section 2), the depth of antecedent soil moisture has only a small impact on all coupling results in Figure 2b - suggesting that antecedent surface and root-zone soil moisture estimates are equally valuable for forecasting storm-scale runoff response. More importantly, the strength of the soil moisture/stormflow coupling captured by independent estimates ($\overline{R_{obs}}$) falls along the upper edge of the range provided by internal LSM predictions ($\overline{R_{LSM}}$). Differences between LSMs also emerge. For example, stormflow-based $\overline{R_{obs}}$ results in Figure 2b are significantly-higher (with 95% confidence) than comparable internal $\overline{R_{LSM}}$ estimates from Noah and Mosaic. Likewise, stormflow-based $\overline{R_{obs}}$ is larger than $\overline{R_{LSM}}$ from VIC (although not by a statistically significant amount). As discussed in the supporting materials, there are credible reasons to suspect that $\overline{R_{obs}}$ values in Figures 2b slightly underestimate the true strength of coupling between soil moisture and event runoff coefficients. Therefore, if anything, Figure 2 underestimates the magnitude of under-coupling in VIC and Noah. In contrast, the higher $\overline{R_{LSM}}$ levels predicted by SAC and CLSM are generally consistent with the $\overline{R_{obs}}$ values (Figure 2b). See Section 5 below for a process-level discussion of these LSM differences.

In addition to pre-storm soil moisture, event runoff coefficients are expected to vary as a function of storm intensity. Figure 3 plots $\bar{R}$-values (based on stormflow-only and root-zone soil moisture) that are sub-set by low (5 to 15 mm d^{-1}) and high (> 25 mm d^{-1}) storm intensity based on the observed daily rainfall accumulation on the first (triggering) day of each storm. Surprisingly, $\overline{R_{obs}}$ is marginally larger for the high-intensity events than for the low-intensity ones. This is at odds with (more intuitive) Noah and CLSM LSM results in which larger events demonstrate less sensitivity to pre-storm soil moisture conditions.

5. Discussion

It is difficult to provide a comprehensive discussion of the model-to-model variations found in Figures 2 and 3. Nevertheless, a useful contrast can be drawn between LSMs with the highest (CLSM) and lowest (Noah) soil moisture/stormflow coupling strengths in Figure 2b. As discussed in Section 2.3, these two LSMs apply contrasting approaches to the modeling of the stormflow runoff response. The response in Noah v2.8 is based on an infiltration-excess representation whereby the fractional conversion of rainfall into stormflow is driven primarily by variations in rainfall intensity [Schaake et al., 1996]. In contrast, CLSM generates surface runoff via a saturation excess process whereby stormflow is generated by rainfall incident upon portions of the landscape that have been saturated from below by a rising water table [Koster et al., 2000]. In the case of CLSM, the efficiency of stormflow generation is tied directly to the saturated land fraction of the basin which, in turn, is tightly linked with basin-averaged surface soil moisture. Figure 2b implies that this type of direct functional relationship between surface soil moisture and stormflow generation is necessary for LSMs to demonstrate sufficient internal coupling to match the levels of coupling obtained from SMAP L4 soil moisture, independent USGS streamflow observations and NLDAS-2 precipitation. A second notable signature is the tendency for the observed coupling to increase as a function of storm intensity (Figure 3). This too is at odds with the theory of infiltration-excess runoff where the relative impact of pre-storm soil moisture is predicted to decrease for high-intensity storm events [Schaake et al., 1996]. However, the observed trend of rainfall intensity on coupling is not statistically significant (see Figure 3) and potentially impacted by our inability to adequately sample across a wider range of storm intensities.

However, the parameterization of single stormflow process is also potentially important. For example, the Noah infiltration excess representation is based on a modification to a curve number approach [Schaake et al., 1996] that can be calibrated to lend varying amounts of weight to pre-storm soil moisture conditions [Massari et al., 2014]. Such parameter modifications could, in principle, correct for the significant under-coupling observed in Figure 2b. Nevertheless, if infiltration-excess runoff approaches are applied in this region, they should, at a minimum, be recalibrated to substantially increase the importance of pre-storm soil moisture.

Among the other LSMs, the 95% confidence intervals for VIC and SAC internal coupling results in Figures 2 and 3 generally overlap those obtained from SMAP_L4 and USGS observations. The single exception being the significantly low coupling observed between VIC surface soil moisture and event runoff coefficients for total runoff results in Figure 2a. On the other hand, Mosaic results generally fall between those of Noah and VIC. The overall trend of low coupling in Noah and Mosaic versus higher coupling in SAC, CLSM and VIC is consistent with variations in model complexity (with Noah and Mosaic utilizing notably-simpler approaches for stormflow generation - see Section 2.3) and is potentially reflective of the origins of SAC and VIC as hydrologic models with a more extensive history of calibration against observed streamflow. This assessment is also consistent with Xia et al. [2012a] who examined the accuracy of LSM runoff prediction versus daily streamflow observations from the NLDAS-2 LSMs and found generally superior results for VIC and SAC relative to Noah and Mosaic.

6. Summary and Conclusions

Accurately representing the relationship between pre-storm soil moisture and subsequent event runoff coefficients is a fundamental requirement of any LSM [Koster and Milly, 1997] and necessary for the successful application of LSMs to hydrologic forecasting. Utilizing a metric developed by Crow et al. [2005; 2017], we demonstrate that, within the study domain illustrated in Figure 1, soil moisture/stormflow coupling strength estimates provided by observations (i.e., SMAP_L4 soil moisture, USGS streamflow, and NLDAS-2 radar/gauge precipitation) is at the top of the inter-model range obtained from various LSMs. An apparent low bias in LSM-based coupling estimates is particularly evident for LSMs (e.g., Noah v2.8) that utilize an infiltration-excess conceptualization of stormflow (Figure 2b). Noah v2.8 also fails to match the observed variation of soil moisture/stormflow coupling with storm intensity (Figure 3). Analogous, although less severe, problems are noted for the simplified stormflow approach applied in Mosaic (Figure 2b). The implication is that these LSMs tend to squander a source of hydrologic predictability by under-utilizing their internal soil moisture estimates for forecasting variations in runoff coefficients during intense storm events. A precise diagnosis of processes (and/or parameterizations) responsible for this under-coupling will require a more incremental approach for generating LSM model variations. Modular LSMs, such as the Noah Multi-parameterization LSM (Noah-MP; Niu et al., 2011), are particularly well-suited for this purpose.

Likewise, several important caveats need to be considered. First, while LSM coupling strengths are insensitive to random errors in LSM forcing data, systematic forcing error may still have an impact. For example, coarse spatial resolution rainfall data induces a conditional bias whereby extreme rainfall rates are systematically underestimated. Since LSM runoff predictions may respond in a nonlinear manner to changes in rainfall intensity, such a conditional bias could conceivably induce systematic changes in internal LSM coupling. Therefore, our results are potentially sensitive to the spatial resolution of the rainfall forcing data used to force the LSM simulations. Second, significant bias in internal LSM results emerges only after baseflow has been separated out of both the modeled and observed streamflow time series (compare Figures 2a and 2b). Due to uncertainty in baseflow separation approaches for observed streamflow, and variations in the definition of runoff components acquired from different LSMs, there is inherent ambiguity in the cross-comparison of stormflow estimates obtained from different sources. Finally, given that the relative importance of various runoff generation processes is known to vary substantially across different climates and land cover types, a wider geographic focus is required before more general conclusions can be drawn.

Acknowledgments

Funding was provided by the NASA SMAP mission and by the NASA Terrestrial Hydrology Program (award 13-THP13-0022). Computational resources were provided by the NASA High-End Computing Program through the NASA Center for Climate Simulation at NASA/GSFC. Daily streamflow observations can be publicly accessed from USGS [2016]. NLDAS-2 precipitation data is available from Xia et al. [2009]. NLDAS-2 soil moisture and runoff results for the Noah, VIC and Mosaic LSMs are available from Xia et al. [2012a; 2012b; 2012c] and SMAP Level-4 data from Reichle et al. [2016]. CLSM daily runoff and soil moisture estimates can be accessed at Reichle and Crow [2018]. Soil moisture and runoff results for the SAC LSM are available from ftp://ldas.ncep.noaa.gov/nldas2/nco_nldas.

7. Work Cited

Burnash, R.J.C. (1995), The NWS river forecast system - Catchment modeling. In *Computer Models of Watershed Hydrology*, V.P. Singh (ed). Water Resources Publications*:* Littleton, CO*;* 311–366.

Clark, M. P., Y. Fan, D. M. Lawrence, J. C. Adam, D. Bolster, D. J. Gochis, R. P. Hooper, M. Kumar, L. R. Leung, D. S. Mackay, R. M. Maxwell, C. Shen, S. C. Swenson, & X. Zeng (2015), Improving the representation of hydrologic processes in Earth System Models, Water Resour. Res., 51, 5929-5956, doi:10.1002/2015WR017096.

Cosgrove, B.A., D. Lohmann, K.E. Mitchell, P.R. Houser, E.F. Wood, J.C. Schaake, A. Robock, C. Marshall, J. Sheffield, Q. Duan, L. Luo, R.W. Higgins, R.T. Pinker, J.D. Tarpley, & J. Meng (2003), Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *J. Geophys. Res*., 108 (D22), 8842, doi:10.1029/2002JD003118.

Crow, W.T., R. Bindlish, & T.J. Jackson (2005), The added value of spaceborne passive microwave soil moisture retrievals for forecasting rainfall-runoff partitioning, *Geophys. Res. Lett*, 32, L18401, doi:10.1029/2005GL023543.

Crow, W.T., F. Lei, C. Hain, M.C. Anderson, R.L. Scott, D. Billesbach, & T. Arkebauer (2015), Robust estimates of soil moisture and latent heat flux coupling strength obtained from triple collocation. *Geophys. Res. Lett.*, 42, 8415-8423, doi:10.1002/2015GL065929.

Crow, W.T., F. Chen, R.H. Reichle, & Q. Liu (2017), L-band microwave remote sensing and land data assimilation improve the representation of pre-storm soil moisture conditions for hydrologic forecasting, *Geophys. Res. Lett*., 44, doi:10.1002/2017GL073642.

Crow, W.T., & R.H. Reichle (2018), *Catchment-Averaged Soil Moisture and Runoff Output from the NASA GEOS Catchment Land Surface Model*, www.figshare.com online data repository, Accessed January 2018, doi:10.6084/m9.figshare.5782206.

Duan, Q., J. Schaake, V. Andréassian, S. Franks, G. Goteti, H.V. Gupta, Y.M. Gusev, F. Habets, A. Hall, L. Hay, T. Hogue, M. Huang, G. Leavesley, X. Liang, O.N. Nasonova, J. Noilhan, L. Oudin, S. Sorooshian, T. Wagener, & E.F. Wood (2006), Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, *J. Hydrol.*, 320(1–2), 3-17, doi:10.1016/j.jhydrol.2005.07.031.

Ek, M.B., K.E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, & J.D. Tarpley (2003), Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, *J. Geophys. Res*., 108, 8851, doi:10.1029/2002JD003296, D22.

Entekhabi, D., & Coauthors (2010), The Soil Moisture Active and Passive (SMAP) Mission, *Proceed. IEEE*, 98, 704-716, doi:10.1109/JPROC.2010.2043918.

Findell, K.L., P. Gentine, B.R. Lintner, & B.P. Guillod (2015), Data Length Requirements for Observational Estimates of Land–Atmosphere Coupling Strength, *J. Hydrometeorol.*, 16, 1778, 1615-1635.

Harmel, R.D., R.J. Cooper, R.M. Slade, R.L. Haney, & J.G. Arnold (2006), Cumulative uncertainty in measured streamflow and water quality data for small watersheds, *Trans. ASABE*, 49(3), 689-701.

Kim, J., M.C. Dwelle, S.K. Kampf, S. Fatichi, & V.Y. Ivanov (2016), On the non-uniqueness of the hydro-geomorphic responses in a zero-order catchment with respect to soil moisture, *Advances in Water Resources*, 92, 73-89, doi:10.1016/j.advwatres.2016.03.019

Koren V., M. Smith, D. Wang, & Z. Zhang (2000), Use of soil property data in the derivation of conceptual rainfall-runoff model parameters. *Proceedings of the 15th Conference on Hydrology*, AMS, Long Beach, CA; 103–116.

Koren V., M. Smith, & M.Q. Duan (2003), Use of a priori parameter estimates in the derivation of spatially consistent parameter sets of rainfall-runoff models. In *Advances in the Calibration of Watershed Models, Water Science and Applications Series*, vol. 6, Q. Duan, S. Sorooshia, H. Gupta, A. Rosseau, R. Turcotte (eds). AGU: Washington, DC; 239–254.

Koster, R.D., & M.J. Suarez (1996), Energy and water balance calculations in the Mosaic SLM, *Technical Report Series on the Global Modelling and Data Assimilation*, NASA Technical Memorandum 104606, Vol. 9, https://gmao.gsfc.nasa.gov/pubs/docs/Koster130.pdf

Koster, R.D., & M.J. Suarez (1994), The components of a 'SVAT' scheme and their effects on a GCM's hydrological cycle, *Adv. Water Resour.*, 17(1), 61-78, doi:10.1016/0309-1708(94)90024-8.

Koster, R.D., & P.C.D. Milly (1997), The interplay between transpiration and runoff formulations in land surface schemes used with atmospheric models, *J. Clim.*, 10, 1578-1591.

Koster, R.D., M. J. Suarez, A. Ducharne, M. Stieglitz, & P. Kumar (2000), A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure, *J. Geophys. Res.*, 105(D20), 24809–24822, doi:10.1029/2000JD900327.

Liang, X., D.P. Lettenmaier, E.F. Wood, & S.J. Burges (1994), A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *J. Geophys. Res.*, 99(D7), 14415–14428, doi:10.1029/94JD00483.

Liang, X., E.F. Wood, & D.P. Lettenmaier (1996), Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification, *Global and Planetary Change*, 13(1), 195-206, doi:10.1016/0921-8181(95)00046-1.

Massari, C., L. Brocca, S. Barbetta, C. Papathanasiou, M. Mimikou, & T. Moramarco (2014), Using globally available soil moisture indicators for flood modelling in Mediterranean catchments, *Hydrol. Earth Syst. Sci.*, 18, 839-853, doi: 10.5194/hess-18-839-2014.

Mirus, B.B., & K. Loague (2013), How runoff begins (and ends): Characterizing hydrologic response at the catchment scale, *Water Resour. Res.*, 49, 2987-3006, doi:10.1002/wrcr.20218.

National Weather Service (2007), National Weather Service Instruction 10-1605. Published online at: https://verification.nws.noaa.gov/content/pm/pubs/directives/10-1605.pdf. Accessed May 2017.

Niu, G.-Y., Z.-L. Yang, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, L. Longuevergne, A. Kumar, K. Manning, D. Niyogi, E. Rosero, M. Tewari, & Y. Xia (2011), The community Noah land surface model with multi-parameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, *J. Geophys. Res*., 116(D12), doi:10.1029/2010JD015139,

Reichle, R.H., G. De Lannoy, R.D. Koster, W.T. Crow, & J.S. Kimball (2016), *SMAP L4 9 km EASE-Grid Surface and Root Zone Soil Moisture Geophysical Data, Version 2*. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. Accessed May 2017. doi:10.5067/YK70EPDHNF0L.

Reichle, R.H., & Coauthors (2017), Assessment of the SMAP Level-4 Surface and Root-Zone Soil Moisture product using in situ measurements, *J. Hydrometeorol.*, in press, doi:10.1175/JHM-D-17-0063.1.

Schaake, J., S. Cong, & Q. Duan (2006), U.S. MOPEX data set, IAHS Publication Series, UCRL-JRNL-221228, https://e-reports-ext.llnl.gov/pdf/333681.pdf.

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

Sloto, R.A., & M.Y. Crouse (1996), HYSEP: A computer program for streamflow hydrograph separation and analysis, *U.S. Geological Survey Water-Resources Investigations Report 96-4040*, 46 p.

U.S. Geological Survey (2016), National Water Information System data available on the World Wide Web (USGS Water Data for the Nation), Accessed May 2017, at URL http://waterdata.usgs.gov/nwis.

Western A.W., & R.B. Grayson (1998), The Tarrawarra dataset: Soil moisture patterns, soil characteristics and hydrological flux measurements, *Water Resour. Res*., 34, 2765-2768.

Xia, Y., & Coauthors (2009a), *NLDAS Primary Forcing Data L4 Hourly 0.125 x 0.125 degree V002*, Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed May 2017, doi:10.5067/6J5LHHOHZHN4.

Xia, Y., & Coauthors (2012a), *NLDAS Noah Land Surface Model L4 Hourly 0.125 x 0.125 degree V002*, Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed May 2017, doi:10.5067/47Z13FNQODKV.

Xia, Y., & Coauthors (2012b), *NLDAS VIC Land Surface Model L4 Hourly 0.125 x 0.125 degree V002*, Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed May 2017, doi:10.5067/ELBDAPAKNGJ9.

Xia, Y., & Coauthors (2012c), *NLDAS Mosaic Land Surface Model L4 Hourly 0.125 x 0.125 degree V002*, Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed May 2017, doi:10.5067/EN4MBWTCENE5.

Xia, Y., & Coauthors (2012a), Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, *J. Geophys. Res*., 117, D03109, doi:10.1029/2011JD016048.

Xia, Y., & Coauthors (2012b), 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, *J. Geophys. Res*., 117, D03110, doi:10.1029/2011JD016051.

Xia, Y., M. Ek, H. Wei, & J. Meng (2012c), Comparative analysis of relationships between NLDAS-2 forcings and model outputs. *Hydrol. Process*., 26: 467–474. doi:10.1002/hyp.8240.

Xia, Y., J. Sheffield, M.B. Ek, J. Dong, N. Chaney, H. Wei, J. Meng, & E.F. Wood (2014), Evaluation of multi-model simulated soil moisture in NLDAS-2, *Journal of Hydrology*, 512, 107-125, ISSN 0022-1694, doi:10.1016/j.jhydrol.2014.02.027.

Xia, Y., M.B. Ek, Y. Wu, T. W. Ford, and S. Quiring (2015), Comparison of NLDAS-2 simulated and NASMD observed daily soil moisture. Part I: Comparison and analysis, *J. Hydrometeor*., doi:10.1175/JHM-D-14-0096.1.

Table 1. Attributes of basins in Figure 1.

Basin Number	USGS Station No.	USGS Station Name	Basin Size (km^2)	Annual *P* (mm)	Runoff Ratio *Q/P*
1	07144780	Ninnescah River AB Cheney Re, KS	2,049	768	0.08
2	07144200	Arkansas River at Valley Center, KS	3,402	842	0.11
3	07152000	Chikaskia River near Blackwell, OK	4,891	896	0.19
4	07243500	Deep Fork near Beggs, OK	5,210	945	0.15
5	07147800	Walnut River at Winfield, KS	4,855	980	0.31
6	07177500	Bird Creek Near Sperry, OK	2,360	1025	0.23
7	06908000	Blackwater River at Blue Lick, MO	2,924	1140	0.29
8	07196500	Illinois River near Tahlequah, OK	2,492	1175	0.29
9	07019000	Meramec River near Eureka, MO	9,766	1187	0.28
10	07052500	James River at Galena, MO	2,568	1255	0.31
11	07186000	Spring River near Wace, MO	2,980	1258	0.27
12	07056000	Buffalo River near St. Joe, AR	2,148	1238	0.37
13	06933500	Gasconade River at Jerome, MO	7,356	1293	0.24
14	07067000	Current River at Van Buren, MO	4,351	1309	0.31
15	07068000	Current River at Doniphan, MO	5,323	1314	0.36
16	07290000	Big Black River NR Bovina, MS	7,227	1368	0.37

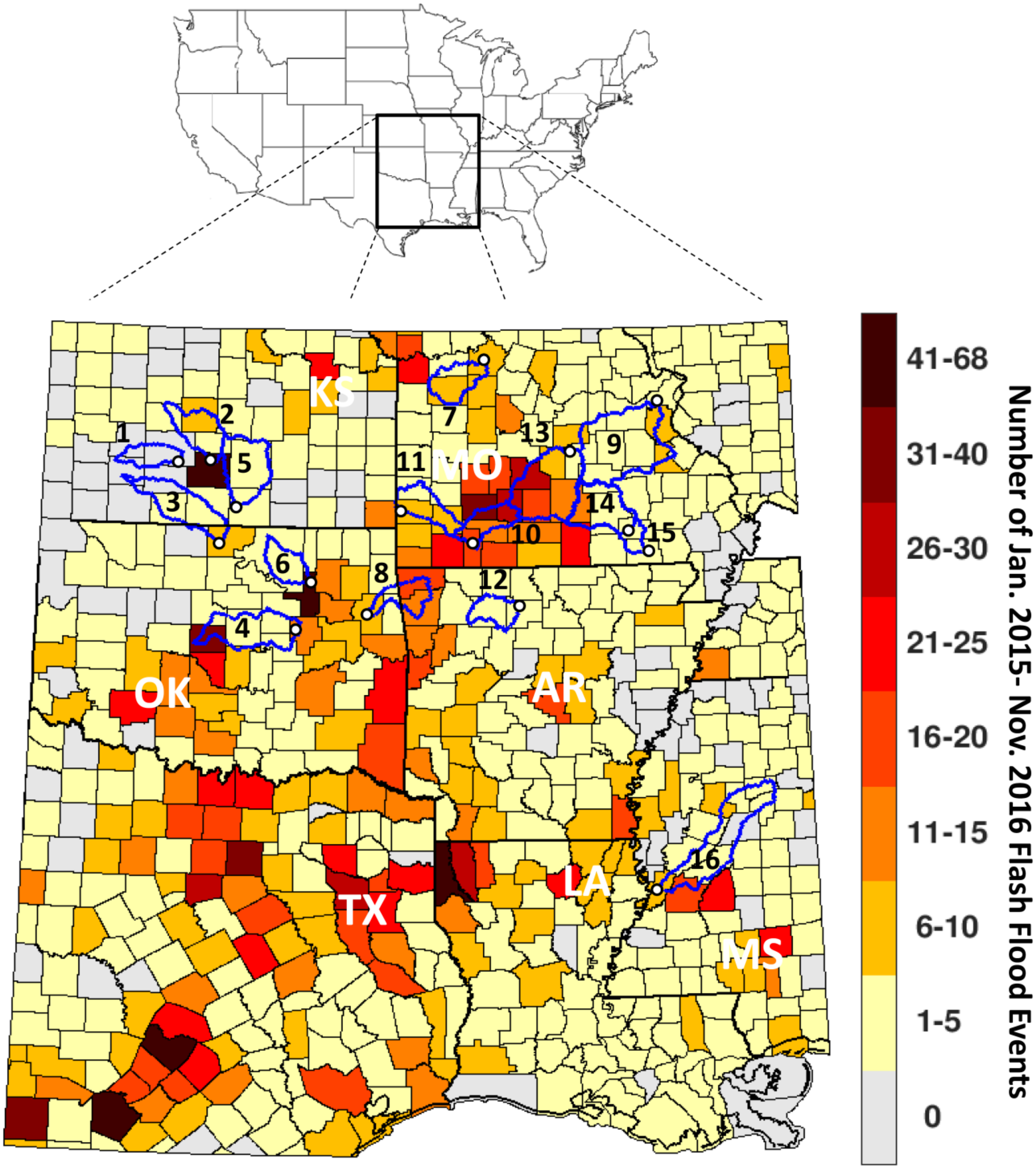


Figure 1. Locations of the 16 study basins within the south-central United States. Color shading represents a (county-scale) map of the total number of flash-flood events observed between Jan. 2015 and Nov. 2016 [NWS, 2007]. Basin numbers refer to the listing order in Table 1. Circles indicate basin outlets and USGS stream gauge locations.

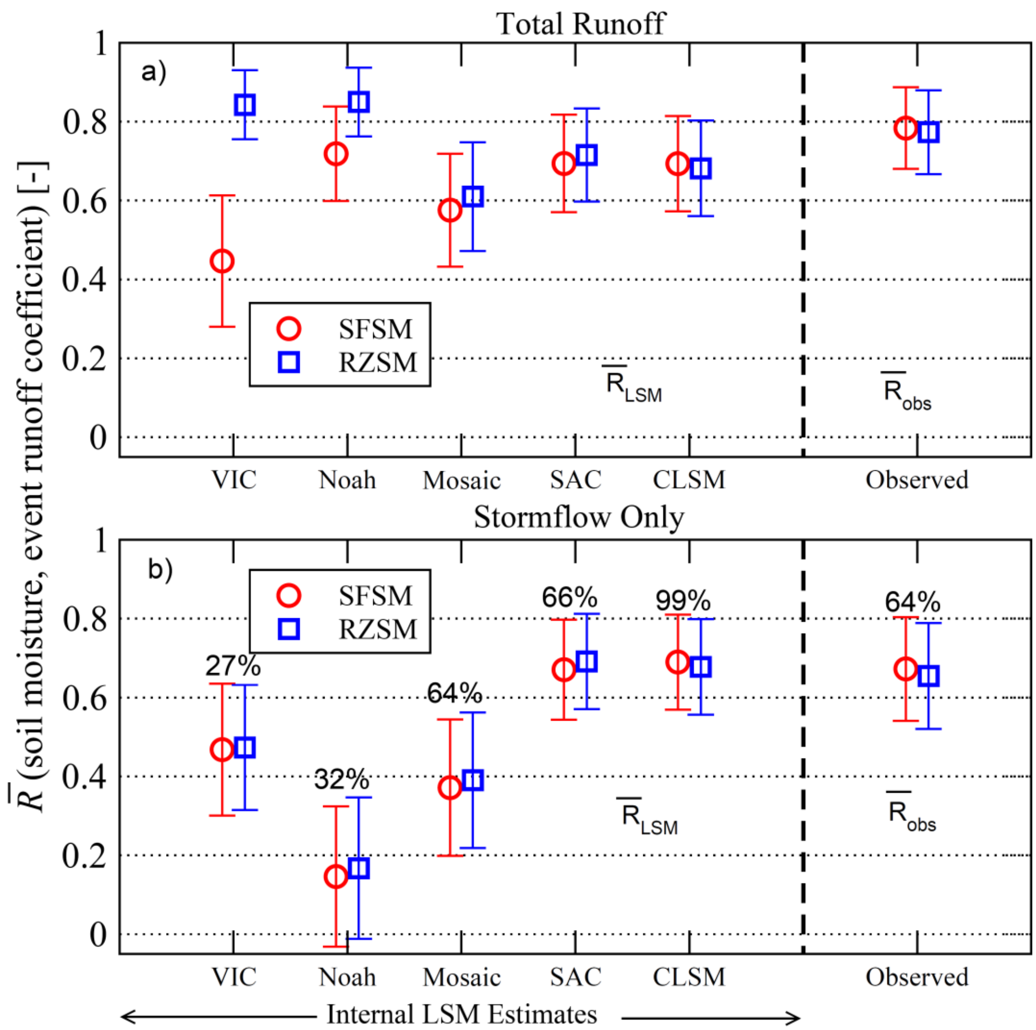


Figure 2. Values of $\bar{R}$ (with 95% confidence intervals) sampled using both a) total streamflow (i.e., stormflow + baseflow) and b) stormflow only. The left-hand side of the figure shows internal $\overline{R_{LSM}}$ predictions while the right-hand side shows $\overline{R_{obs}}$ sampled from independent SMAP_L4 soil moisture, USGS streamflow and NLDAS-2 rainfall observations. Colors/symbols indicate the use of either surface-zone (SFSM) or root-zone (RZSM) soil moisture. Numerical labels in part b) relate the percentage of total streamflow attributed to stormflow. All results are based on a triggering rainfall intensity of 25 mm d^{-1}.

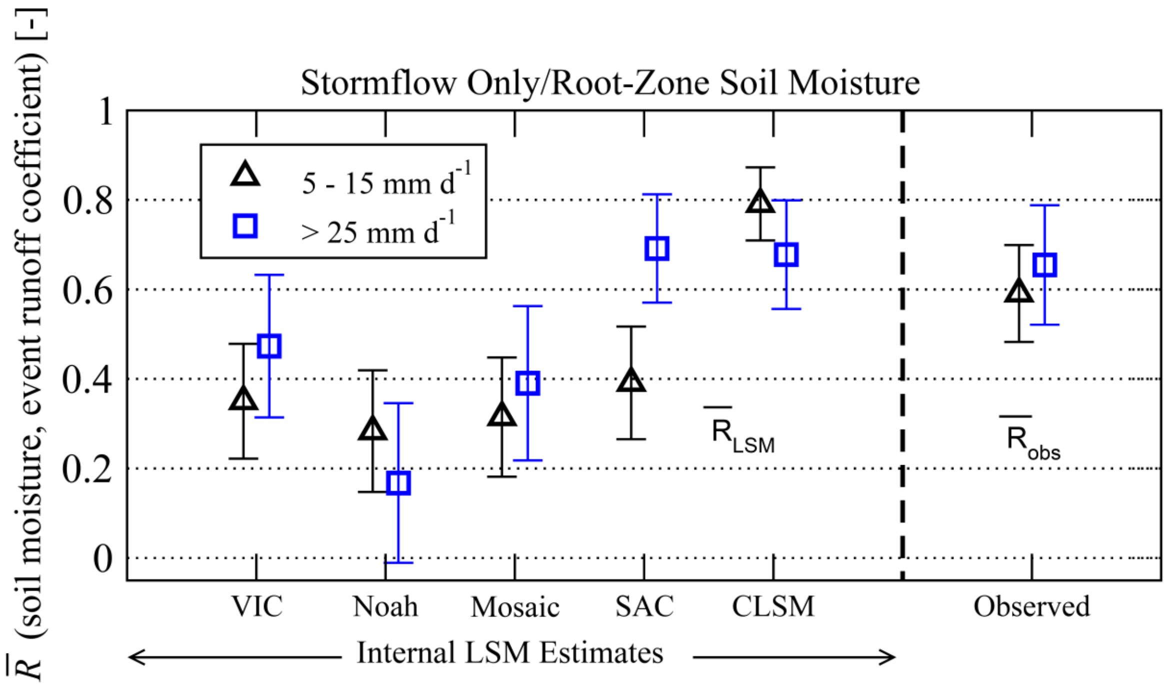
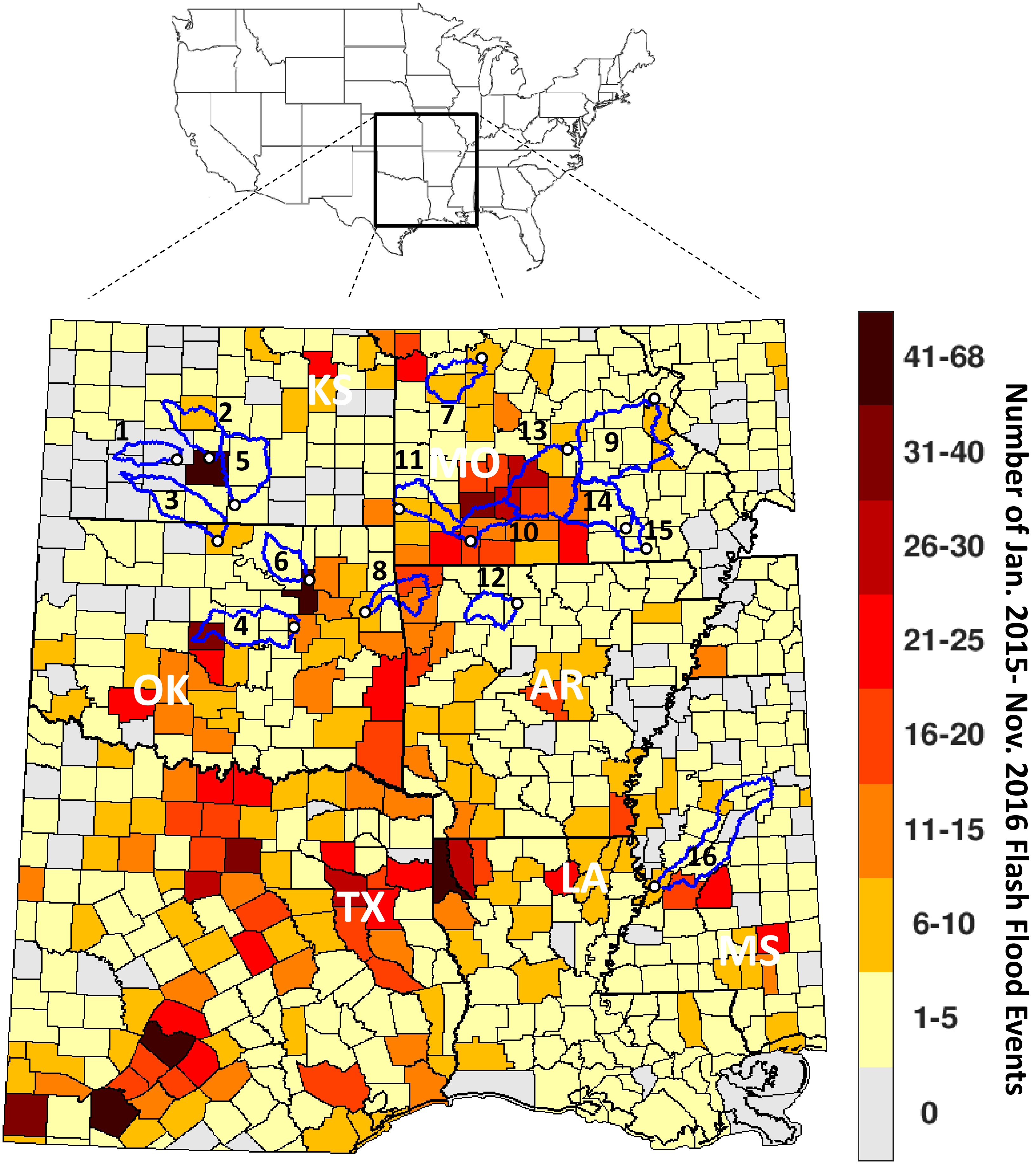


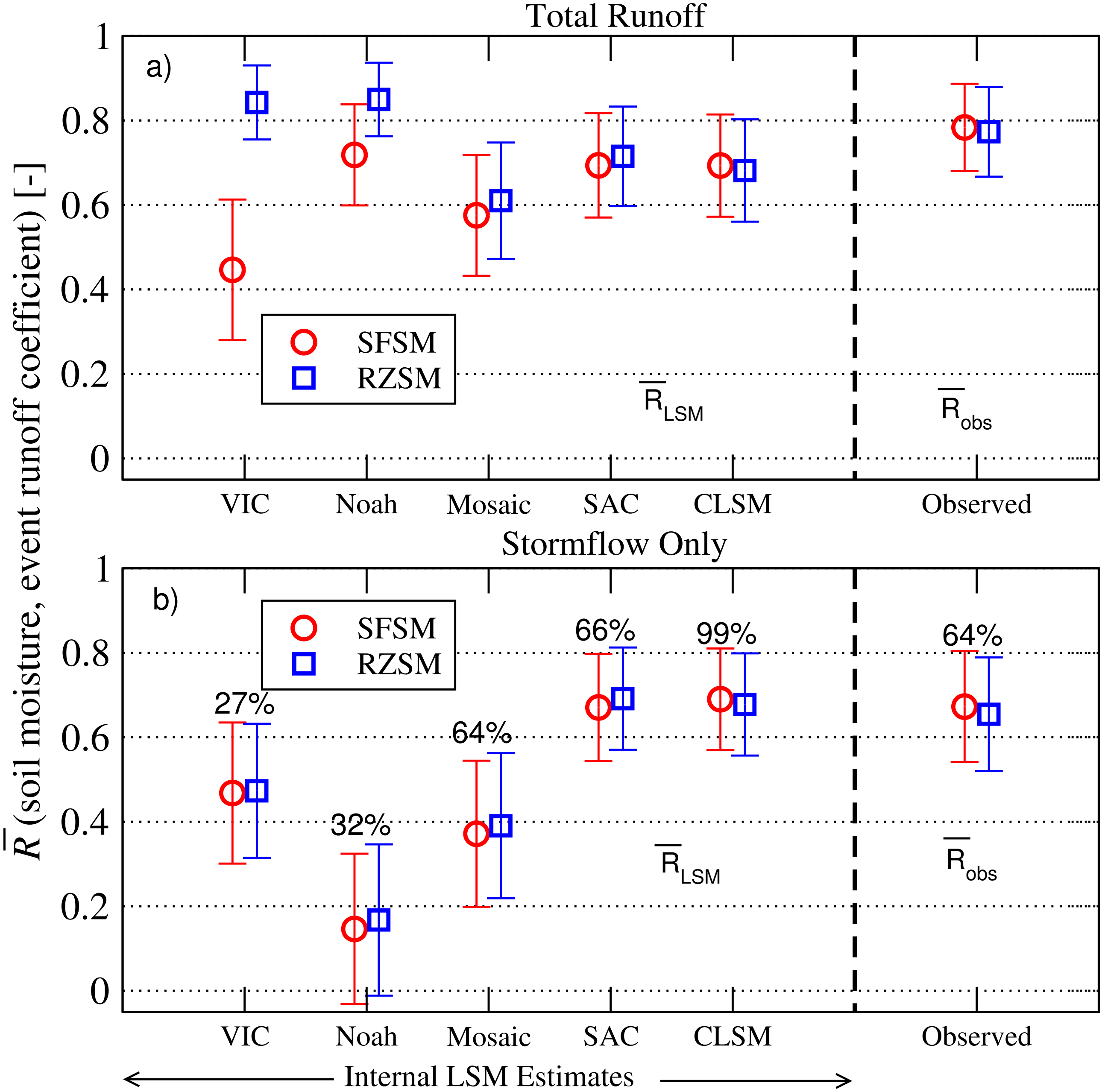
Figure 3. As in Figure 2, except colors/symbols indicate the sampling of storm events with either low (5-15 mm d^{-1}) or high (> 25 mm d^{-1}) triggering rainfall intensities. All results are based on the use of root-zone soil moisture and limited to the stormflow component of total streamflow. Note that high-intensity (> 25 mm d^{-1}) results are identical to "RZSM" results shown in Figure 2b.

Figure 1.



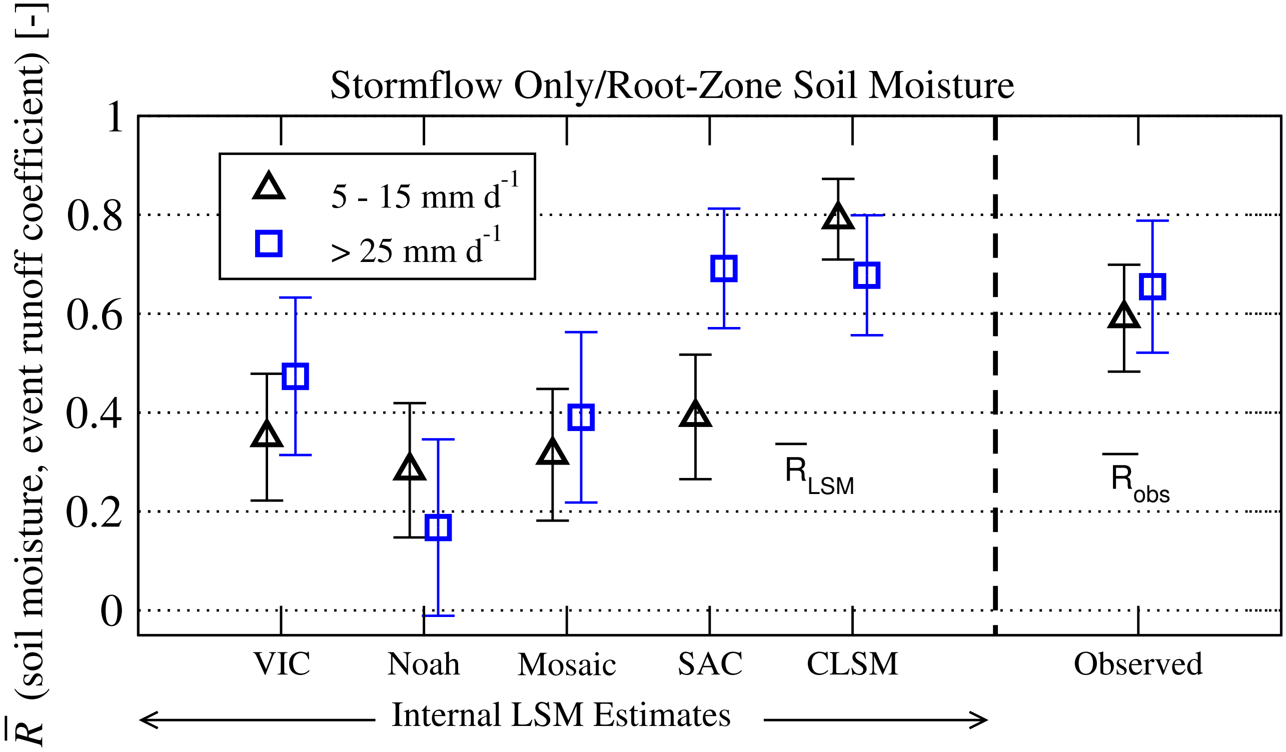
KS
MO
OK
AR
TX
LA
MS
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
41-68
31-40
26-30
21-25
16-20
11-15
6-10
1-5
0
Number of Jan. 2015- Nov. 2016 Flash Flood Events

Figure 2.



Total Runoff
a)
SFSM
RZSM
$\overline{R}_{LSM}$
$\overline{R}_{obs}$
VIC
Noah
Mosaic
SAC
CLSM
Observed
Stormflow Only
b)
27%
32%
64%
66%
99%
64%
Internal LSM Estimates
$\overline{R}$ (soil moisture, event runoff coefficient) [-]

Figure 3.



Stormflow Only/Root-Zone Soil Moisture
R̄ (soil moisture, event runoff coefficient) [-]
1
0.8
0.6
0.4
0.2
0
5 - 15 mm d-1
> 25 mm d-1
R̄LSM
R̄obs
VIC
Noah
Mosaic
SAC
CLSM
Observed
Internal LSM Estimates

Supporting Materials

Appendix A: Interpretation of Coupling Strength Estimates

As described in the main text, our primary goal here is deriving estimates of the correlation event runoff coefficient serial ranks (*RC*) and pre-storm soil moisture serial ranks (*SM*) for a given basin:

$$R = \frac{\mathrm{Cov}(RC,SM)}{\sqrt{\mathrm{Var}(RC)\ \mathrm{Var}(SM)}}. \qquad \text{(A1)}$$

The true value of this rank correlation, R_{true}, is obtained if *RC* and *SM* are free of error. Replicating this true value should be the goal of any credible land surface model (LSM). Note that the overbar on *R*, used in the main text to indicate averaging across multiple basins, in dropped here.

Of course, perfect time series representations of *RC* and *SM* are never available. Instead, we rely on uncertain estimates of these quantities. These estimates can be approximated as:

$$RC_{est} = RC + \xi_{RC} \qquad \text{(A2)}$$

$$SM_{est} = SM + \xi_{SM}$$

where ξ_{RC} and ξ_{SM} are mean-zero, error variables in ranks. Note that scaling gain factors are neglected in (A2) since such factors have no bearing on the correlation-based discussion which follows below.

Without making any statistical assumptions regarding these errors, re-calculating (A1) using estimated values in (A2), yields:

$$R_{est} = \frac{[\mathrm{Cov}(RC,SM) + \mathrm{Cov}(\xi_{RC},\xi_{SM}) + \mathrm{Cov}(RC,\xi_{SM}) + \mathrm{Cov}(\xi_{RC},SM)]}{\sqrt{[\mathrm{Var}(RC+\xi_{RC})]\ [\mathrm{Var}(SM+\xi_{SM})]}}. \qquad \text{(A3)}$$

Equation (A3) can be used to estimate the impact of errors in *RC* and *SM* on ranks correlations sampled from real data. However, several important distinctions should be made between estimates of rank correlation derived from largely independent, observation-based sources (i.e., R_{obs} computed using *SM* from a data assimilation analysis and *external RC* obtained from independent rain and stream gauge observations) and estimates of *R* derived from internally consistent LSM estimates (i.e., R_{LSM} computed from *internal* model estimates of soil moisture, runoff, and precipitation, with the latter also used to force the LSM and generate the LSM soil moisture and runoff estimates). These issues are discussed in depth below.

A.1. Impact of random error

By construction, R_{LSM} is insensitive to purely random errors in the LSM forcing (see main text Section 3.1). Therefore, if (A1) is applied to LSM-internal *RC* and *SM* results provided by a physically realistic and unbiased simulation, the resulting correlation (R_{LSM}) should indeed match R_{true}. In contrast, an observation-based correlation estimate (R_{obs}) is biased low in the presence of independent random error in *RC* and *SM*. This can be illustrated by assuming wholly independent and orthogonal observational errors in (A2) and, thereby, simplifying (A3) to:

$$R_{obs} = \frac{[\mathrm{Cov}(RC,SM)]}{\sqrt{[\mathrm{Var}(RC)+\mathrm{Var}(\xi_{RC})]\,[\mathrm{Var}(SM)+\mathrm{Var}(\xi_{SM})]}}\ . \qquad \text{(A4)}$$

The sole difference between (A4) and (A1) is that the denominator of (A3) is inflated by the additional random error variance associated with uncertain SM_{est} and RC_{est} observations in (A2). Given that $\mathrm{Cov}(RC, SM)$ is almost always positive (see main text), this induces a negative bias into R_{obs} relative to R_{LSM} (or R_{true}).

The magnitude of this bias is determined by the signal-to-noise (SNR) characteristics of SM_{est} and RC_{est} (with lower SNR associated with a larger degradation on R_{obs}). Based on this reasoning, Crow et al. [2017] argued that the size of R_{obs} can be interpreted as a proxy for the skill of various soil moisture products in estimating pre-storm soil moisture. Specifically, they found that the SMAP_L4 product provides more pre-storm soil moisture information for short-term hydrologic forecasts than other remotely sensed product - including the SMAP Level 3 soil moisture product (SMAP_L3). However, since factors other than soil moisture also impact *RC*, even perfect *SM* and *RC* observations should not be expected to yield an R_{obs} of one.

A.2. Impact of non-random error

If errors in (A2) are not wholly random, the interpretation of (A3) is complicated by the non-zero error covariance terms within its numerator. For example, the SMAP_L4 system contains a land surface modeling component and cannot be considered a purely independent observation. In particular, the GEOS-5 precipitation product used to force the SMAP_L4 assimilation model is gauge-corrected using a set of rain gauges which overlap with an analogous correction applied to the NLDAS-2 precipitation product. Consequently, there exists the possibility for cross-correlated error to arise between SMAP_L4-based SM_{est} and NLDAS-2 rainfall accumulation observations used to calculation observation-based RC_{est}. If present, such error correlation would cause the *overestimation* of pre-storm soil moisture (due to the overestimation of pre-storm rainfall) to be associated with the *underestimation* of storm-scale runoff efficiency (due to the continued overestimation of within-storm rainfall used to normalize streamflow) and vice versa. As such, it would lead to $\mathrm{Cov}(\xi_{RC}, \xi_{SM}) \leq 0$ in (A3).

Similar considerations should be made for the $\mathrm{Cov}(RC, \xi_{SM})$ term in (A3). Errors in the SMAP_L3 retrieval product are known to be linked with inter-annual vegetation variability [Dong et al., 2018]. Given that there is overlap in the ancillary vegetation parameters used in the SMAP_L4 and SMAP_L3 retrieval approaches, and inter-annual variability in vegetation can conceivably be linked to surface infiltration properties (and thus *RC*), non-zero $\mathrm{Cov}(RC, \xi_{SM})$

could conceivably arise from pronounced levels of inter-annual vegetation variability. However, this connection is tenuous and our study region is, in fact, characterized by relatively *low* levels of inter-annual vegetation variability during the SMAP data era [Dong et al., 2018]. Therefore, the $\mathrm{Cov}(RC, \xi_{SM})$ term in (A3) is expected to be negligible. Likewise, we are not aware of physical arguments for why error in observed RC_{est} (derived solely from ground-based rain gauge, weather radar and stream gauge observations) would be correlated with true pre-storm *SM* levels. Therefore, the $\mathrm{Cov}(SM, \xi_{RC})$ term in (A3) is also assumed to negligible.

In summary, given that $\mathrm{Cov}(RC, SM) > 0$ (see Figure 2b in the main text), the three (non-random error) tendencies identified here (i.e., $\mathrm{Cov}(\xi_{RC}, \xi_{SM}) \leq 0$, $\mathrm{Cov}(RC, \xi_{SM}) \sim 0$ and $\mathrm{Cov}(SM, \xi_{RC}) \sim 0$) should, if anything, cause R_{obs} to be slightly biased low relative to R_{true}.

A.3. Impact of runoff routing error

A final consideration for calculating R_{obs} using observed streamflow is accounting for the time lag between incident rainfall and observed streamflow at the basin outlet. Here, we assumed a fixed, 6-hour time lag between incident rainfall fall and streamflow response measured at basin outlets (see Section 2.1 in the main text). More complex runoff routing procedures (including, for example, the explicit calibration of basin-dependent time lags) would almost certainly increase R_{obs} but were not applied to avoid the artificial enhancement of R_{obs} via explicit tuning. Therefore, the simplicity of the routing approach used here introduces a potential source of low bias into (positive) R_{obs} values. Note that an analogous issue does not exist for R_{LSM} estimates since LSM *SM* are compared to LSM *RC* derived directly from (un-routed) LSM runoff estimates.

A.4. Summary of impacts

All considerations detailed above suggest that (non-negative) R_{obs} results presented in the main text will, if anything, be slightly biased low relative to reference R_{true} values (hypothetically) sampled from perfect *SM* and *RC* products.

A.5. Work Cited

Dong, J., W.T. Crow, & R. Bindlish (2018). The error structure of the SMAP single- and dual-channel soil moisture retrievals, *Geophysical Research Letters*, 45, 758–765, doi: 10.1002/2017GL075656.