Extracting hydrologic information from the Soil Moisture Active Passive (SMAP) satellite data for improved hydrologic modeling

Yixin Mao

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Bart Nijssen, Chair

Wade T. Crow

Faisal Hossain

Program Authorized to Offer Degree:

Civil and Environmental Engineering

University of Washington

Abstract

Extracting hydrologic information from the Soil Moisture Active Passive (SMAP) satellite data for improved hydrologic modeling

Yixin Mao

Chair of the Supervisory Committee:
Professor WOT Bart Nijssen
Civil and Environmental Engineering

Soil moisture is a key component of the water cycle. The launch of the new-generation satellite mission, the NASA Soil Moisture Active/Passive (SMAP) mission, in 2015 opens up unprecedented opportunities for researchers to learn more about surface soil moisture behavior and to then improve hydrologic modeling. In this dissertation, various ways of extracting hydrologic information from the SMAP surface soil moisture data are explored and their ability to improve hydrologic modeling (both model results and process representation) is assessed. First, data assimilation techniques are applied to incorporate the SMAP data to update modeled soil moisture states. Then, rainfall correction techniques are applied to use SMAP to back-correct rainfall estimates from the Global Precipitation Measurement (GPM) mission. These updated soil moisture states and precipitation estimates are then combined to improve simulated streamflow. This work shows that SMAP soil moisture assimilation only slightly nudges rainfall and streamflow estimates in the correct direction. One main reason for the small hydrologic improvement is that the Kalman-filter-based soil moisture data assimilation techniques are only

able to correct zero-mean random error in a hydrologic simulation system, but not the often more substantial systematic error.

These findings motivate the last part of this dissertation, in which surface soil moisture dynamics are directly derived from SMAP via a data-driven, multivariate regression approach. The SMAP-derived dynamics include surface moisture decay rate, fraction of precipitation retained in the surface layer, and the dependency of the infiltration/runoff partition process on antecedent moisture level. These governing dynamics derived from SMAP are compared with those derived from a model-based global dataset, and inaccuracies in the model setup are pointed out, including slow surface moisture decay, small sensitivity of the infiltration/runoff partition process to the top-layer moisture, and lack of spatial variation in surface soil moisture dynamics. This work demonstrates the potential of using the extracted information from SMAP to evaluate and improve process representation in hydrologic models.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to sincerely thank Dr. Bart Nijssen, my academic advisor. This dissertation would not be possible without his continuous guidance, support and encouragement. I would also like to thank Dr. Wade Crow, who co-led my dissertation project, for sharing his time generously and providing guidance that was essential for me to complete this dissertation. I would like to thank my dissertation committee member, Dr. Faisal Hossain, my Graduate School Representative, Dr. Qiang Fu, and my General Exam Graduate School Representative, Dr. Donald Percival. I am thankful to be part of the UW Computational Hydrology Group, led by Dr. Bart Nijssen, as well as the former UW Land Surface Hydrology Group, led by Dr. Dennis Lettenmaier. Thank you to all my current and former labmates for your support, encouragement and friendship. I would also like to thank Mu Xiao, who helped me start on my research work in my first year of graduate school, and Dr. Dennis Lettenmaier, who guided me to write my first research paper.

Thank you to my family and friends who were always there to give me emotional support throughout my graduate study.

This dissertation was funded in part by NASA Terrestrial Hydrology Program Award NNX16AC50G to the University of Washington and NASA Terrestrial Hydrology Program Award 13-THP13-0022 to the United States Department of Agriculture, Agricultural Research Service. I also received a Pathfinder Fellowship by CUAHSI with support from the National Science Foundation (NSF) Cooperative Agreement No. EAR-1338606, which enabled my in-person visit to Dr. Wade Crow. This work was enabled by the use of the Hyak supercomputer system at the University of Washington, which provided advanced computational resources and data storage.

DEDICATION

For my mom.

Chapter 1. INTRODUCTION

1.1 BACKGROUND

Soil moisture is a key component of the water cycle. Although it only comprises a small part of the global fresh water volume, soil moisture plays an important role in the complex water and energy exchanges near the land surface. Soil moisture participates in exchange processes at the land surface - atmosphere interface such as infiltration, vegetation water uptake, evaporation, and partitioning of incoming energy between sensible heat, latent heat and ground heat fluxes as well as reflected and emitted radiation. It also plays a role between the land surface and deeper layers, such as soil water drainage and groundwater recharge. Soil moisture also partly controls runoff generation both by partitioning precipitation into infiltration and surface runoff during rainfall events, and by determining the rate of subsurface flow [e.g. Freeze and Harlan, 1969; Western et al., 2002; Aubert et al., 2003]. A correct understanding of soil moisture dynamics is therefore crucial for hydrologic science and applications.

In 2015, a new soil moisture satellite mission, the NASA Soil Moisture Active Passive (SMAP) mission [Entekhabi et al., 2010], was launched. SMAP provides high-quality, near-real-time surface soil moisture measurements with near-global coverage. The SMAP product contains unprecedentedly rich information about surface soil moisture. Together with the new-generation satellite precipitation measurements, the Global Precipitation Measurement (GPM) mission [Hou et al., 2014], there is great potential to extract land surface hydrology information from these datasets and to improve hydrologic modeling ability around the globe. This is especially promising for regions with scarce in-situ soil moisture measurements, rain gauges and streamflow measurement sites.

In recent decades, researchers have developed various approaches to extract information contained in large-scale soil moisture measurements (from either SMAP or previous-generation satellite soil moisture products), and to then improve model results directly or to improve the fundamental understanding of hydrologic processes. These approaches include:

1) Soil moisture data assimilation (DA). This is the most direct way to incorporate soil moisture measurements into hydrologic modeling to subsequently improve streamflow

simulation. In DA, measurements are used to update modeled soil moisture states sequentially over time. The updated moisture states then, ideally, lead to improved streamflow simulation. A number of recent studies have applied Kalman-filter-based techniques, which is one of the most widely used data assimilation techniques, to assimilate remotely-sensed soil moisture data to improve simulated streamflow [e.g., Pauwels et al., 2002; Francois et al., 2003; Parajka et al., 2006; Brocca et al., 2010; Wanders et al., 2014; Alvarez-Garreton et al., 2014; Kumar et al., 2014; Lievens et al., 2015; Massari et al., 2015]. However, these studies found different levels of streamflow improvement in response to soil moisture assimilation, ranging from good improvement to no improvement or even degraded streamflow simulations.

2) Rainfall correction. Since soil moisture responds directly to rainfall inputs, soil moisture measurements presumably contain information about antecedent rainfall events. This information can potentially be used to correct inaccurate rainfall estimates (for example, from satellite rainfall observation). Recent studies have explored techniques to back-calculate rainfall or to correct existing rainfall products using satellite soil moisture data with promising results [e.g., Crow et al., 2011; Chen et al., 2012; Brocca et al., 2013; Brocca et al., 2014; Brocca et al., 2016; Koster et al., 2016; Koster et al., 2018]. Combining rainfall correction techniques with soil moisture state updating (as mentioned above), a so-called dual state/rainfall correction system was developed to use soil moisture measurements to simultaneously correct modeled states and rainfall estimates, with the goal of subsequently improving streamflow simulations [e.g., Crow and Ryu, 2009; Chen et al., 2014; Alvarez-Garreton et al., 2016].

3) Soil moisture dynamics extraction. A number of new studies have tried to extract surface soil moisture characteristics from SMAP, such as the soil moisture decay rate [e.g., Shellito et al., 2016; McColl et al., 2017b; Koster et al., 2017; Akbar et al., 2018; Shellito et al., 2018] and soil moisture memory [McColl et al., 2017a]. These extracted soil moisture characteristics are potentially useful for better understanding hydrologic processes at coarse spatial resolution and improving continental-scale hydrologic modeling setups that have a spatial resolution commensurate with the resolution of SMAP. However, little research has been done to use the process understanding drawn from SMAP to improve hydrologic modeling setups. One exception is a recent study by Shellito et al. [2018], who compared the SMAP decay rates with the surface layer decay rates simulated by a land surface model.

1.2 OBJECTIVES AND RESEARCH QUESTIONS

The overarching objective of this dissertation is to assess the ability to use the extracted hydrologic information from the SMAP surface soil moisture data to improve hydrologic modeling. At the beginning phase of my dissertation work, the primary research goal was to directly integrate the SMAP data to improve simulated streamflow results from a physically-based land surface model using DA techniques (which comprises Chapter 2 and Chapter 3 of my dissertation). As the research went on, only slight streamflow improvement was achieved via such a direct data-model integration, and we found that the theory and assumptions behind the DA techniques limit their ability to reduce the often-large systematic bias existing in streamflow simulations. Instead, the reduction of such systematic errors relies on a better overall representation of hydrologic processes in a hydrologic model. Therefore, I shifted my research focus to directly extract information about soil moisture dynamics from SMAP and use this information to enhance our understanding of process representation in hydrologic models (which comprises Chapter 4 of my dissertation). These research questions and efforts are all built upon the literature reviewed above, and the overall research questions of this dissertation are summarized below.

Research questions:

1) To what extent can the SMAP surface soil moisture data improve streamflow simulation via state updating data assimilation techniques, and why?

2) To what extent can the SMAP surface soil moisture data correct the GPM precipitation estimates, and why?

3) To what extend can the SMAP surface soil moisture data improve streamflow simulation via a dual state/rainfall correction system?

4) How can we extract soil moisture dynamics from the SMAP data and use this information to evaluate process representation in hydrologic models?

To address Question 1 (Chapter 2), a Kalman-filter-based state updating system is applied to SMAP in a regional-scale basin, and the improvement in performance of streamflow simulations is assessed. Detailed decomposition of data assimilation steps is examined to provide insights to the underlying factors that affect the streamflow performance. To address Questions 2

and 3 (Chapter 3), an existing dual state/rainfall correction system is extended and applied to the SMAP soil moisture and GPM precipitation data in a regional-scale basin, and the improvement of both precipitation estimates and streamflow simulations is assessed. The reasons for the level of improvement in precipitation and streamflow simulations are investigated. Finally, to address Question 4 (Chapter 4), a unified multivariate regression method is proposed and applied to the global SMAP data to derive the governing factors for surface soil moisture dynamics. The extracted dynamics are then compared to a model-based global surface soil moisture dataset to inform its modeling setup.

REFERENCES

Akbar, R., D. J. Short Gianotti, K. A., McColl, E. Haghighi, G. D. Salvucci, and D. Entekhabi (2018), Estimation of landscape soil water losses from satellite observations of soil moisture, *J. Hydrometeorol.*, 19 (5), 871-889, doi:10.1175/JHM-D-17-0200.1.

Alvarez-Garreton, C., D. Ryu, A. W. Western, W. T. Crow, and D. E. Robertson (2014), The impacts of assimilating satellite soil moisture into a rainfall-runoff model in a semi-arid catchment, *J. Hydrol.*, 519, 2763-2774, doi:10.1016/j.jhydrol.2014.07.041.

Alvarez-Garreton, C., D. Ryu, A. W. Western, W. T. Crow, C.-H. Su, and D. R. Robertson (2016), Dual assimilation of satellite soil moisture to improve streamflow prediction in data-scarce catchments, *Water Resour. Res.*, 52(7), 5357-5375, doi:10.1002/2015WR018429.

Aubert, D., C. Loumagne, and L. Oudin (2003), Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall-runoff model, *J. Hydrol.*, 280(1-4), 145-161, doi:10.1016/S0022-1694(03)00229-4.

Brocca, L., F. Melone, T. Moramarco, W. Wagner, V. Naeimi, Z. Bartalis, and S. Hasenauer (2010), Improving runoff prediction through the assimilation of the ASCAT soil moisture product, *Hydrol. Earth Syst. Sci.*, 14, 1881-1893, doi:10.5194/hess-14-1881-2010.

Brocca, L., T. Moramarco, F. Melone, and W. Wagner (2013), A new method for rainfall estimation through soil moisture observations, *Geophys. Res. Lett.*, 40, 853–858, doi:10.1002/grl.50173.

Brocca, L., L. Ciabatta, C. Massari, T. Moramarco, S. Hahn, S. Hasenauer, R. Kidd, W. Dorigo, W. Wagner, and V. Levizzani (2014), Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data, *J. Geophys. Res. Atmos.*, 119, 5128–5141, doi:10.1002/2014JD021489.

Brocca, L., T. Pellarin, W. T. Crow, L. Ciabatta, C. Massari, D. Ryu, C.-H. Su, C. Rüdiger, and Y. Kerr (2016), Rainfall estimation by inverting SMOS soil moisture estimates: A comparison of different methods over Australia, *J. Geophys. Res. Atmos.*, 121, 12,062–12,079, doi:10.1002/2016JD025382.

Chen F., W. T. Crow, and T. R. H. Holmes (2012), Improving long-term, retrospective precipitation datasets using satellite-based surface soil moisture retrievals and the Soil Moisture Analysis Rainfall Tool, *J. Appl. Remote Sens.*, 6(1), 063604, doi: 10.1117/1.JRS.6.063604.

Chen, F., W. T. Crow, and D. Ryu (2014), Dual forcing and state correction via soil moisture assimilation for improved rainfall–runoff modeling, *J. Hydrometeorol.*, *15*(5), 1832–1848, doi:10.1175/JHM-D-14-0002.1.

Crow, W. T., and D. Ryu (2009), A new data assimilation approach for improving hydrologic prediction using remotely-sensed soil moisture retrievals, *Hydrol. Earth Syst. Sci.*, 12(1-16), doi:10.5194/hess-13-1-2009.

Crow, W. T., M. J. van den Berg, G. J. Huffman, and T. Pellarin (2011), Correcting rainfall using satellite-based surface soil moisture retrievals: The Soil Moisture Analysis Rainfall Tool (SMART), *Water Resour. Res.*, 47, W08521, doi:10.1029/ 2011WR010576.

Entekhabi, D. et al. (2010), The Soil Moisture Active and Passive (SMAP) Mission, *Proceedings of the IEEE*, 98(5), 704-716, doi:10.1109/JPROC.2010.2043918.

Francois, C., Quesney, A., and C. Ottle (2003), Sequential assimilation of ERS-1 SAR data into a coupled land surface-hydrological model using an extended Kalman filter, *J. Hydrometeorol.*, 4(2), 473-487, doi:10.1175/1525-7541(2003)4<473:SAOESD>2.0.CO;2.

Freeze, R. A., and R. L. Harlan (1969), Blueprint for a physically-based, digitally-simulated hydrologic response model, *J. Hydrol.*, 9(3), 237-258, doi:10.1016/0022-1694(69)90020-1.

Hou, A. Y., R. K. Kakar, S. Neeck, A. A. Azarbarzin, C. D. Kummerow, M. Kojima, R. Oki, K. Nakamura, and T. Iguchi (2014), The Global Precipitation Measurement mission, *Bull. Amer. Meteor. Soc.*, 95(5), 701-722, doi: 10.1175/BAMS-D-13-00164.1.

Koster, R. D., L. Brocca, W. T. Crow, M. S. Burgin, and G. J. M. De Lannoy (2016), Precipitation estimation using L-band and C-band soil moisture retrievals, *Water Resour. Res.*, 52, 7213-7225, doi:10.1002/2016WR019024.

Koster, R. D., R. H. Reichle, and S. P. P. Mahanama (2017), A data-driven approach for daily real-time estimates and forecasts of near-surface soil moisture, *J. Hydrometeorol.*, 18 (3), 837-843, doi:10.1175/JHM-D-16-0285.1.

Koster, R. D., W. T. Crow, R. H. Reichle, and S. P. Mahanama (2018), Estimating basin-scale water budgets with SMAP soil moisture data, *Water Resour. Res.*, 54, doi:10.1029/2018WR022669.

Kumar, S. V., C. D. Peters-Lidard, D. M. Mocko, R. Reichle, Y. Liu, K. R. Arsenault, Y. Xia, M. Ek, G. Riggs, B. Livneh, and M. Cosh (2014), Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation, *J. Hydrometeorol.*, 15, 2446-2469, doi:10.1175/JHM-D-13-0132.1.

Lievens, H., et al. (2015), SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia, *Remote Sens. Environ.*, *168*, 146-162, doi:10.1016/j.rse.2015.06.025.

Massari, C., L. Brocca, A. Tarpanelli, and T. Moramarco (2015), Data Assimilation of Satellite Soil Moisture into Rainfall-Runoff Modelling: A Complex Recipe?, *Remote Sens.*, 7, 11403-11433, doi:10.3390/rs70911403.

McColl, K. A., S. H. Alemohammad, R. Akbar, A. G. Konings, S. Yueh, and D. Entekhabi (2017a), The global distribution and dynamics of surface soil moisture, *Nat. Geosci.*, 10, 100-104, doi:10.1038/NGEO2868.

McColl, K. A., W. Wang, B. Peng, R. Akbar, D. J. Short Gianotti, H. Lu, M. Pan, and D. Entekhabi (2017b), Global characterization of surface soil moisture drydowns, *Geophys. Res. Lett.*, 44, 3682–3690, doi:10.1002/2017GL072819.

Parajka, J., V. Naemi, G. Bloschl, W. Wagner, R. Merz, and L. Scipal (2006), Assimilating scatterometer soil moisture data into conceptual hydrologic models at the regional scale, *Hydrol. Earth Syst. Sci.*, 10, 353-368, doi:10.5194/hess-10-353-2006.

Pauwels, R., R. Hoeben, N. Verhoest, F. De Troch, and P. Troch (2002), Improvement of TOPLATS-based discharge predictions through assimilation of ERS-based remotely sensed soil moisture values, *Hydrol. Process.*, 16, 995-1013, doi:10.1002/ hyp.315.

Shellito, P. J., et al. (2016), SMAP soil moisture drying more rapid than observed in situ following rainfall events, *Geophys. Res. Lett.*, 43, 8068-8075, doi:10.1002/2016GL069946.

Shellito, P. J., E. E. Small, and B. Livneh (2018), Controls on surface soil drying rates observed by SMAP and simulated by the Noah land surface model, *Hydrol. Earth Syst. Sci.*, 22, 1649-1663, doi:10.5194/hess-22-1649-2018.

Wanders, N., D. Karssenberg, A. De Roo, S. M. De Jong, and M. F. P. Bierkens (2014), The suitability of remotely sensed soil moisture for improving operational flood forecasting, *Hydrol. Earth Syst. Sci.*, *18*(6), 2343-2357, doi:10.5194/hess-18-2343-2014.

Western, A. W., R. B. Grayson, and G. Blöschl (2002), Scaling of soil moisture: a hydrologic perspective, *Annu. Rev. Earth Planet. Sci.*, *30*(1), 149-180, doi:10.1146/annurev.earth.30.091201.140434.

Chapter 2. A FRAMEWORK FOR DIAGNOSING FACTORS DEGRADING THE STREAMFLOW PERFORMANCE OF A SOIL MOISTURE DATA ASSIMILATION SYSTEM

This chapter has been accepted for publication in its current form in the *Journal of Hydrometeorology*. © American Meteorological Society. Used with permission. The supplemental material for this chapter is provided in appendix A.

Mao Y., W. T. Crow, and B. Nijssen (2018), A framework for diagnosing factors degrading the streamflow performance of a soil moisture data assimilation system, *J. Hydrometeorol.*, *accepted*, doi:10.1175/JHM-D-18-0115.1.

Abstract

Data assimilation (DA) techniques have been widely applied to assimilate satellite-based soil moisture (SM) measurements into hydrologic models to improve streamflow simulations. However, past studies have reached mixed conclusions regarding the degree of runoff improvement achieved via SM state updating. In this study, a synthetic diagnostic framework is designed to 1) decompose the random error components in a hydrologic simulation, 2) quantify the error terms that originate from SM states, and 3) assess the effectiveness of SM DA to correct these random errors. The general framework is illustrated through a case study in which surface Soil Moisture Active Passive (SMAP) data is assimilated into a large-scale land surface model in the Arkansas-Red River basin. The case study includes systematic error in the simulated streamflow that imposes a first-order limit on DA performance. In addition, about 60% of the random runoff error originates directly from rainfall and cannot be corrected by SM DA. In particular, fast-response runoff dominates in much of the basin but is relatively unresponsive to state updating. Slow-response runoff is strongly controlled by the bottom-layer SM and therefore only modestly improved via the assimilation of surface measurements. Combined, the total runoff improvement in the synthetic analysis is small (<10% over the basin). Improvements in the real SMAP-assimilated case are further limited due to systematic error and other factors such as

inaccurate error assumptions and SMAP rescaling. Findings from the diagnostic framework suggest that SM DA alone is insufficient to substantially improve streamflow estimates in large basins.

2.1 INTRODUCTION

Soil moisture (SM) is a key component of the hydrologic cycle. For hydrologic prediction purposes, SM plays an important role in runoff generation both by partitioning infiltration and surface runoff during rainfall events and by controlling the rate of subsurface flow during inter-storm periods [e.g. Freeze and Harlan, 1969; Western et al., 2002; Aubert et al., 2003]. Therefore, improved estimation of SM states in hydrologic models is potentially beneficial for streamflow prediction.

A number of studies have explored the potential to assimilate SM measurements into hydrologic models for runoff simulation improvement, either via synthetic twin experiments [e.g., Reichle et al., 2008; Crow and Ryu, 2009; Chen et al., 2011; Brocca et al., 2012] or the assimilation of real SM measurements, especially remotely sensed products [e.g., Pauwels et al., 2002; Francois et al., 2003; Parajka et al., 2006; Brocca et al., 2010; Wanders et al., 2014; Alvarez-Garreton et al., 2014; Kumar et al., 2014; Lievens et al., 2015; Massari et al., 2015]. However, these studies disagree on the benefits of assimilating SM data for streamflow improvement. Some studies have found improved streamflow simulation with SM assimilation [e.g., Pauwels et al., 2002; Francois et al., 2003; Brocca et al., 2010; Brocca et al., 2012; Wanders et al., 2014; Massari et al., 2015], while others have shown limited improvement or degraded streamflow after SM assimilation, or improved streamflow only under certain conditions. For example, Parajka et al. [2006] found that runoff was not improved, and sometimes even degraded, after satellite SM assimilation; Reichle et al. [2008] found (at best) marginally-improved runoff in a synthetic study; Alvarez-Garreton et al. [2014] found varying levels of improvement when looking at specific flood events; Kumar et al. [2014] found marginal improvements via SM assimilation; Lievens et al. [2015] found improved peak flows but degraded flow at a daily scale; Chen et al. [2011] found moderate streamflow improvement in a synthetic experiment but generally no improvement in a real-data case; and Brocca et al.

[2012] found improved streamflow when assimilating root-zone SM products, but limited success when only assimilating surface SM.

Multiple factors control the magnitude of runoff improvement, including hydrologic model structure, basin characteristics, data assimilation (DA) techniques, error assumptions and the type of SM data product assimilated (see Massari et al. [2015] and Brocca et al. [2010] for useful overviews of these factors). Some past studies have diagnosed performance limitations for their specific DA applications. For example, Chen et al. [2011] argued that the weak vertical coupling in the Soil and Water Assessment Tool (SWAT) limited the success of updating deeper layer SM (and subsequently subsurface flow) in a synthetic twin experiment, and systematic model bias uncorrectable by DA further limited the streamflow improvement in a real data case; Brocca et al. [2012] found much larger streamflow improvement via assimilating root-zone SM product than via assimilating surface SM. Other studies such as Reichle et al. [2008], Alvarez-Garreton et al. [2014] and Massari et al. [2015] discussed the impact of error assumptions, rescaling techniques, study basins, seasonality and hydrologic conditions on streamflow results in their specific cases. All together, these studies pointed out key features that could impact runoff improvement.

Despite this progress, the diagnostic methods used by these past studies are typically only relevant to their specific application, resulting in conclusions that are mostly qualitative. Motivated by a SM DA experiment which yielded no discernible improvement in runoff estimates (see below, especially Figure 2.1), we aim to design a formal diagnostic framework that draws upon synthetic experiments to decompose error terms in streamflow simulation and to provide quantitative understanding of limiting factors in a SM DA system. We formally identify a number of error source terms that occur in the runoff simulation (see Figure 2.2):

1) error in meteorological forcing (primarily precipitation) which is partly reflected in SM states and subsequently in runoff, and partly contributes directly to runoff without going through antecedent states;

2) SM state error caused by an incorrect model representation of SM response to forcing or an incorrect representation of the temporal evolution of SM; and

3) additional runoff error caused by an incorrect model representation of runoff generation processes or streamflow routing.

Built upon this error decomposition, we design a synthetic DA diagnostic framework to address two challenges. First, we distinguish the systematic and random error components (each error term in Figure 2.2 consists of both components), since Kalman-filter-based DA techniques can only correct random error. Second, we further investigate the random error and distinguish the random error associated with SM state variables from other sources of random error. We present a framework that serves as a general diagnostic approach for SM DA problems, which is illustrated by a case study in the Arkansas-Red River basin with a semi-distributed hydrologic model.

Most studies in hydrologic DA focus solely on deterministic analysis results. However, a probabilistic analysis can also be applied to gain additional insight into DA performance. For example, De Lannoy et al. [2006] used several ensemble verification metrics commonly used in meteorology to assess the ensemble representation of SM. Several other DA studies used ensemble verification metrics to verify or calibrate error assumptions [e.g., Pauwels and De Lannoy, 2009; Brocca et al., 2012; Alvarez-Garreton et al., 2014], but typically reported only deterministic results as the final DA performance metric. Throughout this study, we use ensemble metrics not only as a verification and model selection tool, but also to evaluate DA performance from a probabilistic perspective.

The remainder of this paper is organized as follows. Section 2.2 describes our synthetic diagnostic framework as well as the study domain, hydrologic model, and data used for the case study. Section 2.3 presents results from our real-data and synthetic experiments and illustrates the usage of the latter to help diagnose the former. Section 2.4 contains our conclusions.

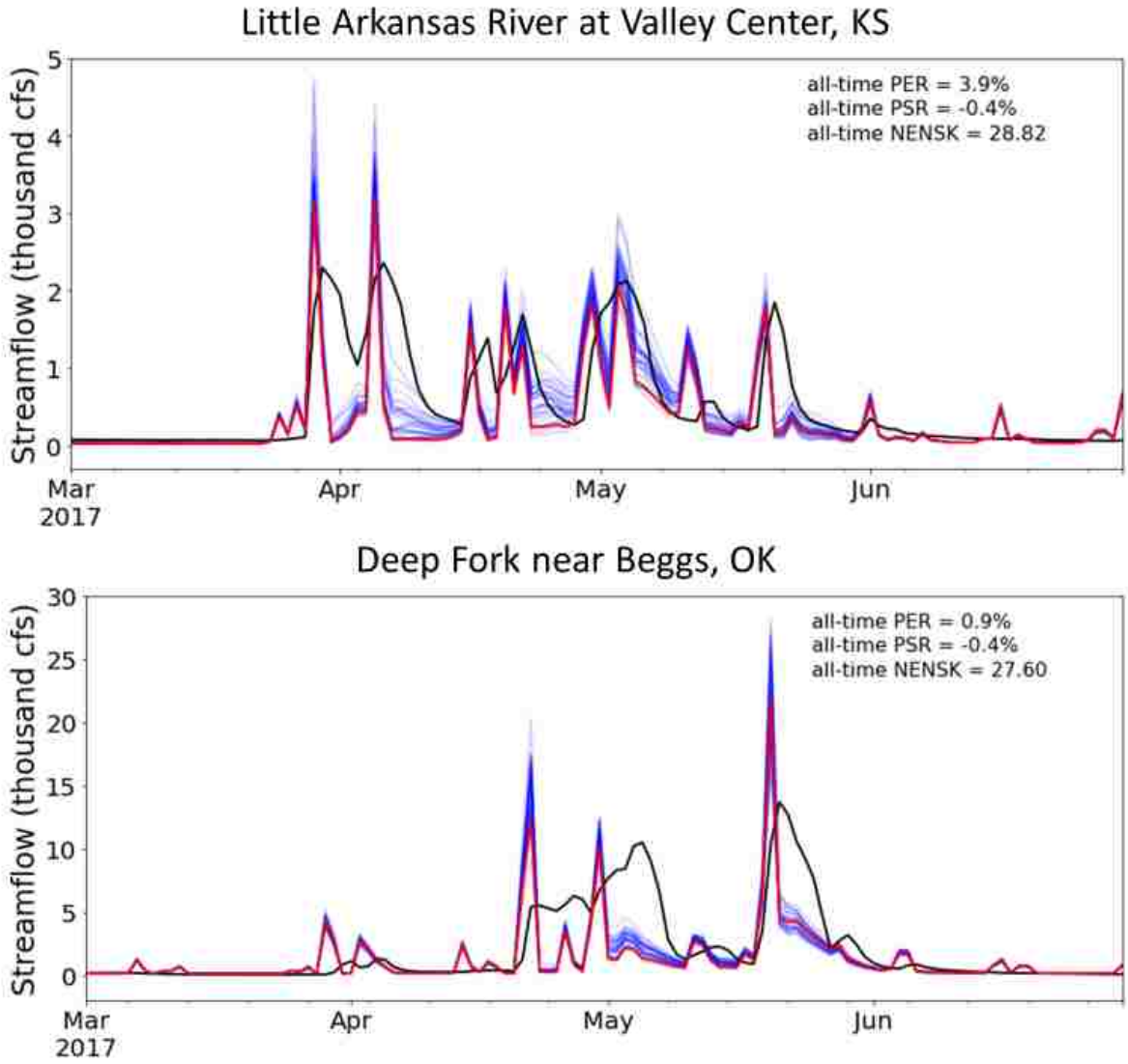

Figure 2.1. Streamflow time series at two example USGS gauges. *Black line*: USGS observed streamflow; *red line*: open-loop simulated streamflow; *blue lines*: DA-updated ensemble streamflow. PER, PSR and NENSK metrics based on the entire simulation period (March 2015 - December 2017) are labeled in each subplot (see Section 2.2.6 and Supplemental Material Section A2 for a full description of evaluation metrics).

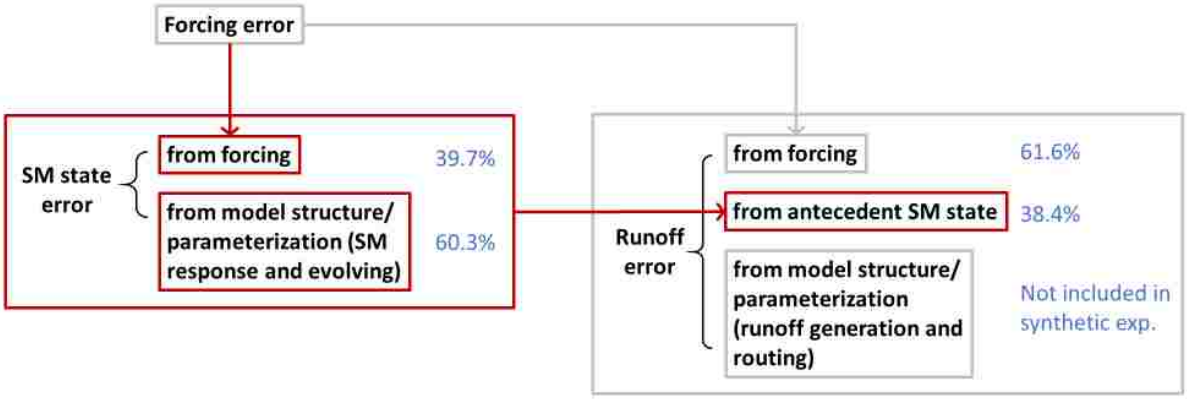

Figure 2.2. Illustration of error decomposition in a streamflow simulation system (see Section 2.1 in the text for detailed explanation of each term). All error terms in the diagram consist of a systematic error component and a random error component. The red highlighted terms are the parts correctable by SM DA. The blue numbers/text describe random error decomposition results from the Arkansas-Red case study.

2.2 METHODS AND DATA

Sections 2.2.1, 2.2.2 and 2.2.3 describe the study domain, hydrologic modeling approach and data used in the case study, respectively, for both the real-data and synthetic experiments. DA techniques and our framework design are then described in Sections 2.2.4 and 2.2.5, respectively. Finally, evaluation metrics are presented in Section 2.2.6.

2.2.1 Case study domain

The Arkansas-Red River basin is located in the south-central United States and covers an area of approximately 605,000 km^2 (Figure 2.3). The basin consists of two major sub-basins, the Arkansas River and the Red River, and represents a major tributary of the Mississippi River. Within the basin, there is a clear east-west climatic gradient (wetter in the east and drier in the west). Most of the basin experiences little snow cover in winter except for mountainous areas on its far western edge. Vegetation cover varies from deciduous forest in the east to wooded grassland, shrubs, crops and grassland in the west.

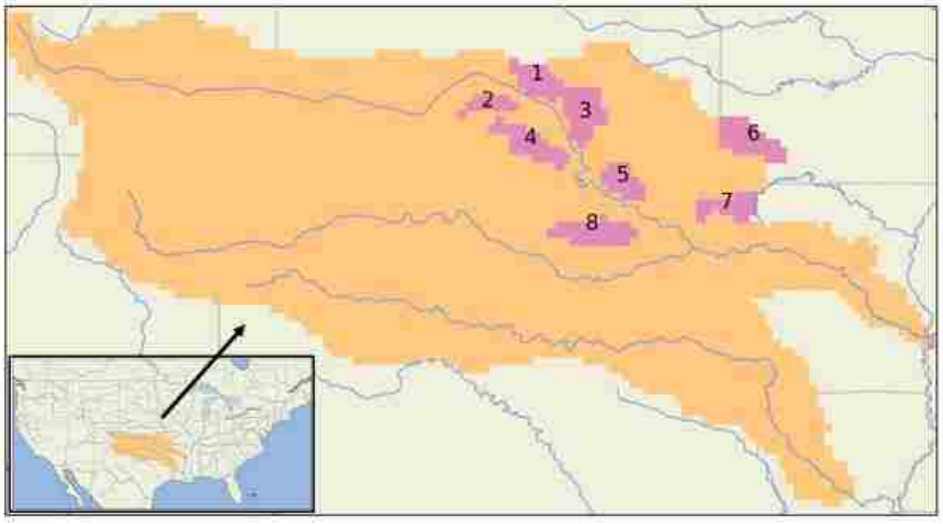

Figure 2.3. Illustration of error decomposition in a streamflow simulation system (see Section 2.1 in the text for detailed

2.2.2 Case study hydrologic modeling and meteorological forcing data

a) Hydrologic model description

Hydrologic model simulations are based on version 5 of the Variable Infiltration Capacity (VIC) model [Liang et al., 1994; Hamman et al., 2018]. The VIC model is a large-scale, semi-distributed model that simulates various land surface processes including both water balance and energy balance components. The VIC model is typically implemented on a grid cell basis, with each grid cell further divided into tiles via statistical distributions to represent sub-grid heterogeneity in elevation and vegetation cover. Particularly relevant to SM DA for runoff improvement, the soil column in each grid cell tile is discretized into L vertical layers, with water within each layer draining into the lower layer via a non-linear gravity-driven drainage curve [Brooks and Corey, 1964]. The sum of water contained in the top (L - 1) soil layers determines the partitioning of rainfall or snowmelt into infiltration and overland runoff (i.e., fast-response runoff generated over saturated soil) by the variable infiltration capacity curve [Zhao et al., 1980; Wood et al., 1992]. Slow-response runoff generated as drainage from the deep soil is determined by the moisture level in the bottom layer via a non-linear recession curve [Todini, 1996; Nijssen et al., 2001].

In this study, the Arkansas-Red River basin was modeled at a spatial resolution of 1/8^o (3999 grid cells in total) with each grid cell further divided into multiple vegetation tiles. We did not consider heterogeneity in sub-grid elevation. All model parameters regarding vegetation cover and soil properties were taken from Maurer et al. [2002] and calibrated for large-basin streamflow. The soil column in each grid cell was divided into 3 vertical layers, with domain-average thicknesses of 0.10 m, 0.40 m and 0.93 m, respectively. The VIC model was run at a 3-hourly timestep in water balance mode.

b) Meteorological forcing data

The real-data experiment was conducted for March 2015 – December 2017, a period when NASA Soil Moisture Active/Passive (SMAP) [Entekhabi et al., 2010] SM retrieval data was available. The North American Land Data Assimilation System Phase 2 (NLDAS-2) meteorological data [Xia et al., 2009] was used to force the simulations. This dataset was chosen because of its temporal coverage, temporal resolution, and data quality. Hourly NLDAS-2 meteorological variables required by VIC were aggregated to a 3-hourly timestep. NLDAS-2 forcing data from 1979-2015 were used to spin up the VIC model.

Since the SMAP retrievals are only available after mid-2015, we selected a longer ten-year period (1980 – 1989) for the synthetic experiments in order to more robustly sample performance evaluation metrics. We used daily precipitation and maximum and minimum temperatures from one of the realizations of a 1/8^o ensemble meteorology product [Newman et al., 2015]. This forcing dataset was used because it considered observation uncertainty, making it suitable for DA applications. Daily wind speed for this period was taken from a 1/8^o gridded meteorological data product by Maurer et al. [2002]. Other meteorological forcing data required by VIC, which includes shortwave and longwave radiation, air pressure and humidity, were calculated using the MTCLIM algorithms and the Tennessee Valley Authority algorithm, and then disaggregated to a 3-hourly timestep following Bohn et al. [2013]. Prior to the synthetic simulation period, VIC was spun-up from 1949-1979 using the Maurer et al. [2002] forcing data.

c) Streamflow routing

For the real-data experiments, we routed grid cell runoff through the stream channel using the RVIC routing model [Hamman et al., 2017], which is a source-to-sink model based on the routing model typically used as a post-processor for VIC model runoff simulations [Lohmann

et al., 1996; 1998]. In the synthetic experiments, locally-generated runoff was evaluated without routing into stream channels.

2.2.3 SMAP retrievals and USGS streamflow data

The SMAP mission provides SM estimates for the top 5 centimeters of the soil column with a revisit time of 2-3 days, a resolution of 36 km, and a 50-hour data latency. Here, SMAP L3 Passive retrievals [O'Neill et al., 2016] were assimilated during the non-winter seasons (March to October). A few SMAP pixels with near-constant retrieval values, indicating an obvious quality flaw, were manually masked out. Daily streamflow data acquired at 8 USGS streamflow sites [USGS, 2018] were used for evaluation in the real-data experiment (see Figure 2.3 and Table 2.1 for site locations). These 8 sites were selected due to their lack of human streamflow regulation and good rain gauge coverage (see Crow et al. [2017] for details).

Table 2.1. List of basins evaluated in the study with their corresponding USGS gaged streamflow site information.

Basin number	USGS station no.	USGS station name
1	07144200	Little Arkansas River at Valley Center, KS
2	07144780	Ninnescah River AB Cheney Re, KS
3	07147800	Walnut River at Winfield, KS
4	07152000	Chikaskia River near Blackwell, OK
5	07177500	Bird Creek Near Sperry, OK
6	07186000	Spring River near Wace, MO
7	07196500	Illinois River near Tahlequah, OK
8	07243500	Deep Fork near Beggs, OK

2.2.4 Ensemble Kalman filter (EnKF)

Our implementation of the standard Ensemble Kalman Filter (EnKF) is described in the Supplementary Materials (Section A1). In the following, we describe implementation details specific to our application.

a) State vector, observation operator and ensemble size

Although SMAP only provides surface measurements, deeper layer SM updates are essential for successful runoff improvement because of their control on runoff generation processes in VIC (discussed further in Section 2.3.2). However, in the Lievens et al. [2015, 2016] SM DA studies using VIC, the authors did not include the bottom-layer state in the Kalman filter to avoid unstable behavior. We will further examine this decision via the synthetic experiments in Section 2.3.3.a. However, unless otherwise noted, all DA results are based on the Lievens et al. [2015, 2016] approach whereby only the first and second soil layers are updated. Furthermore, temperature states were not updated in this study.

Following Gruber et al. [2015], each pixel of observations was assimilated separately (i.e., 1D filtering) without consideration of spatially correlated errors. Gruber et al. [2015] found only limited performance enhancement when transitioning from a 1D to a 2D filter for the assimilation of remotely sensed SM. For the synthetic experiments, synthetic surface SM measurements were generated at the same spatial resolution as the VIC model setup and assimilated into each 1/8° VIC grid cell separately. In such a 1D filter, the measurement vector $\tilde{y}$ (Equation (A2) in Supplemental Material) reduces to a scalar:

$$\tilde{y} = SM_1^{obs} \tag{2.1}$$

where SM_1^{obs} is the synthetic observation of surface SM at 1/8° resolution. The state vector (x in Equations (A1) and (A2) in Supplemental Material) is simply a collection of SM states for the top two layers in all vegetation tiles within a grid cell. The observation operator (H in Equation (A2) in Supplemental Material) calculates the tile-averaged first-layer SM state as a map from the state space to the observation space.

For the real-data experiment, SMAP retrievals were acquired at a coarser spatial resolution (36-km) than the VIC grid (1/8°), and the grid edges do not align. To resolve this spatial mismatch, every VIC grid cell is assigned to the nearest SMAP pixel. Then for each SMAP pixel, SMAP was assimilated to update all corresponding VIC grid cells via EnKF by augmenting the states from all those grid cells in the state vector. In this case, the measurement vector $\tilde{y}$ is still a scalar as in Equation (2.1), and the observation operator H calculates the areal-averaged first-layer SM state from the multiple VIC grid cells. Unless otherwise stated, a 32-member Monte Carlo ensemble was used to generate the EnKF.

Although potential bias in SM states as well as in runoff can, in theory, develop in the EnKF result due to the interaction of zero-mean noise with model nonlinearities, we did not

implement bias correction schemes during EnKF (as in for example Ryu et al. [2009]) since including such a scheme did not improve filter performance in our case (results not shown). This decision does not significantly degrade our EnKF result because, as applied here, the random precipitation perturbations are relatively small and thus produce a correspondingly small EnKF propagation bias.

b) Error assumptions

Measurement errors were estimated to be in the range of 0.03 to 0.045 m^3/m^3 for the SMAP L3 retrievals, which is roughly consistent with those reported by SMAP validation studies [e.g., Colliander et al., 2017; Chan et al., 2017]. For each SMAP pixel (or the synthetic $1/8^o$ measurements in the synthetic case), we linearly mapped its leaf area index (LAI) to the above error range, with larger error values corresponding to denser vegetation cover. A positive relationship between vegetation density and satellite-observed moisture error has been well documented [e.g., Crow et al., 2010]. Generally, eastern portions of the Arkansas-Red River basin have relatively larger measurement errors than western areas due to their denser vegetation cover.

The EnKF method requires the generation of an ensemble to capture the error in model propagation. We perturbed precipitation forcing at a 3-hourly timestep with multiplicative random error. The multiplier was drawn from a log-normal distribution with mean one and a dimensionless standard deviation of 0.3, and high temporal autocorrelation (with the underlying normal random variable generated by a first-order autoregressive process with coefficient 0.9 [-]). The log-normal precipitation error form has been used frequently [e.g., Ciach and Krajewski, 1999; Nijssen and Lettenmaier, 2004; Chen et al., 2014]. The standard deviation used here is smaller than used in some other studies (e.g., Chen et al. [2014] used a standard deviation of 1 [-]) to represent the good-quality, gauge-informed NLDAS-2 forcing dataset. Direct state perturbation was applied to all SM layers and kept constant over space without spatial or temporal autocorrelation (normally distributed with zero mean and a standard deviation of 0.5 mm for all three layers). This state perturbation was tuned in the real-data case to ensure that the overall normalized variance of the filter innovation is approximately one, which, assuming the correct specification of measurement error, is a necessary condition for proper error levels in a filter (see Mehra [1971] and Crow and Bolten [2007]). The state perturbation was assumed to be

highly correlated between vegetation tiles and vertical layers (correlation coefficient = 0.9 [-]). These same model error statistics were also used for the synthetic experiments.

c) SMAP rescaling

The SMAP SM measurements were rescaled prior to their assimilation in the real data experiment. Specifically, we rescaled SMAP L3 to match the seasonal (31-day window) mean and long-term (the 3-year simulation period) standard deviation of the VIC-simulated surface-layer SM time series. Such moment-matching rescaling is one of the standard techniques in DA [e.g., Chen et al., 2011; Brocca et al., 2012; Alvarez-Garreton et al., 2013]. The rescaling was applied to each SMAP pixel separately and to the ascending and descending retrievals separately. The same standard deviation ratio used for SM rescaling was also applied to rescale SMAP retrieval errors into the VIC model regime.

2.2.5 *Experimental design*

In this section we describe the real-data experiment followed by a set of synthetic experiments conducted to diagnose notable results from the real-data experiment. As motivated in the introduction, the synthetic experiments quantify the random error components in a simulation system only since these are the only components potentially correctable by the EnKF.

a) Real-data experiment

A so-called "open-loop" run serves as the baseline for modeling performance without any DA. Here the open-loop run was defined as a single, unperturbed VIC simulation. Deterministic modeling improvement using DA will be evaluated by comparing to the open-loop run.

To evaluate DA performance probabilistically, we generated an ensemble of perturbed (but not updated) VIC simulations to serve as the baseline (hereafter called the "open-loop ensemble"). The error statistics of the perturbations added to the forcings and states in the open-loop ensemble were the same as those applied in the DA procedure (see Section 2.2.4). The size of the open-loop ensemble was also kept the same as the DA ensemble (i.e., 32 members unless otherwise noted).

The DA setup described in Section 2.2.4 was applied to assimilate SMAP retrievals into VIC. The final DA setup for the real-data experiment was informed by findings from the synthetic twin experiment (see Section 2.3.3). The runoff values after DA were routed to the

gauge locations and compared with streamflow observations. Streamflow improvement after DA was evaluated both deterministically and probabilistically.

b) Synthetic experiments

1) Open-loop, synthetic truth, and synthetic measurement generation

In the synthetic experiments, one single perturbed VIC realization was chosen as the synthetic "truth" (following Crow and Reichle [2008] and Kumar et al. [2009]). Synthetic measurements were then generated by degrading the true first-layer SM by adding random observation errors (error statistics consistent with those described in Section 2.2.4). Synthetic measurements were generated at a 24-hour temporal interval which roughly matches the average length of SMAP data gaps. The synthetic open-loop runs were produced as in the real-data case described above.

2) Synthetic perfect-state/forcing experiments

We conducted two sets of model experiments to attribute the total random runoff error to individual error sources.

i) In the "perfect-state" analysis, we substituted the true antecedent states (including all-layer SM states, snow states and all other initial conditions) into the VIC model at a 24-hour interval and evaluated the error reduction of subsequent runoff simulations (relative to the open-loop runoff). Any such error reduction is attributable to the improved specification of antecedent states, where the uncorrected portion of runoff error can be attributed directly to rainfall error. The former is potentially correctable by SM updating (highlighted in red in Figure 2.2), while the latter is not (the grey parts of Figure 2.2). To further decompose the error source from each individual SM layer, we conducted three additional perfect-single-state runs, where we substituted the true *i*th-layer SM (SM*i*) (where $i = 1$, 2 and 3) states into the VIC model while keeping all the other antecedent states the same as in the open-loop baseline. We performed this experiment separately for each of the three soil layers.

ii) In the "perfect-forcing" analysis, we used the true precipitation forcing as input to the VIC model without inserting the true model states and evaluated the resulting error reduction in SM states. This error reduction in SM quantifies the part of the forcing error reflected in SM states and therefore potentially correctable by SM updating (Figure 2.2). SM error reduction was also calculated for each layer separately for additional interpretation.

3) Synthetic identical twin experiment

In addition, a standard synthetic identical twin experiment was conducted to quantify the overall effectiveness of SM DA to correct random error components linked directly to antecedent SM states. Specifically, we assimilated synthetic surface SM retrievals into VIC and evaluated the subsequent improvement in SM states and runoff fluxes relative to the open-loop baseline. The effect of some EnKF methodological choices will be investigated in more detail in Section 2.3.3.a.

2.2.6 *Evaluation metrics*

We evaluated the improvement of SM states and local runoff fluxes in the synthetic experiment, and the improvement of routed streamflow in the real-data experiment. Runoff was aggregated to daily values after which we applied a logarithmic transformation:

$$logRunoff = \log(runoff + 1) \tag{2.2}$$

The logarithmic transformation reduces the dominant influence of a few large flood events on the evaluation metric and the "+ 1" in the transformation has the effect that zero runoff values remain zero after transformation. SM state was not log-transformed.

We used percent root-mean-squared error reduction (PER) for deterministic evaluation. In addition, the Kling-Gupta efficiency (KGE) [Gupta et al. 2009] was calculated for streamflow to quantify the baseline performance. KGE describes overall performance by integrating the evaluation of correlation, variance, and bias, and ranges from negative infinity to 1 with values closer to 1 indicating better performance.

In addition, two probabilistic metrics were used: percent continuous rank probability score reduction (PSR) and normalized ensemble skill (NENSK). PSR measures the reduction of continuous rank probability score (CRPS) which, in turn, measures the deviation of the cumulative distribution function (CDF) of an ensemble from that of a reference (i.e., observation in the real-data case or truth in the synthetic case) [Hersbach, 2000]. NENSK measures the ensemble-mean error normalized by ensemble spread:

$$NENSK = \frac{ENSK}{ENSP} \tag{2.3}$$

where the ensemble skill (ENSK) is the temporal mean of ensemble-mean squared error, and the ensemble spread (ENSP) is the temporal mean of ensemble variance [De Lannoy et al., 2006;

Brocca et al., 2012; Alvarez-Garreton et al., 2014]. In an ideal situation where an ensemble is a correct representation of analysis uncertainty, NENSK should be one [Talagrand et al., 1997; Wilks, 2011]. NENSK > 1 indicates an under-dispersed ensemble while NENSK < 1 indicates an over-dispersed ensemble. Note that NENSK is calculated based on the updated ensemble alone and is not compared to the open-loop baseline. See Supplemental Material (Section A2) for mathematical details concerning the above metrics.

2.3 ARKANSAS-RED CASE STUDY RESULTS AND DISCUSSION

DA results for the real SMAP-assimilated case study will be presented in this section, followed by the diagnostic results from the synthetic framework. Note that we present the real-data results first before the synthetic results (even if the former uses methodological choices informed by the latter) since we aim to use the synthetic framework as a tool to diagnose the streamflow improvement we see in the real-data case. Although the diagnostic conclusions we draw are case-specific, we use this case study as a demonstration of the usage of our general diagnostic framework.

2.3.1 Open-loop streamflow simulation and SMAP assimilation performance

To begin, we present the NLDAS-2 forced, open-loop baseline streamflow simulation evaluated against USGS observations at the 8 sub-basins. This open-loop streamflow performance is compared with the corresponding open-loop runoff performance from the synthetic case. Because synthetic runoff is evaluated at each grid cell, sub-basin median statistics are reported. Since the synthetic and the real-data experiments were conducted for different time periods, the comparison is primarily qualitative. The daily open-loop KGE results are shown in Table 2.2. For all sites, the real-data open-loop case performs comparably to or worse than the corresponding synthetic case. This discrepancy reflects the existence of systematic error in the real-data streamflow simulation, since the synthetic case is designed to only capture the random error components (see Introduction and Section 2.2.5). The systematic error is also visible from the open-loop time series at two example sites in Figure 2.1 at Little Arkansas River at Valley

Center, KS, and Deep Fork near Beggs, OK. For example, the red lines in both show simulated flow recessing too rapidly.

The real-data DA results assimilated by SMAP are presented in Table 2.2 (the "real-data" columns). Note that this experiment was conducted using a DA setup that was informed by our synthetic twin experiment, with 32-member EnKF and no SM3 Kalman update (see Section 2.3.3.a below for a complete discussion of these synthetic experiments). Overall the streamflow correction is small, with PER ranging from -2.0% to 7.0% and PSR close to zero across sub-basins. The example time series in Figure 2.1 illustrate that DA often nudges the simulated flow in the correct direction, but is unable to substantially correct the inaccurate flow behavior. Since the streamflow ensemble produced by DA only accounts for random error, the existence of the systematic error partly results in an ensemble with too little dispersion and consequently a large NENSK (>20 at all sites).

The systematic error we observe here is partly due to model parameterization error. Although the baseline VIC simulation could be improved by parameter calibration for each sub-basin, we chose not to perform this additional calibration for the following reasons: 1) Maurer et al. [2002] calibrated the current VIC parameter set toward large-basin streamflow and this parameter set has been used widely in large-scale VIC applications; 2) although it is possible to calibrate individual unregulated small basins based on streamflow observations, it is challenging to generalize such a calibration scheme to a large domain due to flow regulation, low-quality meteorological forcing data and lack of streamflow observations. Therefore, we chose not to calibrate and to maintain the level of parameterization error we see here, so that the evaluation and investigation in this section is more relevant to large-scale DA applications where similar errors are likely to occur.

Table 2.2. Comparison of runoff metrics in the real-data and the synthetic experiments. NENSK is only shown for the real-data case. Baseflow index (calculated as the fraction of slow-response runoff in total runoff from the synthetic open-loop simulation) of each sub-basin is also shown.

Basin number	Baseflow index	Open-loop KGE		PER				PSR				NENSK
		synthetic	real-data	synthetic			real-data	synthetic			real-data	real-data
				total runoff	fast-response	slow-response		total runoff	fast-response	slow-response		
1	0.28	0.72	0.44	50.7%	3.2%	71.7%	3.9%	8.4%	2.9%	24.6%	-0.4%	28.82
2	0.32	0.86	0.45	8.6%	3.7%	39.5%	-2.0%	8.3%	4.6%	18.8%	0.1%	46.52
3	0.45	0.70	0.30	15.9%	2.5%	20.1%	7.0%	13.0%	2.0%	20.2%	-0.6%	73.91
4	0.31	0.88	0.22	10.2%	4.0%	22.9%	2.7%	9.0%	3.2%	19.8%	-0.2%	114.23
5	0.55	0.83	0.68	11.5%	0.9%	17.9%	1.7%	12.5%	2.0%	20.0%	-0.2%	17.55
6	0.64	0.80	0.29	17.3%	1.3%	19.4%	5.8%	11.6%	1.7%	15.0%	2.2%	33.79
7	0.63	0.83	0.80	16.4%	0.1%	19.2%	3.0%	7.1%	1.0%	9.2%	0.1%	30.20
8	0.55	0.84	0.78	9.3%	0.0%	13.5%	0.9%	8.7%	1.8%	13.9%	-0.4%	27.60

2.3.2 Decomposition of random runoff error via synthetic perfect-state/forcing analysis

As a first step to diagnose the above limited streamflow improvement via SM DA, in this section we decompose the random error components in the runoff simulations, identify their sources, and discuss the theoretical limitation of the error reduction using a set of perfect-state/forcing synthetic experiments.

a) Perfect-state analysis

Results of the perfect-state experiments are shown in Figure 2.4. PER was calculated based on the runoff fluxes immediately after a state substitution and therefore reflects the runoff response to state substitution. Bluer shading in the figure indicates greater runoff improvement in response to the substitution of perfect antecedent state(s), thus indicating that more of the runoff error originates from the error in antecedent state(s).

The first row in Figure 2.4 shows the runoff PER by substituting the true antecedent SM states for all layers (and all other VIC model states). In addition to total runoff, fast-response and slow-response runoff are analyzed separately. Total runoff displays a domain-median PER of 38.4% relative to the open-loop baseline. The runoff error reduction primarily relies on slow-response runoff, whose error is almost eliminated (99.5% domain-median PER). Conversely, fast-response runoff is only modestly improved (13.4% domain-median PER) even if all antecedent VIC states are known perfectly.

This result suggests that only 38.4% of the total runoff error originates from antecedent states, while the remaining 61.6% is derived directly from precipitation forcing error (see "runoff error" box in the lower right side of Figure 2.2; the case-study error decomposition results are marked in blue on the diagram). Therefore, this error decomposition imposes a first-order upper limit for runoff improvement via SM updating. Further reduction of the runoff error that originates directly from the forcing requires additional techniques such as rainfall correction [e.g., Crow et al., 2011]. This upper limit is especially restrictive for fast-response runoff since 86.6% of its error is attributed directly to within-storm rainfall forcing.

We further evaluate the drastically different error decomposition of fast- versus slow-response runoff by calculating the runoff PER from single-layer true state substitutions (rows 2 – 4 in Figure 2.4). The (small) PER of fast-response runoff mainly comes from SM2. The lack of control of SM1 on fast-response runoff error reduction is due to the VIC model structure as well

as its parameterization: VIC combines the moisture amount in the top two layers to determine fast-response runoff. Since the second layer is typically much thicker than the first layer, it is dominant in controlling the overland runoff/infiltration partitioning process. In contrast, slow-response runoff error originates almost entirely from the SM3 state alone, since VIC simulates slow-response runoff as a direct function of the SM3 state. The difference between fast- and slow-response runoff can also explain the spatial gradient in total runoff PER. The wetter and more vegetated areas in the eastern part of the basin are more dominated by subsurface flow, and thus runoff is more strongly linked to states. In contrast, drier areas in the western part of the domain rely more on surface-excess, fast-response runoff, and thus its runoff error primarily originates directly from forcing.

b) Perfect-forcing analysis

Figure 2.5 shows PER in SM when VIC is forced by the synthetic true precipitation. Domain-median PER of total SM (sum of all three layers) is reduced by 39.7%, which represents the random SM error that originates from uncertain forcing. The remaining 60.3% of SM error results from errors in the model representation of SM-related processes itself (see the lower left box in Figure 2.2). In contrast to the forcing error directly contributing to runoff as quantified above in Section 2.3.2.a, SM error identified here that arises from forcing error is potentially correctable by SM updating. We also note that surface-layer PER in response to perfect forcing is smaller than PER in deeper layers (Figure 2.5). This is likely due to the longer memory of deeper layer SM, whose error, over time, reflects the accumulated error in precipitation forcing.

To summarize the perfect-state/forcing analysis, we find that, in our case study, approximately 60% of the random error in runoff directly originates from precipitation forcing and is thus uncorrectable by SM updating. In particular, fast-response runoff error is largely attributable to direct forcing error and therefore cannot be substantially corrected by SM DA, while slow-response runoff is strongly state-controlled and its error reduction relies largely on SM3 correction. The part of the precipitation error that is reflected in SM states is also quantified and is potentially correctable via SM DA. These case-specific conclusions serve to illustrate the usage of the general framework to decompose error sources and quantify theoretical limitations in the application of SM DA to enhance streamflow simulation.

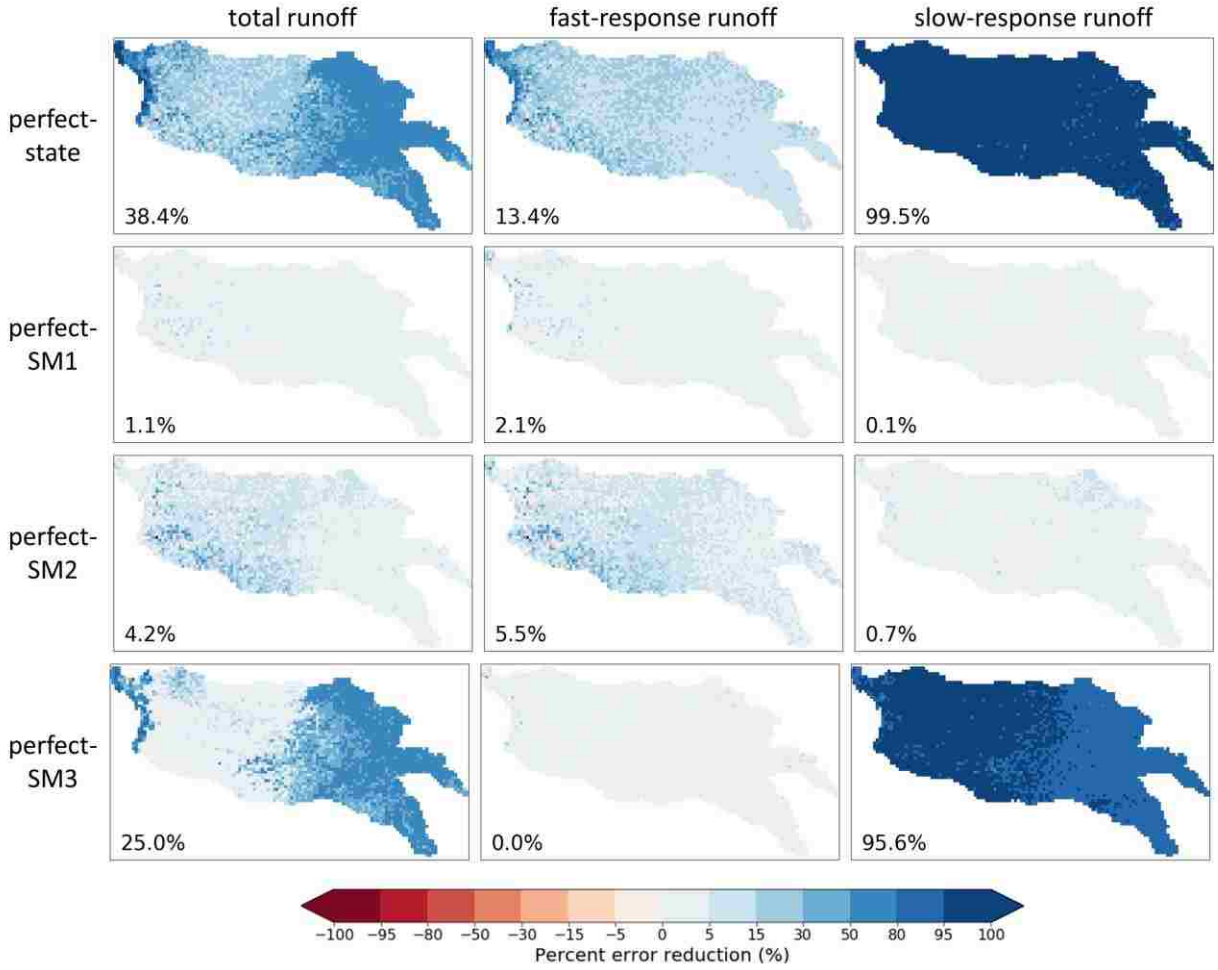

Figure 2.4. Maps of percent error reduction (PER) of runoff for synthetic perfect-state analysis. The first row shows the PER when we substitute true antecedent states of all VIC states. The second through fourth rows show the PER when we only substitute the true SM1 state, SM2 state and SM3 state only, respectively. The columns show PER of total runoff fluxes, fast-response runoff and slow-response runoff, respectively. The number on each panel is the domain-median PER value. Blue shading indicates improved runoff accuracy and red shading indicates degraded runoff accuracy.

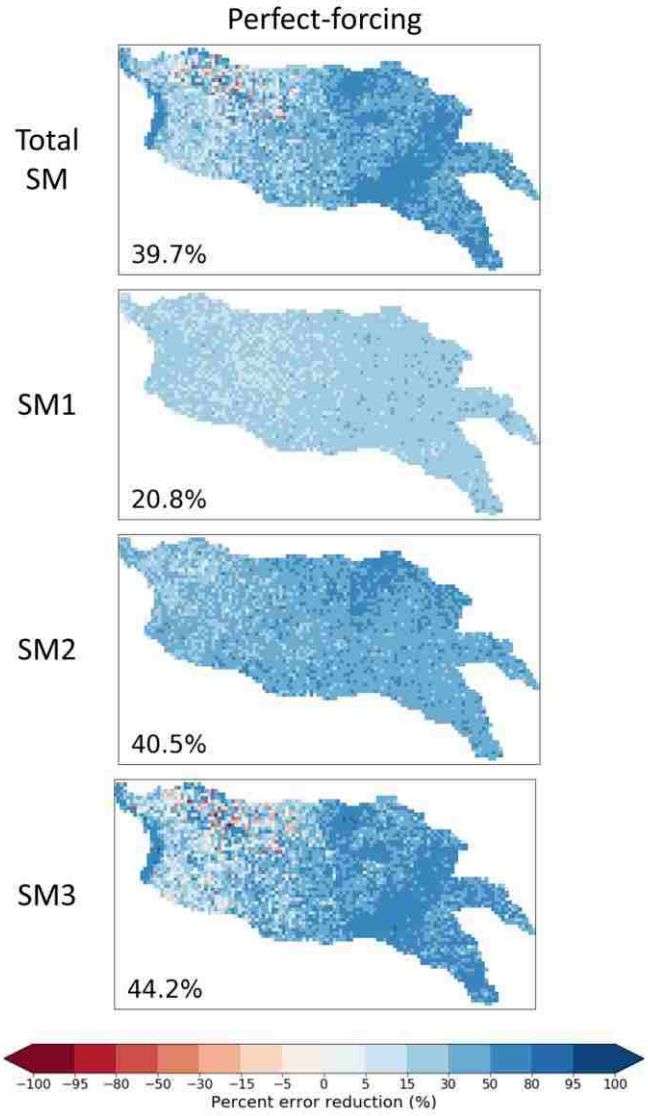

Figure 2.5. Maps of percent error reduction (PER) of SM for the synthetic perfect-state analysis. Each panel shows PER of the total SM (sum of all three layers), SM1, SM2 and SM3, respectively. The number on each panel is the domain-median PER value. Blue shading indicates improved SM accuracy and red shading indicates degraded SM accuracy.

2.3.3 Effectiveness of random error reduction via identical twin experiments and diagnosis of real-data performance

The perfect-state/forcing analysis presented above demonstrates that, in our case study, constraining the VIC SM3 state is critical for reducing random error in runoff via SM DA. In this section, we will first examine factors determining SM3 DA performance in a smaller subdomain via a synthetic identical twin experiment. Then, we will present the synthetic DA performance for the entire domain.

a) Small-domain investigation on SM3 DA performance

We arbitrarily chose a smaller domain to further investigate DA performance (a box around the Little Arkansas sub-basin, see Figure 2.3, sub-basin 1). As mentioned in Section 2.2.4, we examine the effect of including SM3 in the EnKF state vector, which was excluded by Lievens et al. [2015, 2016]. According to Kalman filter theory, we should be able to include any state variable in the state vector, even if it is only weakly correlated or uncorrelated with the available measurements. The EnKF algorithm should be able to correctly determine the level of correlation and use that information in the subsequent update. However, in practice SM3 ensembles tend to be under-dispersed when we update SM3 as part of our EnKF (first row in Figure 2.6 which shows runoff at a single grid cell with NENSK of the updated ensemble greater than 1). Possible reasons for this include: 1) a finite ensemble size, which can result in an overestimate of the error correlation between the surface measurement and SM3, leading to an overconfident update; 2) nonlinearity of the model, which can result in a non-optimal update after propagation. These problems are analogous to three-dimensional atmospheric DA where state vectors are commonly "localized" to exclude states geographically far from observations [e.g., Anderson, 2007]. In our context, excluding SM3 from EnKF essentially applies "vertical localization" since we assume the SMAP surface measurements contain little information about SM3.

To examine this issue, we conducted a set of synthetic DA analyses in the small sub-basin altering: 1) ensemble size and 2) the decision of whether or not to apply "vertical localization", i.e., not including SM3 states in the Kalman state vector. For all these experiments, the assimilated synthetic measurements and the truth were kept unchanged as described in Section 2.2.5. In the synthetic truth, a weak vertical error correlation exists across layers

(because of the way we generated the open-loop baseline and the synthetic truth), but the goal is to examine how much benefit there is in updating SM3.

Figure 2.7 to Figure 2.9 show maps of the DA performance for three different DA setups for the small domain. The first row (default) includes the SM3 state update in EnKF, and the second row excludes SM3 Kalman update. Figure 2.6 shows example time series at one grid cell for the same experiments. In the following we discuss the effect of each DA variation in detail:

1) The effect of ensemble size

Comparing the columns in Figure 2.6 to Figure 2.9, a larger ensemble size clearly benefits the SM3 update both deterministically and probabilistically. Specifically, when using the default DA setup, substantial improvement in both PER and PSR is seen when the ensemble size increases from $N = 16$ to $N = 64$. NENSK also steadily drops closer to one with larger ensemble sizes (row 1 in Figure 2.9). Interestingly, the benefit of a larger ensemble size is much smaller when we do not update SM3 in DA (row 2 in Figure 2.6 to Figure 2.9). We will further discuss the reason for this below.

2) The effect of vertical localization (i.e., exclude SM3 update in EnKF)

Excluding SM3 from Kalman update generally results in better SM3 estimates (comparing row 2 with row 1 in Figure 2.6 to Figure 2.9). This confirms our hypothesis that SM3, which is only weakly correlated with the surface measurement, is prone to overconfident Kalman updates, because the finite ensemble tends to overestimate the vertical correlation. The benefit of vertical localization is less significant for larger ensemble size, which again confirms that it is mainly the finite ensemble size that causes the overconfident SM3 update. Note that even without direct Kalman updating, SM3 is still moderately improved by approximately 20% in terms of PER and PSR due to its positive response to updating SM1 and SM2.

In terms of PER and PSR, vertical localization still enhances SM3 performance when N grows to 32 but shows no extra benefit beyond that. However, for NENSK, SM3 localization continues to benefit when we increase N from 32 to 64 (Figure 2.9). In other words, the ensemble representation of SM3, quantified by NENSK, still suffers slightly from overconfident updates with $N = 64$. This conclusion is based on our ensemble evaluation and cannot be determined solely from a deterministic evaluation.

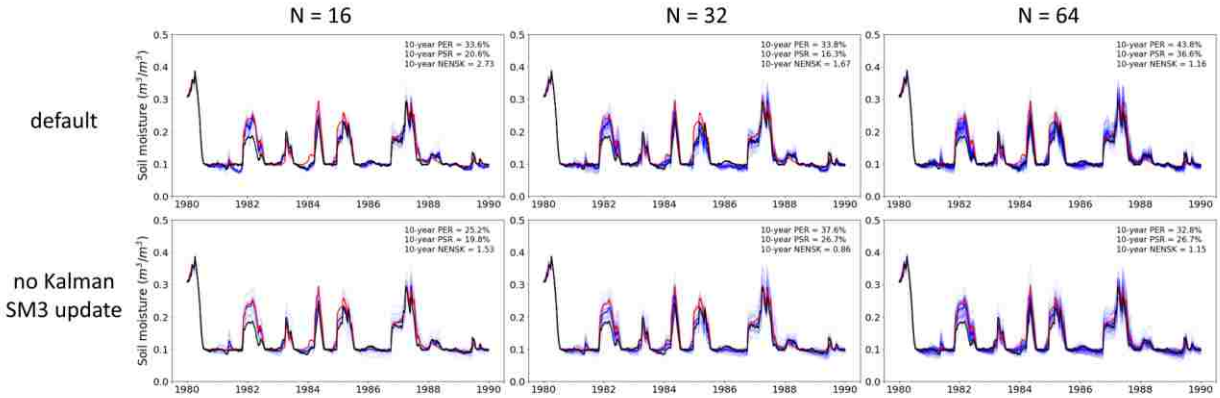

Figure 2.6. Example SM3 time series at one grid cell for the small-domain synthetic DA experiments (37.9375^o N, 97.9375^o W, which is located in the southwest part of the small domain). *Black lines*: synthetic truth; *red lines*: the open-loop analysis; *light blue lines*: the updated ensemble; *blue lines*: updated ensemble-mean. The 10-year PER, PSR and NENSK statistics of the updated ensemble are labeled in each subplot. The columns show ensemble sizes N = 16, 32 and 64, respectively; the rows show the default DA setup and no-Kalman-SM3-update DA run, respectively.

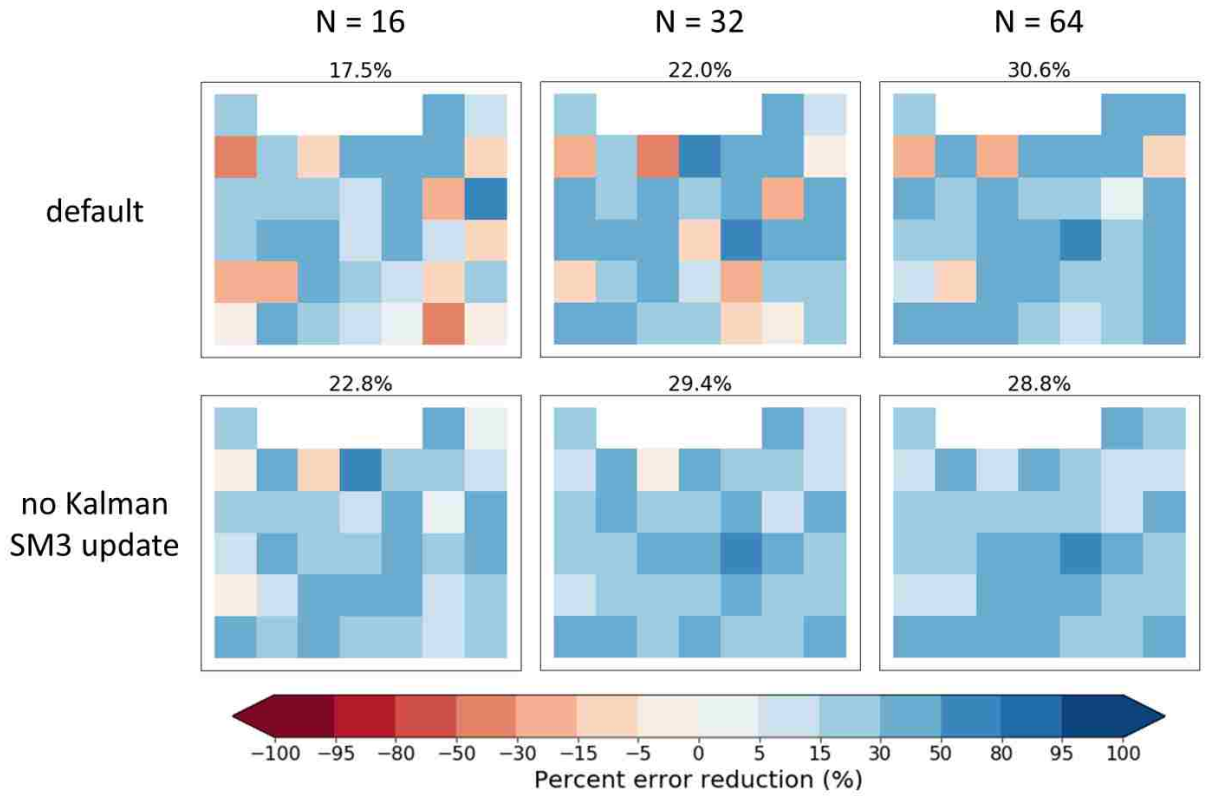

Figure 2.7. Maps of percent error reduction (PER) of SM3 for the small-domain synthetic DA experiments. The number on top of each subplot is the domain-median PER value. The subplot panel layout is the same as in Figure 2.6. Blue shading refers to improved SM3 accuracy and red color shading refers to degraded SM3 accuracy.

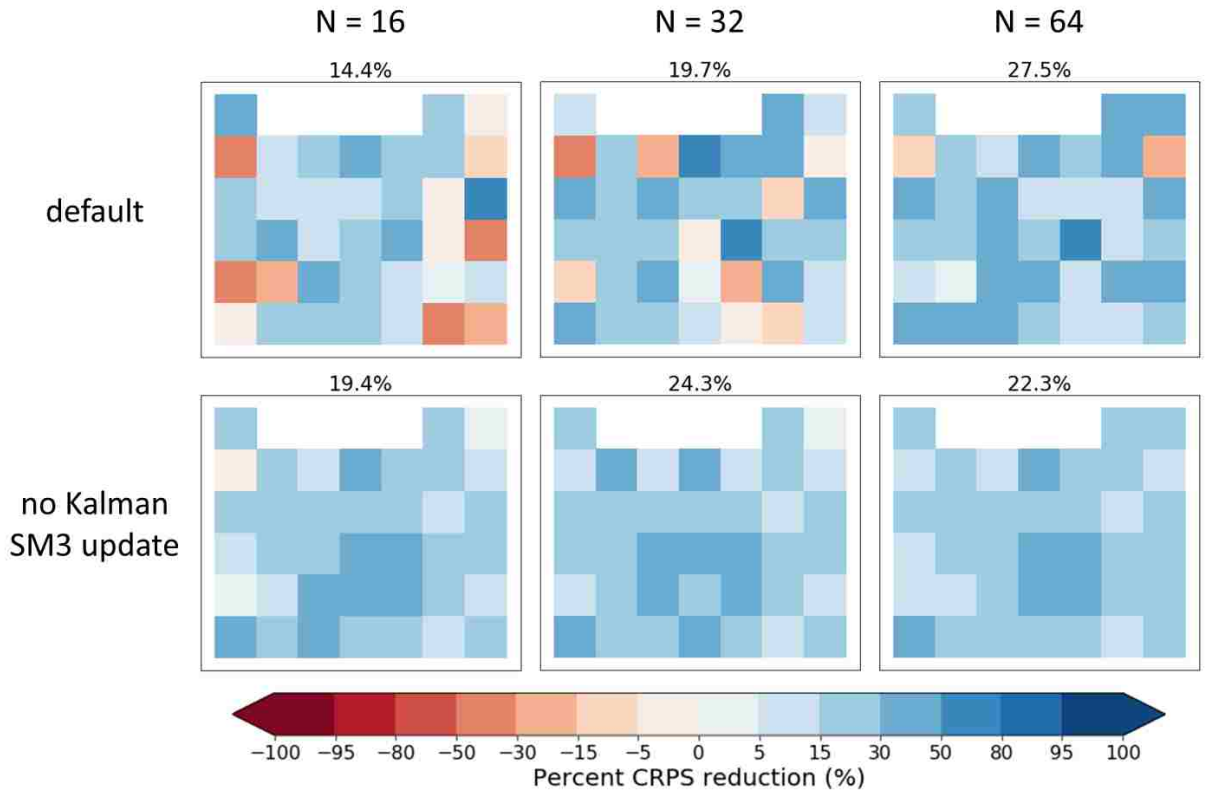

Figure 2.8. Same as Figure 2.7, but for percent CRPS reduction (PSR) of SM3.

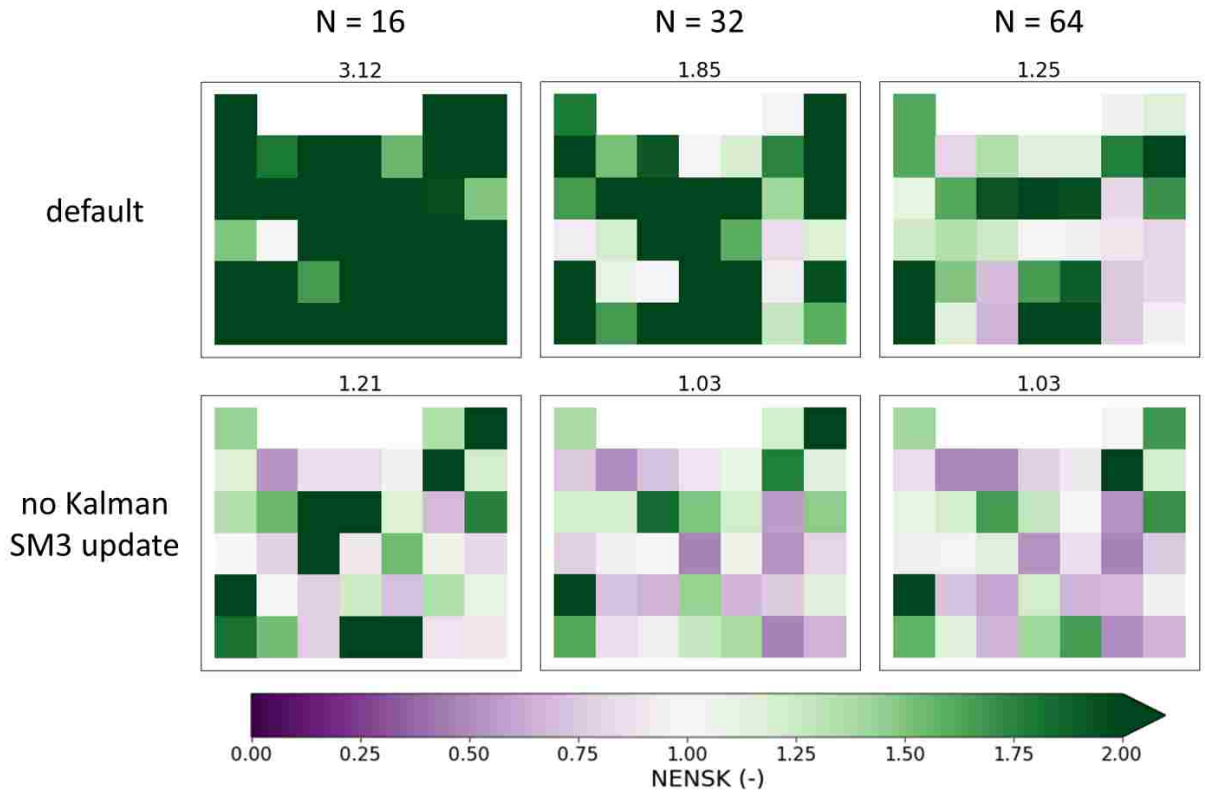

Figure 2.9. Same as Figure 2.7, but for NENSK of SM3. Lighter color (either purple or green) indicates a NENSK closer to one and, therefore, better ensemble representation.

b) Whole-domain (Arkansas-Red) results from the synthetic twin experiment

Based on the previous small-domain studies, our final DA-setup in the case study used 32 ensemble members and no SM3 updating in order to balance DA performance and computational cost. We expect that these choices will provide an optimal EnKF analysis. Figure 2.10 and Figure 2.11 show PER and PSR results, respectively, of the whole-domain synthetic twin experiment. Both SM1 and SM2 states are effectively improved, with > 30% domain-median PER and approximately 30% PSR. The fact that the update of SM2 is only slightly less effective than that of SM1 suggests that SM1 and SM2 are highly coupled in the VIC model, and that the information contained in the surface measurement can be used effectively to update SM2. SM3 states are moderately improved as well, with domain-median PER of 24.8% and PSR of 13.7%. PSR is in general smaller than PER, especially for SM3, possibly related to the challenge of a good ensemble representation in the bottom layer as discussed in Section 2.3.3.a. The

deterministic SM3 improvement shows a clear spatial pattern, with greater improvement in some sub-regions in the east and much less improvement in the west. In our case, the vertical coupling is much weaker in the west of the basin and stronger in some sub-regions in the east (see Figure 2.10). This is consistent with Kumar et al. [2009] who found that the effectiveness of deeper layer SM updates via surface assimilation depends on the degree of vertical coupling between soil layers, which is, in turn, determined by the specific model used, model parameters, and climate.

Even in the synthetic twin experiment containing no systematic error and assuming perfect statistical knowledge of errors, fast-response runoff is essentially unimproved across the entire domain. This is unsurprising, given the finding from the perfect-state analysis in Section 2.3.2that most error in fast-response runoff does not originate from error in the antecedent states. Slow-response runoff, on the other hand, does show improvement, with a domain-median PER value of 20.0% and a value PSR of 13.0%. Combined, total runoff is on average slightly improved (domain-median PER of 7.8% and PSR of 7.1%) and shows the same spatial gradient noted before in the perfect-state experiments (i.e., slight improvement in the east and almost no improvement in the west).

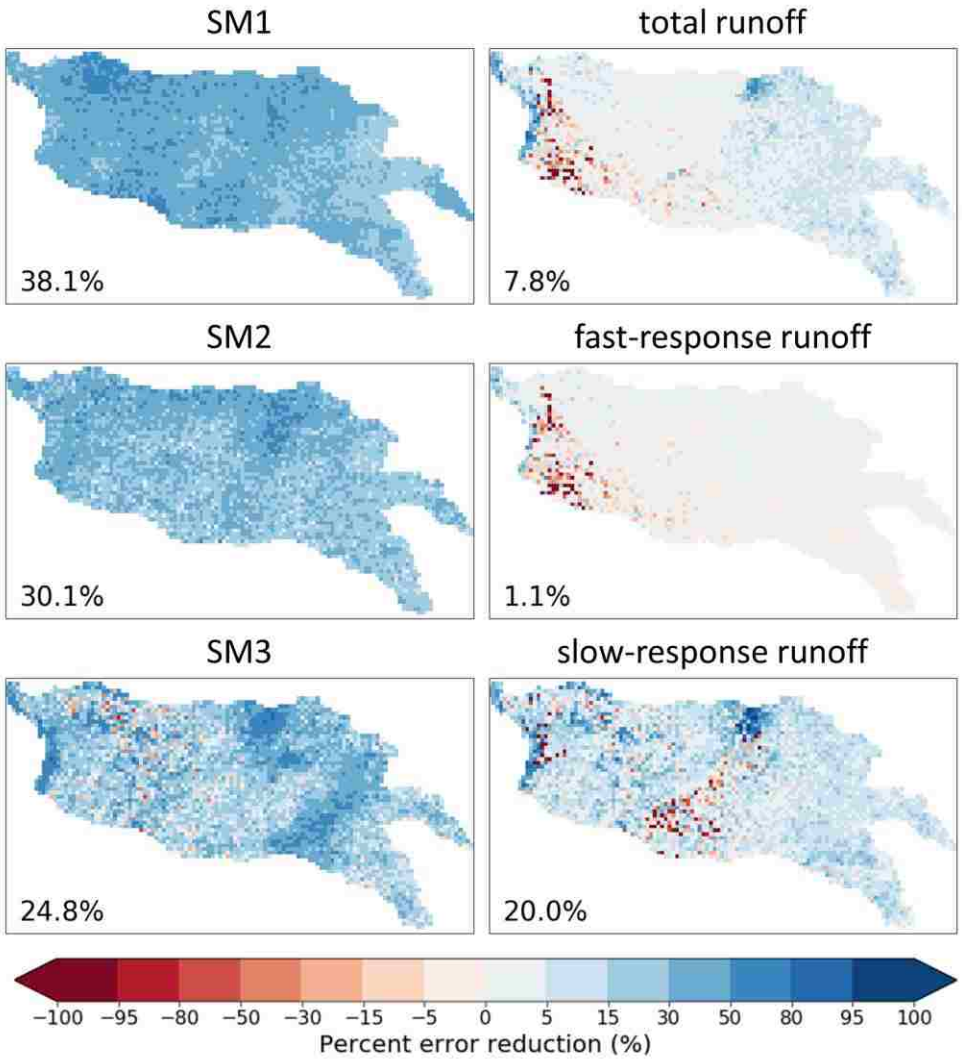

Figure 2.10. Maps of percent error reduction (PER) of DA-updated SM states and daily runoff fluxes from the synthetic experiment. Domain-median PER values are labeled on each subplot.

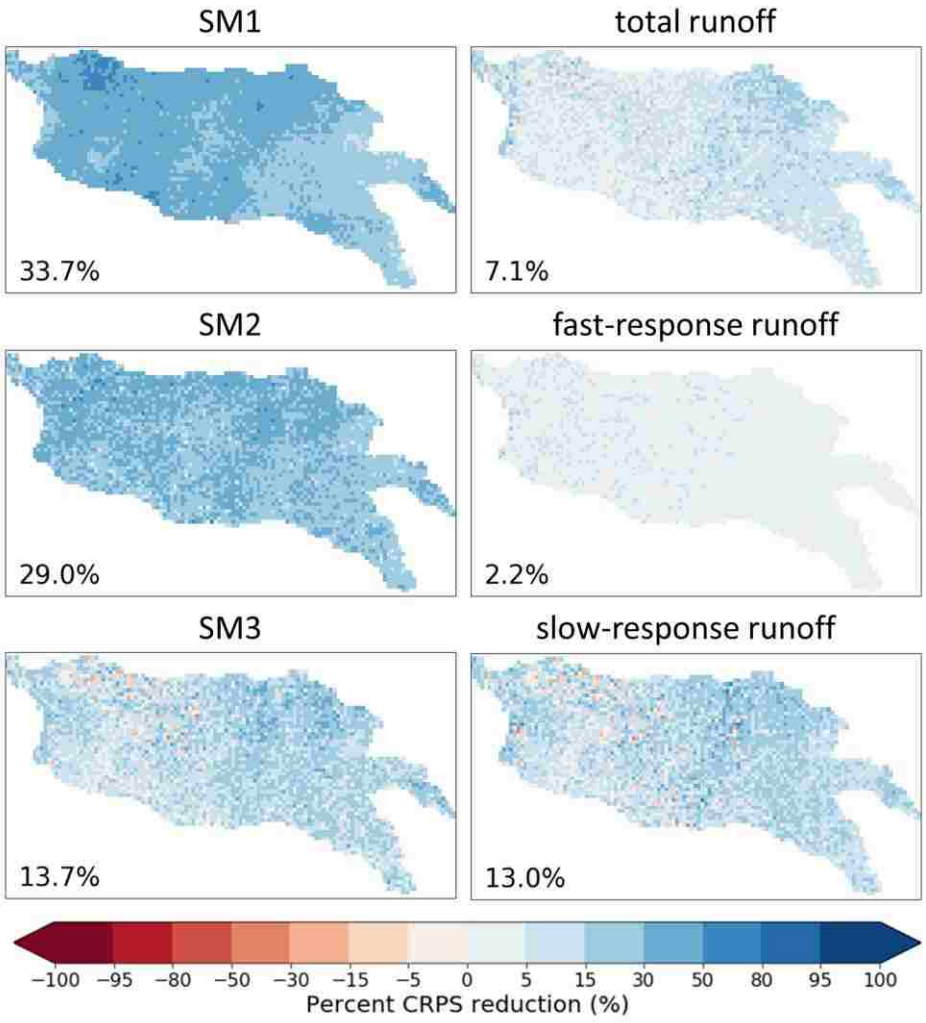

Figure 2.11. Same as Figure 2.10 but for percent CRPS reduction (PSR).

c) Diagnosing the real-data results using the synthetic experiments

We further compare the results from the synthetic SM DA with those from the SMAP-assimilated real-data experiment presented earlier in Section 2.3.1 at the 8 USGS sites (Table 2.2, the "synthetic, total runoff" and "real-data" columns). In general, larger synthetic PER corresponds to larger real-data PER across sub-basins (this correspondence is less clear in PSR partly because real-data PSR is near-zero for most sub-basins). However, the real-data runoff improvement is much smaller than the synthetic improvement at all sub-basins because of significant systematic error existing in the real-data simulation case, which is not correctable by SM DA. SMAP data quality, SMAP rescaling, spatial grid-matching techniques, and incorrect

error assumptions may further affect the real-data case. While these factors are interesting research topics themselves, they have been investigated by previous studies (e.g., Massari et al., 2015; Alvarez-Garreton et al., 2013) and are not the focus of this study.

Finally, we examine the synthetic fast- and slow-response runoff separately for each sub-basin (Table 2.2). For the two example sites shown in Figure 2.1, they have obviously different characteristics: in the Little Arkansas sub-basin there is stronger vertical coupling and therefore better synthetic performance, while in the Deep River sub-basin there is weaker vertical coupling and therefore poorer synthetic performance. The stronger vertical coupling at Little Arkansas is reflected in the real-data time series (Figure 2.1) which shows a baseflow recession more sensitive to the DA update than in the Deep sub-basin. This results in both a larger deterministic streamflow update and a larger ensemble spread that better covers the observation for the Little Arkansas. The flashy streamflow peaks are almost unaffected by DA at both sites because of the weak coupling between runoff and top layer SM.

Assisted by the understanding gained from the synthetic framework, we are able to extend the small-sub-basin results to parts of our study domain where unregulated streamflow is unavailable for model evaluation. We do not expect successful streamflow improvement from SMAP assimilation in the western part of the Arkansas-Red Basin where fast-response runoff is dominant.

2.4 CONCLUSIONS

In this paper, we present a diagnostic framework that builds on a set of synthetic experiments to decompose the error terms in streamflow simulation and to provide quantified understanding of limiting factors in a SM DA system. Specifically, the synthetic framework captures the random error components in the simulation system, distinguishes the runoff error terms that originate from errors in SM states, and assesses the effectiveness of SM DA to correct them.

The usage of the general framework is illustrated through a VIC case study in the Arkansas-Red River basin. A significant level of systematic error was found in the real-data experiment, which imposes a first-order limit on the extent to which SMAP assimilation can improve streamflow. We further found through the synthetic framework that approximately 60%

of the random runoff error in the basin comes directly from precipitation forcing rather than antecedent states and is therefore not correctable by SM DA. In particular, fast-response runoff has weak coupling with antecedent states and is insensitive to state updating. Slow-response runoff, which is a smaller component in total runoff in much of the basin, relies strongly on the bottom-layer SM for its correction. This critical layer is excluded from the EnKF state vector to avoid suboptimal filter results, and its improvement relies on its response to SM1 and SM2 updating and is only moderately improved. Therefore, total runoff improvement is quite low - even in a synthetic twin experiment assuming purely random errors and a perfect statistical parameterization of these errors. Real SMAP assimilation exhibits even less successful streamflow improvement due to the additional presence of systematic error and other factors such as SMAP rescaling and inaccurate error assumptions.

The results from our case study highlight that state correction alone is not sufficient to improve streamflow substantially. Although this conclusion is drawn from the specific case study, it likely holds for a generic streamflow simulation system, especially in large-scale basins, where systematic error and rainfall error are inevitable. Therefore, in addition to more sophisticated SM DA techniques, substantial streamflow improvement would only be possible if future research focuses on reducing these error components that are not correctable by SM DA. Systematic error reduction for large-scale basins is a challenging task but has its potential via, for example, innovative model parameterization techniques. For example, Mizukami et al. [2017] calibrated “transfer functions” instead of model parameters themselves, which can then be applied to distributed model parameterization. The increasing availability of large-scale, satellite-observed distributed hydrologic datasets such as SMAP and Global Precipitation Mission (GPM) data [Hou et al., 2014] is especially promising for informing model representation. Crow et al. [2018], for example, diagnosed systematic coupling errors between antecedent states and runoff generation processes in land surface models using the SMAP data. As for rainfall error, a dual correction system has been developed in the past decade whereby SM measurements are used to simultaneously update antecedent SM states and correct rainfall estimates [e.g., Chen et al., 2014; Alvarez-Garreton et al., 2016]. The findings from this study support the potential of such correction schemes.

Acknowledgements

This work was supported in part by NASA Terrestrial Hydrology Program Award NNX16AC50G to the University of Washington and NASA Terrestrial Hydrology Program Award 13-THP13-0022 to the United States Department of Agriculture, Agricultural Research Service. Yixin Mao also received a Pathfinder Fellowship by CUAHSI with support from the National Science Foundation (NSF) Cooperative Agreement No. EAR-1338606. The VIC model used in the study is available at https://github.com/UW-Hydro/VIC. Specifically, we used VIC version 5.0.1 (doi:10.5281/zenodo.267178) with a modification to the calculation of drainage between soil layers (https://github.com/UW-Hydro/VIC/releases/tag/Mao_etal_stateDA_May2018). The DA code used in this study is available at https://github.com/UW-Hydro/dual_DA_SMAP.

REFERENCES

Alvarez-Garreton, C., D. Ryu, A. W. Western, W. Crow, and D. Robertson (2013), Impact of observation error structure on satellite soil moisture assimilation into a rainfall-runoff model, 20th International Congress on Modelling and Simulation, Adelaide, Australia.

Alvarez-Garreton, C., D. Ryu, A. W. Western, W. T. Crow, and D. E. Robertson (2014), The impacts of assimilating satellite soil moisture into a rainfall-runoff model in a semi-arid catchment, *J. Hydrol.*, 519, 2763-2774, doi:10.1016/j.jhydrol.2014.07.041.

Alvarez-Garreton, C., D. Ryu, A. W. Western, W. T. Crow, C.-H. Su, and D. R. Robertson (2016), Dual assimilation of satellite soil moisture to improve streamflow prediction in data-scarce catchments, *Water Resour. Res.*, 52(7), 5357-5375, doi:10.1002/2015WR018429.

Anderson, J. L. (2007), Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter, *Physica D*, 230(1-2), 99-111, doi:10.1016/j.physd.2006.02.011.

Aubert, D., C. Loumagne, and L. Oudin (2003), Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall-runoff model, *J. Hydrol.*, 280(1-4), 145-161, doi:10.1016/S0022-1694(03)00229-4.

Bohn, T. J., B. Livneh, J. W. Oyler, S. W. Running, B. Nijssen, and D. P. Lettenmaier (2013), Global evaluation of MTCLIM and related algorithms for forcing of ecological and

hydrological models, *Agric. For. Meteorol.*, 176, 38-49, doi:10.1016/j.agrformet.2013.03.003.

Brocca, L., F. Melone, T. Moramarco, W. Wagner, V. Naeimi, Z. Bartalis, and S. Hasenauer (2010), Improving runoff prediction through the assimilation of the ASCAT soil moisture product, *Hydrol. Earth Syst. Sci.*, 14, 1881-1893, doi:10.5194/hess-14-1881-2010.

Brocca, L., T. Moramarco, F. Melone, W. Wagner, S. Hasenauer, and S. Hahn (2012), Assimilation of surface-and root-zone ASCAT soil moisture products into rainfall–runoff modeling, *IEEE Trans. Geosci. Remote Sens.*, 50(7), 2542-2555, doi: 10.1109/TGRS.2011.2177468.

Brooks, R. H., and A. T. Corey (1964), *Hydraulic properties of porous media*, Hydrology papers, Colorado State University, Fort Collins, CO.

Chan, S. et al. (2017), Development and validation of the SMAP enhanced passive soil moisture product, Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International, doi: 10.1109/IGARSS.2017.8127512.

Chen, F., W. T. Crow, P. J. Starks, and D. N. Moriasi (2011), Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture, *Adv. Water Resour.*, 34, 526-536, doi:10.1016/j.advwatres.2011.01.011.

Chen, F., W. T. Crow, and D. Ryu (2014), Dual forcing and state correction via soil moisture assimilation for improved rainfall–runoff modeling, *J. Hydrometeorol.*, *15*(5), 1832–1848, doi:10.1175/JHM-D-14-0002.1.

Ciach, G. J., and W. F. Krajewski (1999), Radar-rain gauge comparisons under observational uncertainties, *J. Appl. Meteorol.*, *38*, 1519-1525, doi:10.1175/1520-0450(1999)038<1519:RRGCUO>2.0.CO;2.

Colliander, A. et al. (2017), Validation of SMAP surface soil moisture products with core validation sites, *Remote Sens. Environ.*, 191, 215-231, doi:10.1016/j.rse.2017.01.021.

Crow, W. T., and J. D. Bolten (2007), Estimating precipitation errors using spaceborne surface soil moisture retrievals, Geophys. Res. Lett., 34, L08403, doi:10.1029/2007GL029450.

Crow, W. T., and R. H. Reichle (2008), Comparison of adaptive filtering techniques for land surface data assimilation, *Water Resour. Res.*, 44, W08423, doi:10.1029/2008WR006883.

Crow, W. T., and D. Ryu (2009), A new data assimilation approach for improving hydrologic prediction using remotely-sensed soil moisture retrievals, *Hydrol. Earth Syst. Sci.*, 12(1-16), doi:10.5194/hess-13-1-2009.

Crow, W.T., D. G. Miralles, and M. H. Cosh (2010), A quasi-global evaluation system for satellite-based surface soil moisture retrievals, *IEEE Trans. Geosci. Remote Sens.*, 48(6), 2516-2527, doi:10.1109/IGARSS.2008.4779051.

Crow, W. T., M. J. van den Berg, G. J. Huffman, and T. Pellarin (2011), Correcting rainfall using satellite-based surface soil moisture retrievals: The Soil Moisture Analysis Rainfall Tool (SMART), *Water Resour. Res.*, 47, W08521, doi:10.1029/ 2011WR010576.

Crow, W. T., F. Chen, R. H. Reichle, and Q. Liu (2017), L band microwave remote sensing and land data assimilation improve the representation of prestorm soil moisture conditions for hydrologic forecasting, *Geophys. Res. Lett.*, 44, 5495-5503, doi:10.1002/2017GL073642.

Crow, W. T., F. Chen, R. J. Reichle, Y. Xia, and Q. Liu (2018), Exploiting soil moisture, precipitation, and streamflow observations to evaluate soil moisture/runoff coupling in land surface models, *Geophys. Res. Lett.*, 45, 4869–4878, doi:10.1029/2018GL077193.

De Lannoy, G. J. M., P. R. Houser, V. R. N. Pauwels, and N. E. C. Verhoest (2006), Assessment of model uncertainty for soil moisture through ensemble verification, *J. Geophys. Res.*, 111, D10101, doi: 10.1029/2005JD006367.

Entekhabi, D. et al. (2010), The Soil Moisture Active and Passive (SMAP) Mission, *Proceedings of the IEEE*, 98(5), 704-716, doi:10.1109/JPROC.2010.2043918.

Francois, C., Quesney, A., and C. Ottle (2003), Sequential assimilation of ERS-1 SAR data into a coupled land surface-hydrological model using an extended Kalman filter, *J. Hydrometeorol.*, 4(2), 473-487, doi:10.1175/1525-7541(2003)4<473:SAOESD>2.0.CO;2.

Freeze, R. A., and R. L. Harlan (1969), Blueprint for a physically-based, digitally-simulated hydrologic response model, *J. Hydrol.*, 9(3), 237-258, doi:10.1016/0022-1694(69)90020-1.

Gruber, A., W. T. Crow, W. Dorigo, and W. Wagner (2015), The potential of 2D Kalman filtering for soil moisture data assimilation, *Remote Sens. Environ.*, 171, 137-148, doi: 10.1016/j.rse.2015.10.019.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez (2009), Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, 377, 80-91, doi: 10.1016/j.jhydrol.2009.08.003.

Hamman, J., B. Nijssen, A. Roberts, A. Craig, W. Maslowski, and R. Osinski (2017), The coastal streamflow flux in the Regional Arctic System Model, *J. Geophys. Res.*, 122(3), 1683-1701, doi: 10.1002/2016JC012323.

Hamman, J., B. Nijssen, T. J. Bohn, D. R. Gergel, and Y. Mao (2018), The Variable Infiltration Capacity model version 5 (VIC-5): infrastructure improvements for new applications and reproducibility, *Geosci. Model Dev.*, 11, 3481-3496, doi: 10.5194/gmd-11-3481-2018.

Hersbach, H. (2000), Decomposition of the continuous ranked probability score for ensemble prediction systems, *Weather Forecast.*, 15(5), 559-570, doi: 10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Hou A. Y., R. K. Kakar, S. Neeck, A. A. Azarbarzin, C. D. Kummerow, M. Kojima, R. Oki, K. Nakamura, and T. Iguchi (2014), The Global Precipitation Measurement mission, *Bull. Amer. Meteor. Soc., 95(5), 701-722, doi: 10.1175/BAMS-D-13-00164.1.*

Kumar, S. V., R. H. Reichle, R. D. Koster, W. T. **Crow,** and C. D. Peters-Lidard (2009), Role of subsurface physics in the assimilation of surface soil moisture observations. *J. Hydrometeorol.*, 10(6), 1534–1547, doi:10.1175/2009JHM1134.1.

Kumar, S. V., C. D. Peters-Lidard, D. M. Mocko, R. Reichle, Y. Liu, K. R. Arsenault, Y. Xia, M. Ek, G. Riggs, B. Livneh, and M. Cosh (2014), Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation, *J. Hydrometeorol.*, 15, 2446-2469, doi:10.1175/JHM-D-13-0132.1.

Liang, X., D. P. Lettenmaier, E. F. Wood, and 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.

Lievens, H., et al. (2015), SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia, *Remote Sens. Environ.*, *168*, 146-162, doi:10.1016/j.rse.2015.06.025.

Lievens, H., G. J. M. De Lannoy, A. Al Bitar, M. Drusch, G. Dumedah, H.-J. Hendricks Franssen, Y. H. Kerr, S. K. Tomer, B. Martens, O. Merlin, M. Pan, J. K. Roundy, H. Vereecken, and J. P. Walker (2016), Assimilation of SMOS soil moisture and brightness

temperature products into a land surface model, *Remote Sens. Environ.*, 180, 292-304, doi: 10.1016/j.rse.2015.10.033.

Lohmann, D., R. Nolte-Holube, and E. Raschke (1996), A large-scale horizontal routing model to be coupled to land surface parametrization schemes, *Tellus*, 48(A), 708-721, doi: 10.1034/j.1600-0870.1996.t01-3-00009.x.

Lohmann, D., E. Raschke, B. Nijssen, and D. P. Lettenmaier (1998), Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model, *Hydrol. Sci. J.*, 43(1), 131-141, doi: 10.1080/02626669809492107.

Massari, C., L. Brocca, A. Tarpanelli, and T. Moramarco (2015), Data Assimilation of Satellite Soil Moisture into Rainfall-Runoff Modelling: A Complex Recipe?, *Remote Sens.*, 7, 11403-11433, doi:10.3390/rs70911403.

Maurer, E.P., A.W. Wood, J.C. Adam, D.P. Lettenmaier, and B. Nijssen (2002), A long-term hydrologically-based data set of land surface fluxes and states for the conterminous United States, *J. Clim.*, 15(22), 3237-3251, doi:10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2.

Mehra, R. K. (1971), On-line identification of linear dynamic systems with applications to Kalman filtering, *IEEE Trans. Autom. Control.*, 16(1), 12-21, doi:10.1109/TAC.1971.1099621.

Mizukami, N., M. P. Clark, A. J. Newman, A. W. Wood, E. D. Gutmann, B. Nijssen, O. Rakovec, and L. Samaniego (2017), Towards seamless large-domain parameter estimation for hydrologic models, *Water Resour. Res.*, 53, 8020–8040, doi:10.1002/2017WR020401.

Newman, A. J., M. P. Clark, and J. Craig (2015), Gridded ensemble precipitation and temperature estimates for the contiguous United States, *J. Hydrometeorol.*, 16(6), 2418-2500, doi:10.1175/JHM-D-15-0026.1.

Nijssen, B., G. M. O'Donnell, D. P. Lettenmaier, D. Lohmann, and E. F. Wood (2001), Predicting the discharge of global rivers, *J. Clim.*, *14*(15), 3307-3323, doi:10.1175/1520-0442(2001)014<3307:PTDOGR>2.0.CO;2.

Nijssen, B., and D. P. Lettenmaier (2004), Effect of precipitation sampling error on simulated hydrological fluxes and states: anticipating the Global Precipitation Measurement satellites, *J. Geophys. Res.*, 109, D02103, doi:10.1029/2003JD003497.

O'Neill, P. E., S. Chan, E. G. Njoku, T. Jackson, and R. Bindlish (2016), SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Version 4, Boulder, Colorado USA, NASA National Snow and Ice Data Center Distributed Active Archive Center, Accessed 2018-01-18, doi:10.5067/OBBHQ5W22HME.

Parajka, J., V. Naemi, G. Bloschl, W. Wagner, R. Merz, and L. Scipal (2006), Assimilating scatterometer soil moisture data into conceptual hydrologic models at the regional scale, *Hydrol. Earth Syst. Sci.*, 10, 353-368, doi:10.5194/hess-10-353-2006.

Pauwels, R., R. Hoeben, N. Verhoest, F. De Troch, and P. Troch (2002), Improvement of TOPLATS-based discharge predictions through assimilation of ERS-based remotely sensed soil moisture values, *Hydrol. Process.*, 16, 995-1013, doi:10.1002/ hyp.315.

Pauwels, V. R. N., and G. J. M. De Lannoy (2009): Ensemble-based assimilation of discharge into rainfall-runoff models: A comparison of approaches to mapping observational information to state space, *Water Resour. Res.*, 45(8), W08428, doi: 10.1029/2008WR007590.

Reichle, R. H., W. T. Crow, and C. L. Keppenne (2008), An adaptive ensemble Kalman filter for soil moisture data assimilation, *Water Resour. Res.*, *44*(3), 1-13, doi:10.1029/2007WR006357.

Ryu, D., W. T. Crow, X. Zhan, and T. J. Jackson (2009), Correcting unintended perturbation biases in hydrologic data assimilation, *J. Hydrometeorol.*, 10, 734-750, doi: 10.1175/2008JHM1038.1.

Talagrand, O., R. Vautard, and B. Strauss (1997), Evaluation of probabilistic prediction systems, technical report, Eur. Cent. for Medium-Range Weather Forecast., Reading, UK.

Todini, E. (1996), The ARNO rainfall-runoff model, *J. Hydrol.*, *175*(1-4), 339-382, doi:10.1016/S0022-1694(96)80016-3.

United States Geological Survey (USGS) (2018), USGS Surface-water daily data for the nation. [Available at https://waterdata.usgs.gov/nwis/dv/?referred_module=sw.]

Wanders, N., D. Karssenberg, A. De Roo, S. M. De Jong, and M. F. P. Bierkens (2014), The suitability of remotely sensed soil moisture for improving operational flood forecasting, *Hydrol. Earth Syst. Sci.*, *18*(6), 2343-2357, doi:10.5194/hess-18-2343-2014.

Western, A. W., R. B. Grayson, and G. Blöschl (2002), Scaling of soil moisture: a hydrologic perspective, *Annu. Rev. Earth Planet. Sci.*, *30*(1), 149-180, doi:10.1146/annurev.earth.30.091201.140434.

Wilks, D. S. (2011), Statistical methods in the atmospheric sciences (3rd edition), Elsevier/Academic Press, Amsterdam; Boston.

Wood, E.F, D.P. Lettenmaier, and V.G. Zartarian (1992), A land-surface hydrology parameterization with subgrid variability for general circulation models, *J. Geophys. Res.*, 97(D3), 2717-2728, doi:10.1029/91JD01786.

Xia, Y. et al., NCEP/EMC (2009), NLDAS Primary Forcing Data L4 Hourly 0.125 x 0.125 degree V002, Edited by David Mocko, NASA/GSFC/HSL, Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed 2018-02-27, doi:10.5067/6J5LHHOHZHN4.

Zhao, R.-J., Y.-L. Zhang, L.-R. Fang, X.-R. Liu, and Q.-S. Zhang (1980), The Xinanjiang model, *Hydrol. Forecast. Proceed. Oxford Symposium*, IASH, 129, 351-356.

Chapter 3. DUAL STATE/RAINFALL CORRECTION VIA SOIL MOISTURE ASSIMILATION FOR IMPROVED STREAMFLOW SIMULATION: EVALUATION OF A LARGE-SCALE IMPLEMENTATION WITH SMAP SATELLITE DATA

This chapter is to be submitted to *Hydrology and Earth System Sciences* as

Mao Y., W. T. Crow, and B. Nijssen (2018), Dual state/rainfall correction via soil moisture assimilation for improved streamflow simulation: Evaluation of a large-scale implementation with SMAP satellite data, *Hydrology and Earth System Sciences*, manuscript in preparation.

The supplemental material for this chapter is provided in appendix B.

Abstract

Soil moisture (SM) measurements contain information about both pre-storm hydrologic states and within-storm rainfall estimates, both of which are essential for accurate streamflow simulation. In this study, an existing dual state/rainfall correction system is extended and implemented in a regional-scale basin with a semi-distributed land surface model. The latest Soil Moisture Active Passive (SMAP) satellite surface SM retrievals are assimilated to simultaneously correct antecedent SM states in the model and rainfall estimates from the latest Global Precipitation Measurement (GPM) mission. While the GPM rainfall is corrected slightly to moderately especially for larger events, the correction is smaller than reported in past studies because of the improved baseline quality of the new generation GPM satellite product. The streamflow is corrected slightly to moderately via dual correction across 8 Arkansas-Red subbasins, with larger correction at subbasins with poorer GPM rainfall and baseline streamflow simulation, and smaller correction at subbasins with better GPM rainfall and baseline streamflow quality. Overall, although the dual data assimilation scheme is able to nudge streamflow simulations toward the correct direction, it only corrects a relatively small portion of the total streamflow error. Systematic modeling error accounts for a large portion of the streamflow error,

which are uncorrectable by standard data assimilation techniques. These findings suggest that we may be approaching the limit of using SM data assimilation to correct random errors in streamflow simulation. More substantial streamflow correction would rely on future research effort on reducing systematic error and developing higher-quality satellite rainfall products.

3.1 INTRODUCTION

Accurate streamflow simulation is important for water resources management such as flood control and drought monitoring. Reliable streamflow simulation requires accurate soil moisture (SM) conditions since they control the partitioning of infiltration and surface runoff during rainfall events as well as longer-memory subsurface flow [Freeze and Harlan, 1969; Western et al., 2002; Aubert et al., 2003]. Good streamflow simulations also require realistic within-storm rainfall time series estimates.

SM measurements, if available, contain information about both antecedent hydrologic states and preceding rainfall events. With the advance of in-situ and satellite-measured SM products, researchers have started to explore the potential of using SM measurements to improve both aspects. For example, a number of studies have attempted to assimilate SM measurements to improve antecedent SM states in hydrologic models via Kalman-filter-based techniques [e.g., Francois et al., 2003; Brocca et al., 2010, 2012; Wanders et al., 2014; Alvarez-Garreton et al., 2014; Lievens et al., 2015, 2016; Massari et al., 2015; Mao et al., 2018]. Other studies explored approaches to use SM measurements to back-calculate rainfall or to correct existing rainfall products [e.g., Crow et al., 2011; Chen et al., 2012; Brocca et al., 2013; Brocca et al., 2014; Brocca et al., 2016; Koster et al., 2016].

In the recent decade, a so-called dual state/rainfall correction system has been developed that combines the state update and rainfall correction schemes with the goal of subsequently improving streamflow simulations [e.g., Crow and Ryu, 2009; Chen et al., 2014; Alvarez-Garreton et al., 2016]. Specifically, SM measurements (typically from satellite observation) are used to simultaneously update model states and correct a rainfall product (also typically satellite-observed). The updated antecedent states and corrected rainfall are then brought together as inputs into the hydrologic model to produce improved streamflow simulation (see Figure 3.1 for illustration of the dual correction system). These past studies in general found further-improved

streamflow performance via such a dual correction system compared to either the state-update-only or the rainfall-correction-only scheme [Crow and Ryu, 2009; Chen et al., 2014; Alvarez-Garreton et al., 2016], with the rainfall correction contributing more during high flow simulations and the state update during low flows [also see Massari et al., 2018].

While these past studies had encouraging findings, they applied the dual correction system only to catchment-scale, lumped hydrologic models. In this study, a semi-distributed land surface model, the Variable Infiltration Capacity (VIC) model, is implemented instead. The VIC model, compared to the previous lumped models, includes a more detailed representation of both energy and water balance processes [Liang et al., 1994; Hamman et al., 2018]. The macroscale grid-based VIC also better matches the spatial resolution of satellite SM measurements, and provides the potential for large-scale streamflow analysis.

We also evaluate the performance of the dual correction method using the latest satellite SM and precipitation data products. The aforementioned dual correction studies used previous-generation satellite products such as the Advanced Scatterometer (ASCAT) satellite SM data, the Soil Moisture Ocean Salinity (SMOS) satellite SM data and the Tropical Rainfall Measuring Mission (TRMM) precipitation data. Instead, we use data products from the more recent Global Precipitation Measurement (GPM) mission [Hou et al., 2014] and the NASA Soil Moisture Active Passive (SMAP) mission [Entekhabi et al., 2010]. Both the SMAP and GPM products provide real-time measurements over much of the global land surface, making them especially useful for regions with scarce in-situ rainfall and SM observations.

The main objective of this study is to assess the effectiveness of such a dual correction system to improve streamflow simulations with the latest satellite SM and precipitation products. To address this main objective, we introduce a number of methodological advances. We extend the system to provide a probabilistic streamflow estimate via ensemble simulations (past studies focused solely on deterministic improvement), we update the rainfall correction scheme to take advantage of new features of the satellite data, and we investigate the potential cross-correlation of errors in the dual system and validate the theoretical correctness of the system design. These methodological contributions will be presented throughout the paper.

The remainder of this paper is organized as follows. Section 3.2 describes the dual correction system and our novel methodological contributions, as well as the study domain,

hydrologic model and datasets used. Results are presented in Section 3.3. Section 3.4 discusses a few remaining issues and takeaways from the study, and Section 3.5 summarizes conclusion.

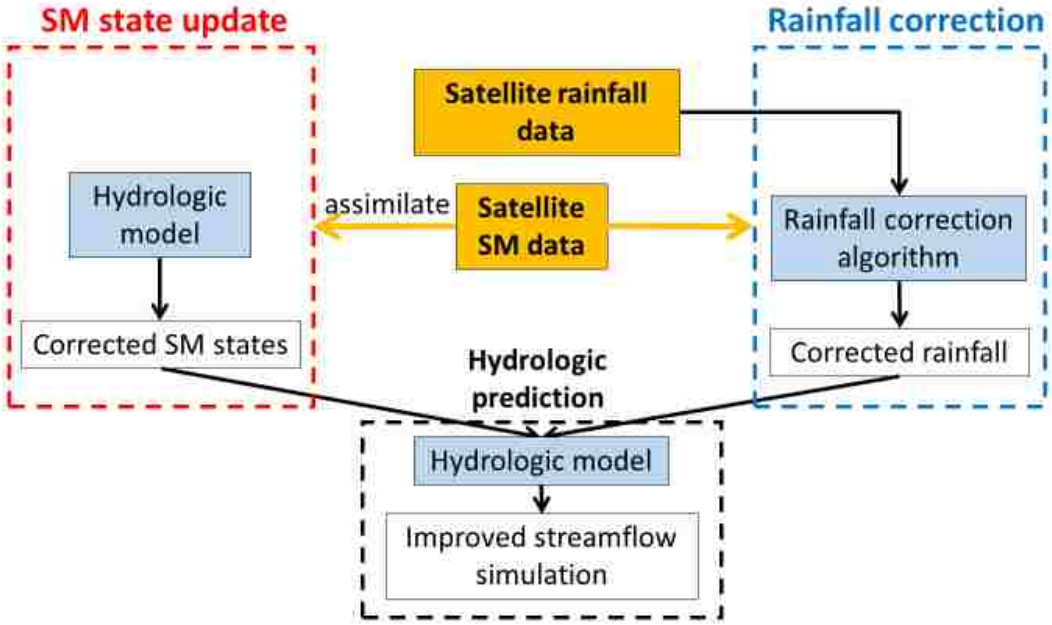

Figure 3.1. The dual state/rainfall correction framework applied in this study. Satellite-based soil moisture (SM) data is integrated into two correction schemes: 1) a hydrologic model to correct modeled states (shown in the red box on the left), and 2) a rainfall correction algorithm to correct rainfall (shown in the blue box on the right). Finally, these two contributions are fused to improve streamflow simulations (shown in the black box at the bottom).

3.2 METHODS

3.2.1 Study domain

The dual state/rainfall correction system is applied in the Arkansas-Red River basin (approximately 605,000 km^2) located in the south-central United States (Figure 3.2). This basin consists of the Arkansas River and the Red River, both converging eastward into the Mississippi River. This domain has a strong climatic gradient (i.e., wetter in the east and drier in the west,

see Figure 3.2 for climatology-averaged precipitation). The basin experiences little snow cover in winter except for the mountainous areas along its western edge. Vegetation cover is denser in the east (deciduous forest) than in the west (wooded grassland, shrubs, crops and grassland).

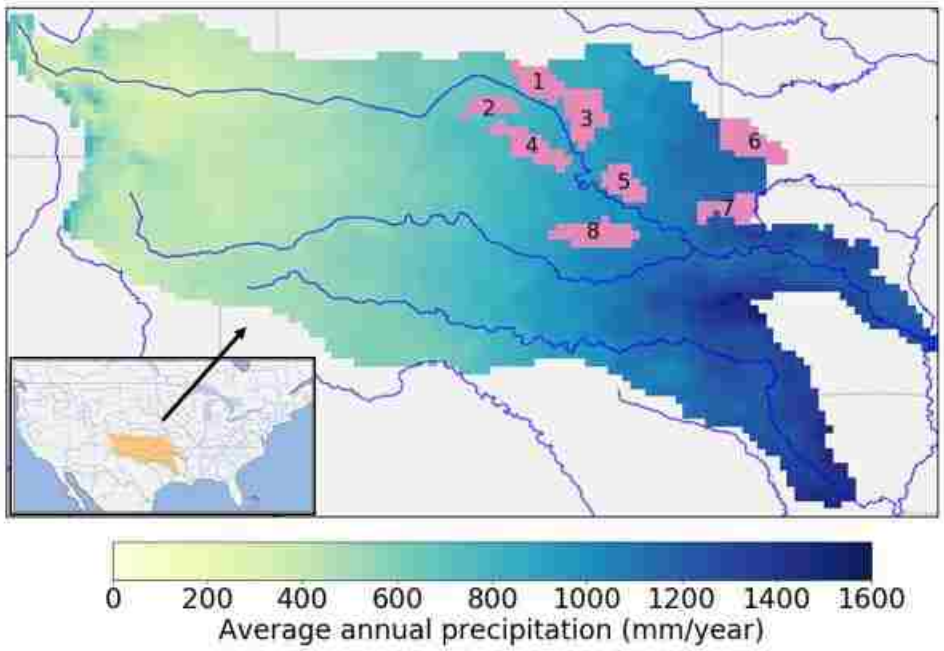

Figure 3.2. The Arkansas-Red River basin with climatology-averaged annual precipitation (calculated from NLDAS-2 precipitation data over 1979-2017). The pink shaded areas show the upstream subbasins of the 8 USGS streamflow sites evaluated in this study, with basin numbers labeled on the plot (see Table 3.3 for basin numbers and corresponding sites).

Table 3.3. List of USGS streamflow sites used for verification.

Basin number	USGS station no.	USGS station name	Short name
1	07144200	Little Arkansas River at Valley Center, KS	L Arkansas
2	07144780	Ninnescah River AB Cheney Re, KS	Ninnescah
3	07147800	Walnut River at Winfield, KS	Walnut
4	07152000	Chikaskia River near Blackwell, OK	Chikaskia
5	07177500	Bird Creek Near Sperry, OK	Bird
6	07186000	Spring River near Wace, MO	Spring
7	07196500	Illinois River near Tahlequah, OK	Illinois
8	07243500	Deep Fork near Beggs, OK	Deep

3.2.2 Data

3.2.2.1. SMAP satellite SM data

The SMAP mission provides SM estimates for the top 5 centimeters of the soil column, with an average revisit time of 2-3 days, a resolution of 36 km and a 50-hour data latency. Both ascending (PM) and descending (AM) retrievals from the SMAP L3 Passive product [O'Neill et al., 2016] (data Version 4) from Mar 31, 2015 to December 31, 2017 were used in this study. A few SMAP pixels with obvious quality flaws (i.e., near-constant retrieval values) were manually masked out. The internal quality flags provided by the SMAP mission were not applied in this study to preserve the measurements in the east half of the domain, where the data quality of the entire region is flagged as unrecommended due to relatively heavy vegetation cover. The native 36 km SMAP retrievals were used throughout the study without spatial remapping or temporal aggregation.

3.2.2.2 GPM satellite precipitation data

The Integrated Multi-satellitE Retrievals for GPM (IMERG) Level 3 Version 05 Early Run precipitation data was used in this study [Huffman et al., 2018]. IMERG merges multiple satellite observations and provides a near-global, precipitation product with a spatial resolution of 0.1º [Huffman et al., 2015]. The Early Run product was used in this study, since its short latency (4 hours) makes it suitable for near-real-time assimilation applications. We aggregated the original 30-minute precipitation product to our 3-hourly modeling timestep and remapped it to the 1/8th model resolution.

3.2.2.3. Other meteorological forcing data

Other than precipitation, the VIC model requires air temperature, shortwave and longwave radiation, air pressure, vapor pressure and wind speed as forcing inputs. These variables were obtained from the 1/8th-gridded North American Land Data Assimilation System Phase 2 (NLDAS-2) meteorological forcing data product [Xia et al., 2009]. In this study, the original hourly NLDAS-2 meteorological variables were aggregated to the 3-hourly modeling timestep.

3.2.2.4. Validation data

Daily streamflow data at 8 USGS streamflow sites in the study domain [USGS, 2018] were used to evaluate the streamflow time series from the dual correction system (see Figure 3.2 and Table 3.3 for site locations). These 8 sites were selected for their lack of human regulation and their dense rain gauge coverage (see Crow et al. [2017] for details). We separately evaluated the rainfall correction scheme, in which the gauge-informed NLDAS-2 precipitation data was treated as the benchmark.

3.2.3 Hydrologic modeling

We used Version 5.0.1 of the VIC model (doi:10.5281/zenodo.267178) with a modification to the calculation of drainage between soil layers (https://github.com/UW-Hydro/VIC/releases/tag/Mao_etal_stateDA_May2018). VIC is a large-scale, semi-distributed model that simulates various land surface processes. In this study, the VIC model was implemented in the Arkansas-Red River basin at a 1/8º spatial resolution (3999 grid cells in total), with each grid cell further divided into multiple vegetation tiles via statistical distributions. Each grid cell was simulated by VIC separately using a soil column discretized into 3 vertical layers (with domain-average thicknesses of 0.10 m, 0.40 m and 0.93 m, respectively). Runoff can be generated by fast-response surface runoff and by slow-response runoff from the bottom soil layer. All vegetation cover and soil property parameters in the model were taken from Maurer et al. [2002], which were calibrated against streamflow observations at the most downstream outlet of the combined Arkansas and Red River basins. The simulation period was from March 2015 to December 2017 when both the SMAP and GPM products are available. The VIC model was spun-up by running the period 1979-2015 twice.

The local runoff simulated by VIC at each grid cell was routed through the stream channels using the RVIC routing model [Hamman et al., 2017]. RVIC is an adapted version of the routing model developed by Lohmann et al. [1996, 1998].

3.2.4 The dual correction system

In this section, we describe the components of the dual correction system. First, we describe our methodological updates to the rainfall correction scheme, followed by a description

of the state update scheme. Next, we present how the two schemes are fused to produce the final ensemble streamflow analysis.

3.2.4.1. The SMART rainfall correction scheme updates and adaption

The Soil Moisture Analysis Rainfall Tool (SMART) rainfall correction algorithm [Crow et al., 2009; 2011; Chen et al., 2012] is based on sequential assimilation of SM measurements into a simple Antecedent Precipitation Index (API) model:

$$API_t = \gamma API_{t-1} + P_t \tag{3.1}$$

where t is a timestep index; P is the original IMERG precipitation observation; γ is a loss coefficient. Here SMAP was rescaled to the API regime through cumulative distribution function (CDF) matching over the 2.5-year simulation period prior to assimilation. We implemented a 3-hourly version of SMART (instead of the daily version in past studies) to receive the 3-hourly IMERG rainfall input and both the ascending (PM) and descending (AM) SMAP retrievals at the correct time of day. We also extended the ensemble Kalman filter (EnKF) version of SMART introduced by Crow et al. [2011] to an ensemble Kalman smoother (EnKS), in which the API increment, δ, is not only calculated at timesteps when SMAP is available, but also back-filled during measurement gaps (see Supplemental Material Section B1 for mathematical details of SMART EnKS).

The SMART algorithm then uses the API increment δ to estimate the rainfall correction amount via a simple linear relation. Here we implemented an ensemble version of rainfall correction that produces an ensemble of corrected rainfall estimates (instead of a single deterministic rainfall time series as in the past SMART studies):

$$P_{corr,t}^{(j)} = P_{pert,t}^{(j)} + \lambda \delta_t^{(j)} \tag{3.2}$$

where the superscript *(j)* denotes the *j*th ensemble member; P_{corr} is the corrected precipitation; P_{pert} is the original IMERG precipitation perturbed in the EnKS algorithm; λ is a scaling factor that can either be calibrated or set to a prescribed constant. In this study we only applied rainfall correction at timesteps when the original IMERG rainfall observation is non-zero, while the zero-rainfall timesteps remain unchanged. This modification was designed to mitigate the degradation introduced by SMART at low-rainfall timesteps, and is particularly suitable for IMERG correction due to its enhanced rain/no rain detection quality compared to the previous generations of satellite precipitation products [Gebregiorgis et al., 2018]. We will discuss the

impact of this modification further in Section 3.3.1. Finally, negative P_{corr} values resulted from Equation (3.2) are reset to zero, and the final corrected precipitation time series is (multiplicatively) rescaled to be unbiased over the entire simulation period against the original IMERG observation.

In this study, the 0.1° IMERG was first remapped to the coarser 36-km SMAP resolution such that the SMART algorithm was run at each of the 36-km pixels individually. The corrected 36-km rainfall was then downscaled to its native 0.1° resolution while maintaining the original finer-grid IMERG spatial pattern, and further regridded to the VIC 1/8° modeling resolution. An ensemble size of $M = 32$ was applied. The API loss coefficient (γ in Equation (3.1)) was manually tuned to a constant 0.98 [3 hours^{-1}] such that the un-corrected API time series has high correlation with SMAP retrievals, which ensures that the API model roughly captures the dynamics of SMAP. We experimented two approaches for estimating the λ parameter in Equation (3.2): 1) calibrating a temporally constant λ at each SMAP pixel separately to optimize the rainfall correlation with the NLDAS-2 benchmark rainfall, and 2) setting λ to a spatial constant of 0.1, which is applicable for any region that may not have a good rain gauge coverage. The impact of the two λ choices will be further discussed in Section 3.3.1.

In our SMART EnKS, the original IMERG precipitation was multiplicatively perturbed by log-normally distributed noise with mean one and standard deviation of one. SMAP measurement error ranges from 0.03 to 0.045 m^3/m^3 across domain. This range was estimated from the SMAP errors reported by the ground validation studies [e.g., Colliander et al., 2017; Chan et al., 2017], and the error for each pixel was linearly mapped from its leaf area index (LAI) value into the range with denser vegetation cover corresponding to larger SMAP error. The *API* state was directly perturbed by zero-mean Gaussian noise with variance 0.3 mm^2 over the entire domain such that the normalized filter innovation has variance of approximately one (which is a necessary condition for proper error assumptions in a Kalman filter; see Mehra [1971] and Crow and Bolten [2007]). See Supplemental Material Section B1 for mathematical details of these error assumptions.

3.2.4.2. State updating via EnKF

As illustrated in Figure 3.1 (the red box on the left), the SMAP SM retrievals were also assimilated into the VIC model to update model states using the EnKF method. Since the

standard EnKF technique has been widely used in many hydrologic state update studies [e.g., Francois et al., 2003; Brocca et al., 2010, 2012; Wanders et al., 2014; Alvarez-Garreton et al., 2014; Lievens et al., 2015, 2016; Massari et al., 2015; Mao et al., 2018], we only briefly describe the specific implementation in this study.

The EnKF implementation in this study mostly follows Mao et al. [2018] with slight modifications. EnKF was implemented as a 1D filter for each 36-km SMAP pixel separately. Each SMAP pixel was assimilated to update the SM states of multiple underlying finer 1/8^{o} VIC grid cells. Therefore, the state vector is a stack of SM states of the top two layers in all vegetation tiles within all these grid cells. Note that the bottom-layer SM state was not updated during EnKF (following Lievens et al. [2015; 2016] and Mao et al. [2018]), although it does respond to the update of the upper two layers through drainage. The observation operator in EnKF maps the state vector to surface SM measurement by averaging the first-layer SM states from the multiple VIC grid cells underlying a single SMAP pixel. An ensemble size of 32 was used, and the resulting ensemble of updated states was preserved to produce probabilistic estimate of states.

The SMAP retrievals were rescaled to match the 2.5-year mean and standard deviation of the VIC-simulated surface-layer SM time series prior to assimilation. The error assumptions of IMERG precipitation and unscaled SMAP retrievals were assumed to be the same as in SMART (described in Section 3.2.4.1). The VIC SM states were directly perturbed (to represent model structure/parameterization error) by zero-mean Gaussian noise with standard deviation of 0.5 mm for all VIC soil layers across the study domain (following Mao et al. [2018]). Although VIC modelling errors are likely to contain some spatial auto-correlation, consideration of the spatial error correlation in our case did not result in better filter performance (results not shown; this finding is consistent with Gruber et al. [2015] who showed limited benefit of a 2D filter with auto-correlated spatial error over a 1D filter with spatially white error). Therefore, modeling errors were assumed to be spatially white and corrected via a 1D EnKF analysis. We will further discuss this choice in Section 3.4.

3.2.4.3. Fusing the state update and the rainfall correction schemes

After carrying out the state update scheme and the rainfall correction scheme separately, the updated model states and the corrected rainfall forcing were fused to produce final

streamflow results (see the black box in the bottom of Figure 3.1 for illustration). We first randomly paired ensemble members of corrected rainfall and updated VIC states, and an arbitrary number of pairs were selected to produce the final streamflow ensemble. Here 32 pairs were selected to balance the computational cost and statistical stability. For each pair, the VIC model was re-run with the updated states inserted sequentially over time and forced by the corrected rainfall. Other meteorological forcings were kept unchanged, i.e., from NLDAS-2. The grid-cell-based runoff output from VIC for each pair was then routed to the 8 USGS gauge locations, resulting in an ensemble of streamflow time series for evaluation. To separate the contribution of the state update and rainfall correction to the final streamflow improvement, two additional sets of streamflow simulations were carried out. The first is "state-updated streamflow", where VIC was re-run with the updated state ensemble only but forced by the original IMERG precipitation. The resulting streamflow reflects only the impact of state updating. The second is "rainfall-corrected streamflow", where VIC was forced by the corrected rainfall ensemble but without inserting the EnKF-updated states. The resulting streamflow therefore reflects only the effect of rainfall correction.

One potential issue with the merging of the two schemes is that, since the state update scheme and the rainfall correction scheme are informed by the same SM measurement data, cross-correlated error exists in the two schemes. Such cross-correlated error could potentially be amplified when fusing the two schemes together, degrading the streamflow results. In fact, Massari et al. [2018] intentionally avoided combining the state and rainfall correction schemes due to this legitimate concern. To formally investigate this issue, we performed a set of synthetic experiments where we compared the following two scenarios: 1) a single set of synthetically generated SM measurements were assimilated into the state and rainfall correction schemes, which mimics the real dual correction system; 2) two SM measurements with mutually independent error were assimilated separately into the two schemes, which is rid of any potential error cross-correlation in the system. The results show that the two scenarios achieve very similar streamflow correction performance. This validated that it is safe to assimilate a single SM measurement product into both schemes without significantly degrading the final streamflow performance (see Supplemental Material Section B2 for more details on the cross-correlated error experiments).

3.2.5 *Evaluation strategies and metrics*

We evaluated the rainfall correction results in addition to the dual-corrected streamflow results. Both deterministic and probabilistic metrics were applied for the two evaluations.

The 1/8th gridded, gauge-informed NLDAS-2 precipitation data was remapped to the 36 km SMART resolution as the benchmark for evaluating rainfall. For all metrics, all precipitation datasets were aggregated to multiple temporal accumulation periods (the native 3-hour period without aggregation; 1-day; 3-day) for evaluation. Deterministically, the ensemble-mean SMART-corrected rainfall was compared to the original IMERG precipitation (remapped to 36 km), and its improvement was evaluated in terms of 1) correlation coefficient (r) of time series; 2) Percent error reduction (PER) in terms of the root-mean-squared error (RMSE) compared to the original IMERG precipitation; 3) Categorical skill metrics, including false alarm ratio (FAR), probability of detection (POD) and threat score (TS) [Wilks, 2011; Crow et al., 2011; Chen et al., 2012; Brocca et al., 2016]. Probabilistically, normalized ensemble skill (NENSK) was calculated, which measures the ensemble-mean error normalized by ensemble spread:

$$NENSK = \frac{ENSK}{ENSP} \tag{3.3}$$

where the ensemble skill (ENSK) is calculated as the temporal mean of ensemble-mean squared error, and the ensemble spread (ENSP) is calculated as the temporal mean of ensemble variance [De Lannoy et al., 2006; Brocca et al., 2012; Alvarez-Garreton et al., 2014; Mao et al., 2018]. Ideally, if an ensemble time series correctly represent the uncertainty of analysis, NENSK should be 1 [Talagrand et al., 1997; Wilks, 2011]. NENSK > 1 indicates an under-dispersed ensemble while NENSK < 1 indicates an over-dispersed ensemble. Note that NENSK is calculated based on the ensemble after correction alone as a verification of uncertainty representation, and is not compared to the baseline.

The dual-corrected streamflow was evaluated at the 8 USGS sites. Deterministically, the ensemble-median corrected streamflow was compared to the baseline streamflow (or the so-called “open-loop” streamflow, which is simply the single VIC simulation forced by IMERG precipitation without any correction) in terms of 1) PER; and 2) the Kling-Gupta efficiency (KGE) [Gupta et al. 2009] which combines the performance of correlation, variance and bias. Ensemble-median instead of ensemble-mean streamflow was used to give more stable evaluation

results in the case of a skewed streamflow ensemble caused by model nonlinearity. The probabilistic metric NENSK was also calculated for streamflow ensemble.

3.3 RESULTS

The rainfall improvement realized via SMART will be evaluated first. Next, the streamflow improvement from the dual correction system will be discussed.

3.3.1 SMART rainfall correction

3.3.1.1. The impact of SMART methodological choices

The impact of the λ parameter on SMART rainfall correction results is first shown in Figure 3.3 by comparing the first two columns, where Column 1 shows the improvement of correlation coefficient *r* after SMART correction with λ tuned at each pixel to optimize *r* (with respect to the NLDAS-2 benchmark), and Column 2 shows that with domain-constant $\lambda = 0.1$. Overall the *r* improvement is relatively small even with optimal λ, which we will further discuss in Section 3.3.1.2 below. Simply setting $\lambda = 0.1$ results in slightly smaller correlation improvement compared to the optimal λ case for all temporal accumulation periods (3-hour, 1-day and 3-day), especially for locations in the east and west ends of the domain. The central part of the domain is not obviously impacted by setting $\lambda = 0.1$. In general, these reductions are small, and since it is a more generally applicable case, we present all the results hereafter (including rainfall correction and dual streamflow correction results) with $\lambda = 0.1$.

Figure 3.3 also compares the impact of implementing the smoother version of SMART (Column 2) with the EnKF version (Column 3). Note that for EnKF, both δ and P in Equation (3.2) were aggregated to 3-day windows prior to correction to ensure SM data availability in every correction window. EnKF results in less *r* improvement overall, which confirms the benefit of applying SMART using a smoothing approach.

Finally, the impact of our modification of only correcting rainfall at non-zero IMERG timesteps is demonstrated by the domain-median categorical metrics (Figure 3.4). If all timesteps are corrected (Figure 3.4 Column 1), FAR is largely degraded (by 0.1 – 0.4) at low rainfall thresholds especially with shorter accumulation periods (3-hour and 1-day; see Figure 3.4a). This

is likely due to the issue of SMART misinterpreting SM retrieval noise as small rainfall correction [Chen et al., 2014]. POD is improved at these low thresholds (Figure 3.4b), but is not able to compensate the large FAR degradation. Therefore, TS, which accounts for both false alarms and missed events, is also obviously degraded at low thresholds (by as large as 0.2 at 3-hourly). In contrast, when applying our tactic of only correcting rainfall at non-zero IMERG timesteps (Figure 3.4 Column 2), such FAR degradation is significantly mitigated (note that the y-axes in the two columns in Figure 3.4 are different). While it does sacrifice POD at low thresholds (Figure 3.4e), the overall TS is improved over most of the thresholds, especially at higher thresholds, for 1-day and 3-day aggregation. As mentioned in Section 3.2.4.1, the success of this SMART modification is likely due to the improved rain/no rain detection quality of the baseline IMERG precipitation product compared to its predecessors. Researchers found that the IMERG product exhibits definite improvement in terms of miss-rain, false-rain and hit rate compared to TRMM TMPA-RT over the Continental U.S. [Gebregiorgis et al., 2018]. Given the good baseline rainfall quality, SMART's tendency to mis-correct low events degrades, rather than helps, rainfall estimates.

3.3.1.2. Rainfall correction evaluation

Besides correlation and categorical metrics, PER exhibits slight to moderate improvement after correction, with a domain-median of ~8% error reduction at 1-day and 3-day accumulation periods (Figure 3.5 Column 1). The positive PER is consistent with the positive (although small) improvement of the categorical metrics at high-event thresholds (Figure 3.4 Column 2), since PER is more sensitive to high rainfall values. 3-hourly PER is less successful (Figure 3.5a), suggesting that the deterministic correction is more effective at an accumulation period that more closely matches the SMAP retrieval interval. Similarly, correlation and categorical results also show better correction at 1-day and 3-day than at 3-hour accumulation periods (Figure 3.3 Column 2 and Figure 3.4 Column 2).

Overall, SMART improves the IMERG rainfall product in terms of various deterministic metrics, especially at larger events. However, the magnitude of rainfall improvement we find here is generally smaller than that found in previous SMART studies, especially in terms of correlation *r* (improved over much of the domain with median improvement of 0.01 to 0.02). The relatively smaller improvement seen here is likely due to the improved accuracy of the baseline

IMERG precipitation product. Table 3.4 summarizes the past SMART studies in literature, including the baseline and benchmark rainfall products, the SM product assimilated, baseline correlation *r* and its improvement, and baseline RMSE and its reduction (PER). It is clear that, over the past decade, the quality of baseline rainfall product has been significantly enhanced, from TRMM 3B40-RT used in Crow et al. [2009] and Crow et al. [2011] with $r = \sim 0.5$, to TRMM 3B42-RT used in Brocca et al. [2016] with $r = \sim 0.6 – 0.7$, to IMERG used in our study with *r* over 0.8. A direct TRMM-IMERG intercomparison study also confirmed the improved accuracy of IMERG over the Continental U.S. in terms of correlation, RMSE, bias and categorical metrics [Gebregiorgis et al., 2018]. The marginal value of SMART is known to decrease as a function of increased baseline rainfall accuracy [Crow et al., 2011]. This trend is clearly demonstrated from the SMART studies in Table 3.4. Although SMAP presumably provides more reliably SM measurements as well relative to the older satellite SM products used in previous SMART applications, its benefit does not appear large enough to outweigh this trend. Another consideration is that the correlation achieved by SMART may be approaching that of the NLDAS-2 benchmark (which does not have perfect accuracy), which may make our SMART evaluation less accurate.

Finally, NENSK shown in Figure 3.5 (Column 2) gives insight to the probabilistic representation of the SMART ensemble. At a 3-hour timestep, NENSK is less than one for most of the domain, indicating an over-dispersed ensemble on average. However, when evaluating at 1-day and 3-day accumulation periods, NENSK is closer to one, indicating a better representation of the uncertainty of rainfall estimates. As we aggregate over longer accumulation windows (e.g., 3-day), NENSK becomes slightly > 1 (i.e., under-dispersed ensemble), since the SMART algorithm only assumes random rainfall error but not systematic bias, and therefore slightly underestimates the uncertainty range over longer-term periods.

In summary, SMART is able to use the SMAP retrievals to correct IMERG rainfall at relatively larger events, with slight to moderate deterministic improvement. Correction at low events is less successful and degraded slightly. Overall, correction is more effective when rainfall estimates are temporally aggregated to periods that better match SMAP retrieval intervals (1-day to 3-day accumulation periods), while the raw 3-hourly correction is less successful. Similarly, the ensemble estimates represent the rainfall uncertainty better at 1-day to 3-day accumulation periods than at 3-hourly timestep.

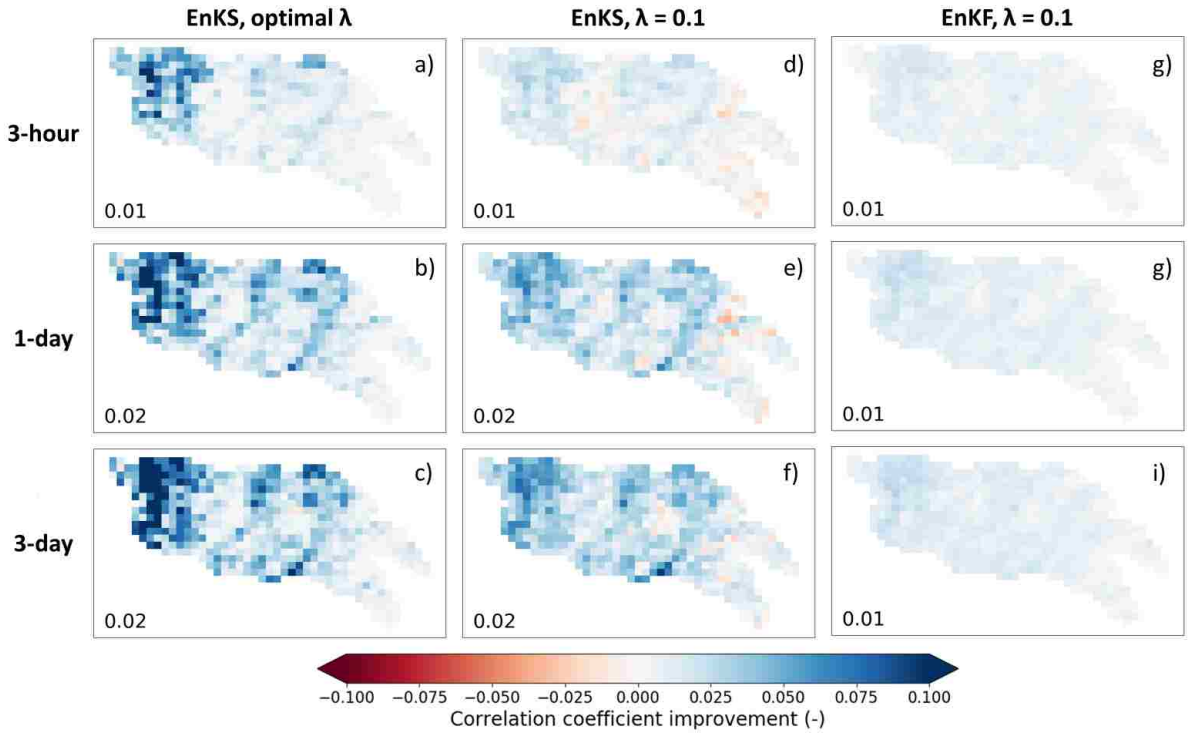

Figure 3.3. Maps of correlation coefficient improvement after SMART rainfall correction. Each column shows the following SMART experiments, respectively: EnKS with λ tuned to optimize correlation coefficient with regards to the NLDAS-2 benchmark (*a*, *b*, *c*); EnKS with constant λ = 0.1 (*d*, *e*, *f*); EnKF with constant λ = 0.1 (*g*, *h*, *i*). Each row shows results based on different temporal accumulation period: 3-hourly, 1-day and 3-day aggregation, respectively. The number on the lower left corner of each subplot shows the domain-median correlation improvement.

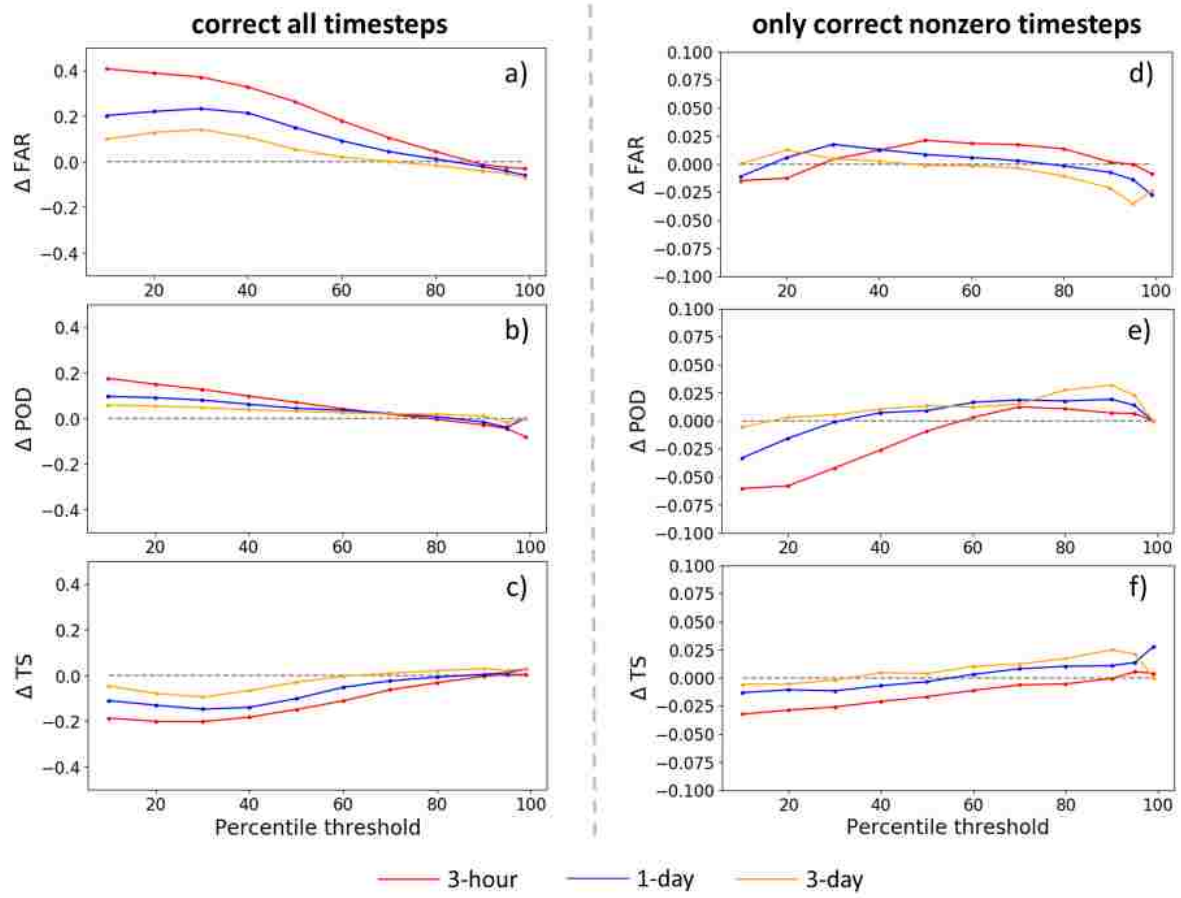

Figure 3.4. Change in categorical metrics (FAR, POD and TS) before and after SMART correction for 3-hourly, 1-day and 3-day accumulation periods. Metrics at different event thresholds are shown on the *x* axis. The left column (*a*, *b*, *c*) is for SMART with rainfall corrected at all timesteps; the right column (*d*, *e*, *f*) is for SMART with rainfall corrected only at non-zero timesteps. Note that the y-axis range is different for the two columns.

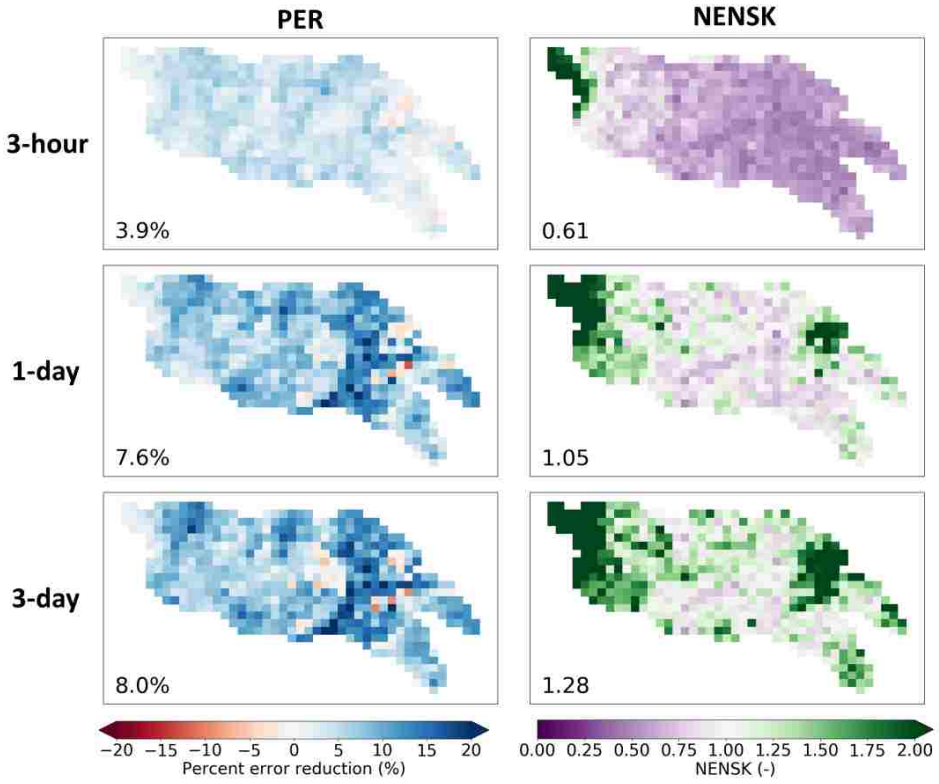

Figure 3.5. Maps of SMART rainfall correction results (with λ = 0.1, EnKS, and rainfall corrected only at non-zero timesteps). Each column shows the following metrics, respectively: percent RMSE reduction (PER) (*a*, *b*, *c*), and ensemble NENSK (*d*, *e*, *f*). Each row shows results based on different temporal accumulation period: 3-hourly, 1-day and 3-days, respectively. The number on the lower left corner of each subplot shows the domain-median statistic.

Table 3.4. Review of SMART rainfall correction results in literature along with the results in this study.

Literature	Baseline rainfall product	Benchmark rainfall product	SM product	Domain	Accumulation period	Baseline correlation *r*	*r* improvement	Baseline RMSE (mm)	PER
Crow et al. [2009]	TRMM 3B40RT	CPC rain gauge analysis	AMSR-E	Southern Great Plain	3-day	~ 0.5	~ + 0.2	13.0	~ 30%
				CONUS	3-day	~ 0.55	~ + 0.05	11.8	~ 15%
Crow et al. [2011]	TRMM 3B40RT	CPC rain gauge analysis	AMSR-E	CONUS	3-day	~ 0.55	~ + 0.1	13.1	~ 20%

Chen et al. [2012]	Princeton Global Forcing Dataset	CPC rain gauge analysis	SMMR, SMM/I, ERS	Global	10-day	~ 0.35	~ + 0.15	-	-
Brocca et al. [2016]	TRMM 3B42RT	AWAP rain gauge product	SMOS	Australia	1-day	0.62	+0.01	5.6	7%
					5-day	0.71	+0.05	14.0	14%
This study	IMERG Early Run	NLDAS-2	SMAP L3 Passive	Arkansas-Red	1-day	0.80	+0.02	6.1	8%
					3-day	0.82	+0.02	11.0	8%

3.3.2 Streamflow from the dual correction system

The final daily streamflow performance from the dual correction system are listed in Table 3.5 (the "*dual*" columns) for each subbasin, both in terms of deterministic ensemble-median performance and probabilistic NENSK verification. Overall, we see small to moderate streamflow improvement after dual correction with large variability across subbasins. Specifically, PER ranges from approximately 6% to 34% and KGE improvement ranges from slightly negative to +0.95 across all subbasins. If using the open-loop KGE (listed in Table 3.5) as a measure of baseline streamflow performance, we observe that at subbasins with better baseline performance (i.e., Ninnescah, Walnut and Chikaskia, all with positive baseline KGE), the relative improvement after the dual correction is generally smaller. In other words, the dual correction system is more effective when the baseline model performance is poor.

Table 3.5 also lists the streamflow improvement from each of the correction schemes alone (the "*state update only*" and "*rainfall correction only*" columns) to further interpret the correction results. For the better-baseline subbasins (the three with positive KGE as mentioned above, as well as Little Arkansas with slightly negative baseline KGE), the contribution of state updating in general outweighs that of rainfall correction. Conversely, for the rest of the worse-baseline subbasins (Bird, Spring, Illinois and Deep, all with relatively large negative baseline KGE), the majority of the improvement is attributed to SMART rainfall correction scheme. To further understand the interrelationship between model baseline performance, rainfall forcing error and correction performance, we forced the VIC model by the NLDAS-2 benchmark rainfall (without state update) and assessed the subsequent streamflow improvement (listed in Table 3.5

"*NLDAS2-forced*" columns). This quantifies the maximum potential streamflow improvement achievable by rainfall correction. While almost all subbasins show an obvious streamflow improvement simply by switching to the NLDAS-2 rainfall forcing, the improvement magnitude has a strikingly clear negative correlation with basement performance. In particular, the four worse-baseline subbasins exhibit large streamflow improvement, with PER over 65% and KGE improvement that brings the negative baseline KGE to much closer to zero or positive. This result suggests that the worse baseline at these subbasins is largely caused by the bad IMERG rainfall forcing. While state update is still beneficial at these subbasins (through direct correction of random error of model structure/parameterization that reflects in SM states, and through correcting the part of rainfall error that is reflected in SM states), the SMART rainfall correction scheme is particularly necessary.

In contrast to the worse-baseline subbasins, the better-baseline subbasins demonstrate a reduced capability of streamflow improvement via rainfall correction, and therefore rely more on the state update scheme for enhancing streamflow. In particular, the best-baseline subbasin, Chikaskia, experiences degraded streamflow KGE and slightly positive PER when forced by the NLDAS-2 rainfall. This seems inconsistent with the slightly positive KGE and bigger PER contributed by SMART at this basin (Table 3.5). One possible reason is that at this subbasin NLDAS-2 benchmark rainfall is not obviously better than the IMERG baseline. Therefore, switching to NLDAS-2 rainfall forcing does not benefit streamflow, but SMART is still able to extract information from SMAP to slightly correct IMERG.

It is worth noting that the dual correction scheme presented in this study is designed to only correct the random error existing in the simulation system, but not the systematic error or overall bias. Figure 3.6 shows example time series of the baseline (prior to the ingestion of any SMART or state correction), USGS-observed and dual-corrected streamflow at three subbasins with various levels of baseline performance. It is clear from the time series that, although the dual system often nudges the simulated streamflow toward the correct direction (especially at high-flow periods) and results in overall improved evaluation statistics, obvious systematic error (in the model structure/parameterization as well as rainfall forcing) exists that drives the streamflow simulation away from the observed behavior. Such systematic error, although difficult to quantify exactly, cannot be corrected by the means introduced here. The NENSK statistic partly reflects such systematic error. NENSK is significantly above one at most

subbasins, indicating an under-dispersed ensemble on average. In other words, at most subbasins the ensemble spread created by the dual system only represents the random uncertainty around the open-loop baseline, but not the systematic error which accounts for much of the total error.

The level of systematic error is tied closely to the quality of hydrologic model parameters, often estimated through calibration. The VIC parameters used in this study were taken from Maurer et al. [2002] and derived at large-scale watershed outlets. To further examine the effect of systematic error on data assimilation, we calibrated the model parameters for the 8 subbasins using streamflow acquired from USGS (Table 3.3). Specifically, VIC parameters that control infiltration, soil conductivity and baseflow generation as well as the recession rate of the grid-cell-scale unit hydrograph in RVIC were calibrated using the MOCOM multi-objective autocalibration method [Yapo et al., 1998]. Basin-constant parameters were calibrated toward USGS streamflow time series during 2015 to 2017, forced by the baseline IMERG precipitation, to optimize daily KGE and monthly bias. Only a subset of the 8 subbasins were able to achieve better-than-baseline streamflow results via this traditional calibration method, mainly due to the large IMERG forcing error at some subbasins that makes the calibration scheme not able to find a reasonable parameterization. Here, we show three example subbasins with relatively good calibration outcome as demonstration (see Figure 3.7). The Chikaskia subbasin, whose baseline performance improves from the original daily KGE of 0.67 to 0.78 after calibration, experiences slightly decreased streamflow improvement compared to before calibration, with smaller PER (from 15.0% to 10.7%) and slightly smaller KGE improvement (from +0.07 to +0.01). Visually from the time series, the open-loop streamflow behavior is slightly improved through calibration (e.g., at the beginning of April 2017), and the difference in the correction amount before and after calibration is relatively small (comparing Figure 3.6a and Figure 3.7a). NENSK for Chikaskia goes up slightly away from one (i.e., more under-dispersed). The baseline simulation of the second subbasin, the Deep Fork, improves from KGE = -0.77 to 0.08, which is clear from the visual time series as well (comparing Figure 3.6b and Figure 3.7b). The dual correction with calibrated parameters shows slightly bigger PER and smaller KGE improvement than before. NENSK again becomes slightly worse (more under-dispersed) after calibration. The third basin, the Illinois River, again experiences slightly smaller PER and KGE improvement and worse NENSK after calibration (comparing Figure 3.6c and Figure 3.7c). These results suggest that a better model parameterization (and therefore better baseline simulation) typically results in

similar or smaller added value of dual correction. This means that the dual system is able to nudge streamflow relatively effectively (or slightly more effectively) even when model parameterization is poor. However, it is also clear that the portion of error corrected by the dual system is typically smaller than that achievable by a better model parameterization.

Table 3.5. Daily streamflow results from the dual correction system at the 8 USGS subbasins. In addition to the deterministic KGE improvement, PER and probabilistic NENSK results from the dual system ("*dual*" columns), the table also lists the open-loop streamflow KGE ("*open-loop KGE*" column), KGE improvement and PER as a result of state update or rainfall correction scheme alone ("*state update only*" and "*rainfall correction only*" columns, respectively), and KGE improvement and PER when forced by the NLDAS-2 benchmark precipitation without state update ("*NLDAS-2 forced*" column).

	Open-loop KGE	KGE improvement				PER				NENSK
		Dual	State update only	Rainfall correction only	NLDAS2-forced	Dual	State update only	Rainfall correction only	NLDAS2-forced	Dual
L Arkansas	-0.12	+0.17	+0.23	-0.01	+0.57	7.3%	10.8%	1.2%	40.0%	1.98
Ninnescah	0.25	+0.15	+0.06	+0.16	+0.20	14.0%	5.5%	13.7%	30.4%	0.35
Walnut	0.54	-0.02	-0.03	+0.03	-0.23	5.8%	5.7%	2.8%	23.3%	2.70
Chikaskia	0.67	+0.07	+0.05	+0.02	-0.45	15.0%	11.1%	6.6%	2.2%	1.96
Bird	-1.49	+0.95	+0.58	0.63	+0.95	33.5%	17.0%	25.8%	68.9%	2.01
Spring	-3.64	+0.83	+0.65	+0.33	+3.93	13.2%	8.7%	7.0%	83.4%	13.11
Illinois	-1.91	+0.50	+0.36	+0.26	+2.72	17.6%	7.4%	12.9%	81.8%	13.78
Deep	-0.77	+0.49	+0.39	+0.37	+1.55	20.8%	13.1%	21.2%	68.3%	2.34

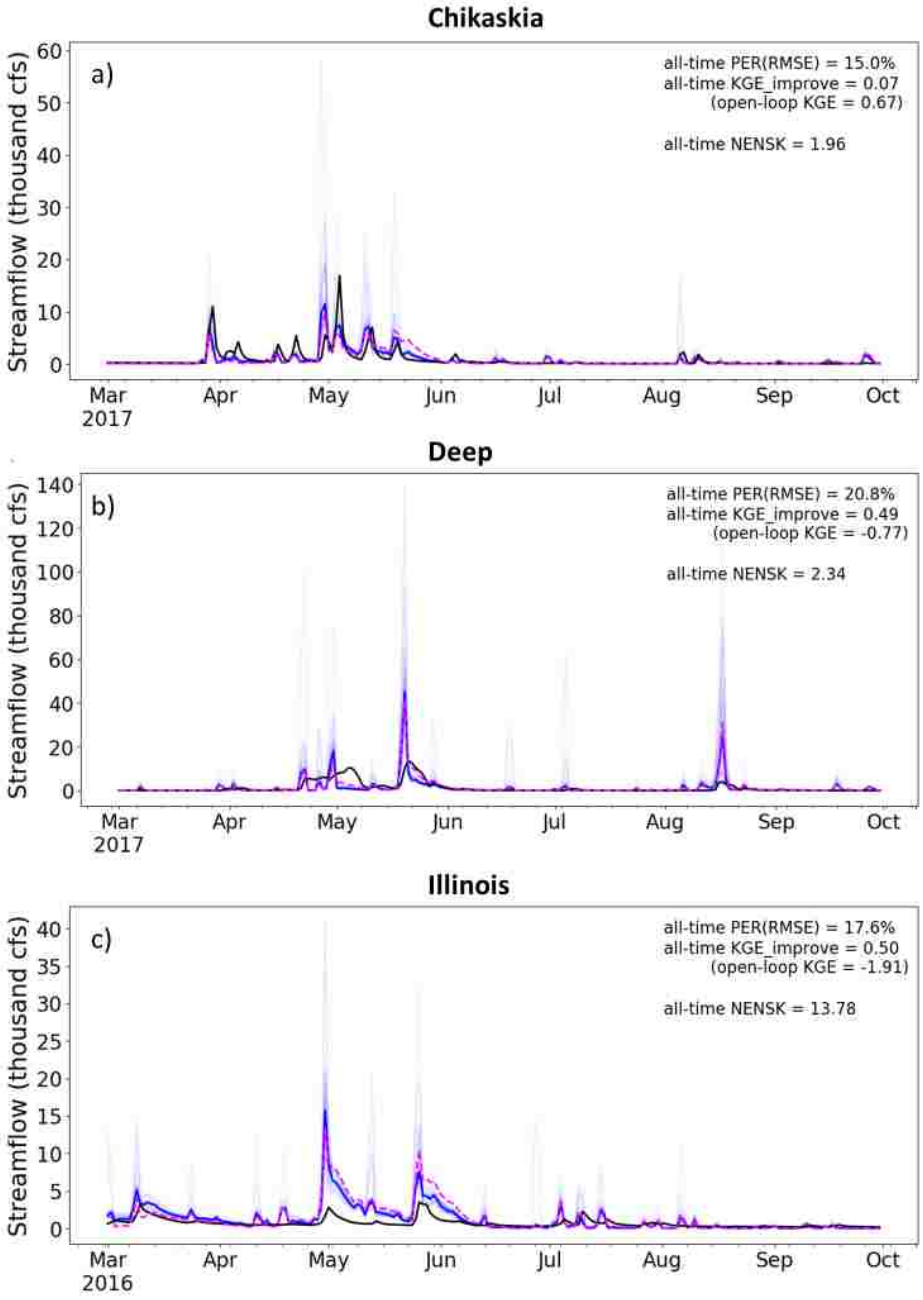

Figure 3.6. Example time series of streamflow results from the dual correction system. *Black line*: USGS observed streamflow; *magenta line*: baseline VIC simulation; *light blue lines*: ensemble updated streamflow results; *solid blue line*: ensemble-mean updated streamflow. Only part of the simulation period is shown for clear display. Statistics shown on each panel are based on the entire simulation period (approximately 2.5 years).

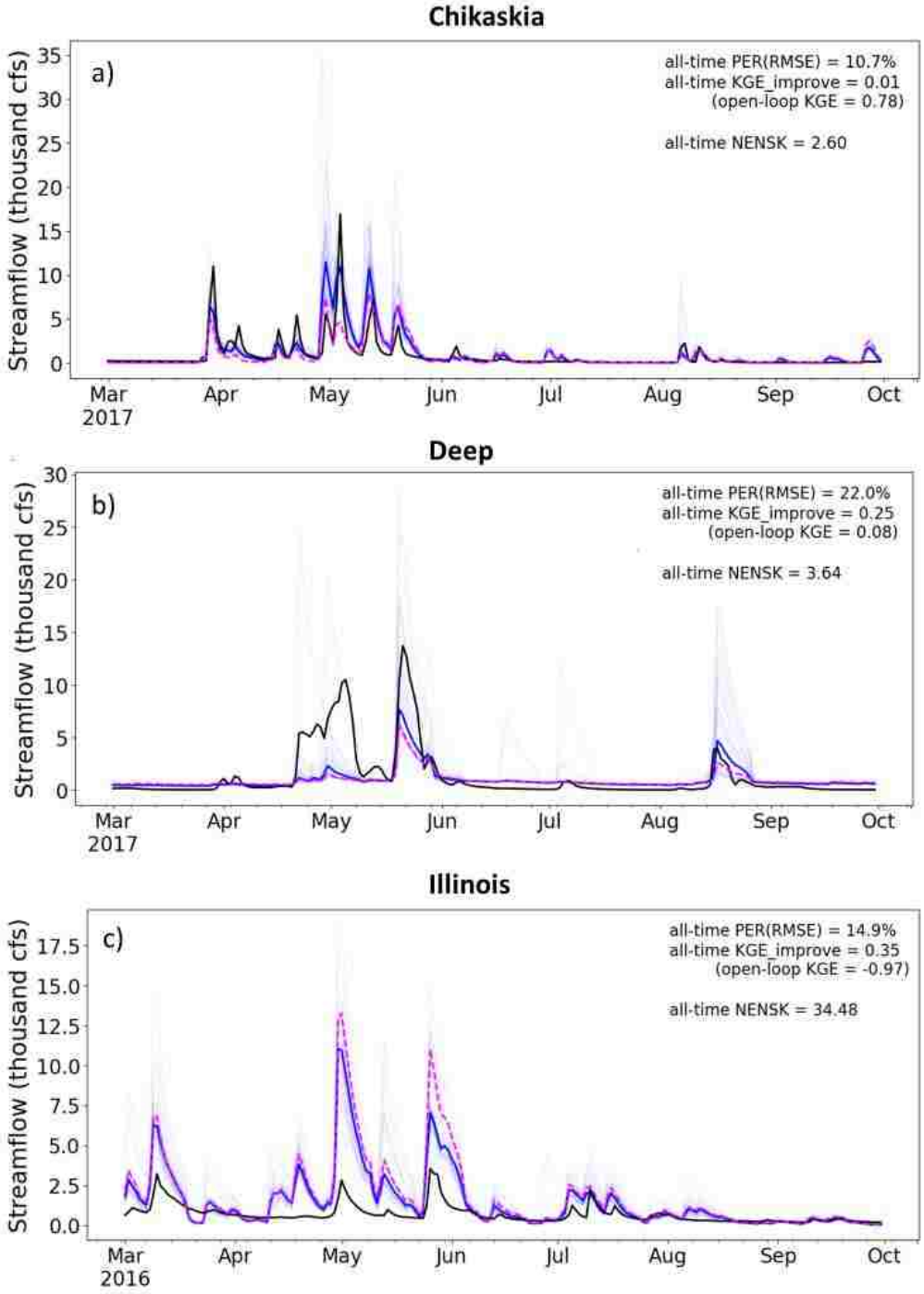

Figure 3.7. Time series of simulated baseline, corrected and observed streamflow at three example subbasin outlets with calibrated model parameters. All lines and notations are the same as in Figure 3.6.

3.4 DISCUSSION

With the VIC implementation of the dual correction system in this study, there is potential to correct streamflow at large-scale stream outlets. While the infrastructure for large-scale system has already been developed here with distributed hydrologic model and satellite data, only streamflow at small subbasins was evaluated due to the availability of unregulated streamflow observations. More research is needed to investigate the impact of error spatial correlation on downstream streamflow performance before applying the system to large-scale river outlets. Specifically, while the 1D filter with spatially white noise (as applied in this study) may be appropriate for small-basin correction, ignoring the spatial structure of errors could potentially have a more profound impact on the correction performance of large river outlets with a large number of grid cells routed and aggregated. Some past studies have investigated issues related to spatial patterns of error in hydrologic data assimilation. For example, Reichle and Koster [2003] investigated the impact of spatial error correlation in the model SM states on its assimilation performance; Gruber et al. [2015] examined the impact of a 2D filter with spatially auto-correlated error versus a 1D filter on SM updating quality; Pan et al. [2009] and Pan and Wood [2009; 2010] evaluated the surface SM assimilation performance with VIC by comparing a 1D filter, a 2D filter and a multiscale autoregressive filtering approach, as well as with precipitation ensemble generated with consideration of spatial and temporal uncertainty structure. However, these studies focus on the performance of SM simulations alone. Direct assessment of the impact of spatial error patterns on the routed streamflow assimilation results is needed. Also, evaluation of the probabilistic representation of ensemble streamflow at large river outlets is necessary, especially since the ignorance of spatial error patterns may potentially cause noises to be canceled out at large outlets, causing incorrect probabilistic streamflow estimates.

We would also like to discuss a few takeaways based on the findings from this study. We have shown that the dual correction approach is able to correctly nudge streamflow simulation a little, especially during relatively high flow events in area with poor IMERG data. However, the magnitude of correction is generally small. There are two main reasons for the small correction we see. First, the latest generation of satellite rainfall products (e.g., IMERG) has significantly improved quality compared to its predecessors, which makes it a much more difficult task to correct rainfall using SM measurements and therefore limits the level of rainfall correction.

Second, the dual approach, or any other Kalman-filter based data assimilation approaches, is designed to only correct the zero-mean random error component in the total streamflow error, but not the systematic error or bias. But we have shown that systematic error, typically originated from inaccurate model structure and/or parameterization and large rainfall bias, can account for a big portion of the total error. The existence of systematic error is particularly problematic from a probabilistic perspective, since the ensemble streamflow produced by the dual system only represents random error, therefore largely underestimates simulation uncertainty.

Given the above considerations, we may be approaching the limit of using data assimilation techniques to correct random streamflow errors. This insight provides a few guidelines for future research focus that aims to further improve streamflow simulation systems:

1) More sophisticated data assimilation techniques, though may help a little, is not very likely to substantially reduce streamflow error further, since they only deal with the random error components in the system which in many cases only account for a relatively small portion of the total error;

2) Instead, approaches to reduce systematic errors in streamflow simulation are needed. To date this is still a challenging task in large-scale hydrologic modeling, since calibration is difficult to perform with limited streamflow data and large number of distributed parameters. With the availability of the globally observed, distributed and near-real-time SMAP and IMERG products, more creative methods need to be developed to extract useful information from the big volume of observed data. For example, characteristics of SM dynamics and its response to rainfall can be directly extracted from the datasets themselves, which can potentially inform hydrologic model structure/parameterization. These areas of research are less studied but has the potential to improve hydrologic modeling beyond correcting random errors;

3) it is worthwhile to continue to develop future generation of higher-quality, near-real-time rainfall products, since rainfall plays a very important role in streamflow simulations and its error is not easily and substantially reduced by the current correction methods that use SM measurement information.

3.5 CONCLUSION

In this paper, we applied a dual state/rainfall correction data assimilation system in the Arkansas-Red River basin. Built upon the dual system developed in past studies, we have made several methodological advances. First, we implemented the dual correction system with a more complexed, semi-distributed land surface model, the VIC model, and applied it in a regional-scale basin. Second, the latest satellite products, the SMAP SM product and the IMERG rainfall product, were incorporated into the system. Third, the existing dual correction algorithm was updated to better suit the new satellite data, and also to produce an ensemble streamflow product that gives probabilistic estimates.

Our results show that, overall, IMERG rainfall (evaluated toward the gauge-informed NLDAS-2 benchmark over the entire Arkansas-Red domain) and streamflow (evaluated toward 8 USGS streamflow sites) were improved to some extent but not substantially via dual correction. For rainfall, the slight-to-moderate improvement is primarily from the correction of larger events via SMART, while smaller event estimates are slightly degraded. Rainfall correction is more effective at daily to multi-daily time scale than at 3-hourly scale. The ensemble produced by the correction scheme represents the rainfall uncertainty relatively well at daily to multi-daily scale. For streamflow, the dual correction is able to reduce the random errors in simulated streamflow across the 8 test subbasins, ranging from slight to moderate error reduction. The subset of subbasins with worse baseline performance, mainly due to worse IMERG rainfall quality, exhibit larger relative improvement via dual correction, and the correction is mainly contributed by the SMART rainfall correction scheme. The other subbasins with relatively better IMERG and baseline streamflow performance show less improvement via dual correction, and the improvement is more primarily attributed to state updating. The ensemble streamflow produced by the system largely underestimates the error uncertainty, because the dual system only accounts for the random error components in the system, but not the systematic error (e.g., resulted from incorrect model structure and parameterization). We have also demonstrated that model parameterization error that commonly exists in large-scale distributed models in general does not degrade (and sometimes actually benefits) the performance of the dual correction scheme.

Based on the findings from this study, we might be reaching the limit of using SM data assimilation techniques to reduce random errors in simulated streamflow. More sophisticated data assimilation techniques may improve the correction results a little further, but more substantial streamflow reduction would require future research efforts on reducing systematic modeling error via, e.g., innovative ways of achieving better model representation, as well as higher-quality satellite rainfall products.

Acknowledgements

This work was supported in part by NASA Terrestrial Hydrology Program Award NNX16AC50G to the University of Washington and NASA Terrestrial Hydrology Program Award 13-THP13-0022 to the United States Department of Agriculture, Agricultural Research Service. Yixin Mao also received a Pathfinder Fellowship by CUAHSI with support from the National Science Foundation (NSF) Cooperative Agreement No. EAR- 1338606. We would also like to thank Andrew Wood from NCAR for help on calibration. The VIC model used in the study can be found at https://github.com/UW-Hydro/VIC. Specifically, we used VIC version 5.0.1 (doi:10.5281/zenodo.267178) with a modification to the calculation of drainage between soil layers (https://github.com/UW-Hydro/VIC/releases/tag/Mao_etal_stateDA_May2018). The DA code used in this study is available at https://github.com/UW-Hydro/dual_DA_SMAP.

REFERENCES

Alvarez-Garreton, C., D. Ryu, A. W. Western, W. T. Crow, and D. E. Robertson (2014), The impacts of assimilating satellite soil moisture into a rainfall-runoff model in a semi-arid catchment, *J. Hydrol.*, 519, 2763-2774, doi:10.1016/j.jhydrol.2014.07.041.

Alvarez-Garreton, C., D. Ryu, A. W. Western, W. T. Crow, C.-H. Su, and D. R. Robertson (2016), Dual assimilation of satellite soil moisture to improve streamflow prediction in data-scarce catchments, *Water Resour. Res.*, 52(7), 5357-5375, doi:10.1002/2015WR018429.

Aubert, D., C. Loumagne, and L. Oudin (2003), Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall-runoff model, *J. Hydrol.*, 280(1-4), 145-161, doi:10.1016/S0022-1694(03)00229-4.

Brocca, L., F. Melone, T. Moramarco, W. Wagner, V. Naeimi, Z. Bartalis, and S. Hasenauer (2010), Improving runoff prediction through the assimilation of the ASCAT soil moisture product, *Hydrol. Earth Syst. Sci.*, 14, 1881-1893, doi:10.5194/hess-14-1881-2010.

Brocca, L., T. Moramarco, F. Melone, W. Wagner, S. Hasenauer, and S. Hahn (2012), Assimilation of surface-and root-zone ASCAT soil moisture products into rainfall–runoff modeling, *IEEE Trans. Geosci. Remote Sens.*, 50(7), 2542-2555, doi: 10.1109/TGRS.2011.2177468.

Brocca, L., T. Moramarco, F. Melone, and W. Wagner (2013), A new method for rainfall estimation through soil moisture observations, *Geophys. Res. Lett.*, 40, 853–858, doi:10.1002/grl.50173.

Brocca, L., L. Ciabatta, C. Massari, T. Moramarco, S. Hahn, S. Hasenauer, R. Kidd, W. Dorigo, W. Wagner, and V. Levizzani (2014), Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data, *J. Geophys. Res. Atmos.*, 119, 5128–5141, doi:10.1002/2014JD021489.

Brocca, L., T. Pellarin, W. T. Crow, L. Ciabatta, C. Massari, D. Ryu, C.-H. Su, C. Rüdiger, and Y. Kerr (2016), Rainfall estimation by inverting SMOS soil moisture estimates: A comparison of different methods over Australia, *J. Geophys. Res. Atmos.*, 121, 12,062–12,079, doi:10.1002/2016JD025382.

Chan, S. et al. (2017), Development and validation of the SMAP enhanced passive soil moisture product, Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International, doi: 10.1109/IGARSS.2017.8127512.

Chen F., W. T. Crow, and T. R. H. Holmes (2012), Improving long-term, retrospective precipitation datasets using satellite-based surface soil moisture retrievals and the Soil Moisture Analysis Rainfall Tool, *J. Appl. Remote Sens.*, 6(1), 063604, doi: 10.1117/1.JRS.6.063604.

Chen, F., W. T. Crow, and D. Ryu (2014), Dual forcing and state correction via soil moisture assimilation for improved rainfall–runoff modeling, *J. Hydrometeorol.*, *15*(5), 1832–1848, doi:10.1175/JHM-D-14-0002.1.

Colliander, A. et al. (2017), Validation of SMAP surface soil moisture products with core validation sites, *Remote Sens. Environ.*, 191, 215-231, doi:10.1016/j.rse.2017.01.021.

Crow, W. T., and J. D. Bolten (2007), Estimating precipitation errors using spaceborne surface soil moisture retrievals, *Geophys. Res. Lett.*, 34, L08403, doi:10.1029/2007GL029450.

Crow, W. T., and D. Ryu (2009), A new data assimilation approach for improving hydrologic prediction using remotely-sensed soil moisture retrievals, *Hydrol. Earth Syst. Sci.*, 12(1-16), doi:10.5194/hess-13-1-2009.

Crow W. T., G. J. Huffman, R. Bindlish, and T. J. Jackson (2009), Improving satellite-based rainfall accumulation estimates using spaceborne surface soil moisture retrievals, *J. Hydrometeorol.*, 10, 199-212, doi: 10.1175/2008JHM986.1.

Crow, W. T., M. J. van den Berg, G. J. Huffman, and T. Pellarin (2011), Correcting rainfall using satellite-based surface soil moisture retrievals: The Soil Moisture Analysis Rainfall Tool (SMART), *Water Resour. Res.*, 47, W08521, doi:10.1029/ 2011WR010576.

Crow, W. T., F. Chen, R. H. Reichle, and Q. Liu (2017), L band microwave remote sensing and land data assimilation improve the representation of prestorm soil moisture conditions for hydrologic forecasting, *Geophys. Res. Lett.*, 44, 5495-5503, doi:10.1002/2017GL073642.

De Lannoy, G. J. M., P. R. Houser, V. R. N. Pauwels, and N. E. C. Verhoest (2006), Assessment of model uncertainty for soil moisture through ensemble verification, *J. Geophys. Res.*, 111, D10101, doi: 10.1029/2005JD006367.

Entekhabi et al. (2010), The Soil Moisture Active and Passive (SMAP) Mission, *Proceedings of the IEEE*, 98(5), 704-716, doi:10.1109/JPROC.2010.2043918.

Francois, C., Quesney, A., and C. Ottle (2003), Sequential assimilation of ERS-1 SAR data into a coupled land surface-hydrological model using an extended Kalman filter, *J. Hydrometeorol.*, 4(2), 473-487, doi:10.1175/1525-7541(2003)4<473:SAOESD>2.0.CO;2.

Freeze, R. A., and R. L. Harlan (1969), Blueprint for a physically-based, digitally-simulated hydrologic response model, *J. Hydrol.*, 9(3), 237-258, doi:10.1016/0022-1694(69)90020-1.

Gebregiorgis, A. S., P.-E. Kirstetter, Y. E. Hong, J. J. Gourley, G. J. Huffman, W. A. Petersen, X. Xue, and M. R. Schwaller (2018), To what extent is the day 1 GPM IMERG satellite precipitation estimate improved as compared to TRMM TMPA-RT?, *J. Geophys. Res. Atmos.*, 123, 1694–1707, doi:10.1002/2017JD027606.

Gruber, A., W. T. Crow, W. Dorigo, and W. Wagner (2015), The potential of 2D Kalman filtering for soil moisture data assimilation, *Remote Sens. Environ.*, 171, 137-148, doi: 10.1016/j.rse.2015.10.019.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez (2009), Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, 377, 80-91, doi: 10.1016/j.jhydrol.2009.08.003.

Hamman, J., B. Nijssen, A. Roberts, A. Craig, W. Maslowski, and R. Osinski (2017), The coastal streamflow flux in the Regional Arctic System Model, *J. Geophys. Res.*, 122(3), 1683-1701, doi: 10.1002/2016JC012323.

Hamman, J. J., B. Nijssen, T. J. Bohn, D. R. Gergel, and Y. Mao (2018), The Variable Infiltration Capacity Model, Version 5 (VIC-5): Infrastructure improvements for new applications and reproducibility, *Geosci. Model Dev.*, 11, 3481-3496, doi: 10.5194/gmd-11-3481-2018.

Hou, A. Y., R. K. Kakar, S. Neeck, A. A. Azarbarzin, C. D. Kummerow, M. Kojima, R. Oki, K. Nakamura, and T. Iguchi (2014), The Global Precipitation Measurement mission, *Bull. Amer. Meteor. Soc., 95(5), 701-722, doi: 10.1175/BAMS-D-13-00164.1.*

Huffman, G. J., D. T. Bolvin, and E. J. Nelkin (2015), Integrated Multi-Satellite Retrievals for GPM (IMERG) Technical Documentation. Tech. Doc., NASA GSFC. [Available online at https://docserver.gesdisc.eosdis.nasa.gov/public/project/GPM/IMERG_doc.05.pdf.]

Huffman, G. J., E. F. Stocker, D. T. Bolvin, and E. J. Nelkin (2018), last updated 2018: IMERG L3 Early Run Data Sets. NASA/GSFC, Greenbelt, MD, USA, Accessed 2018-08-29, https://gpm1.gesdisc.eosdis.nasa.gov/opendap/hyrax/GPM_L3/GPM_3IMERGHHL.05/.

Koster, R. D., L. Brocca, W. T. Crow, M. S. Burgin, and G. J. M. De Lannoy (2016), Precipitation estimation using L-band and C-band soil moisture retrievals, *Water Resour. Res.*, 52, 7213–7225, doi:10.1002/2016WR019024.

Liang, X., D. P. Lettenmaier, E. F. Wood, and 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.

Lievens, H., et al. (2015), SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia, *Remote Sens. Environ.*, *168*, 146-162, doi:10.1016/j.rse.2015.06.025.

Lievens, H., G. J. M. De Lannoy, A. Al Bitar, M. Drusch, G. Dumedah, H.-J. Hendricks Franssen, Y. H. Kerr, S. K. Tomer, B. Martens, O. Merlin, M. Pan, J. K. Roundy, H. Vereecken, and J. P. Walker (2016), Assimilation of SMOS soil moisture and brightness temperature products into a land surface model, *Remote Sens. Environ.*, 180, 292-304, doi: 10.1016/j.rse.2015.10.033.

Lohmann, D., R. Nolte-Holube, and E. Raschke (1996), A large-scale horizontal routing model to be coupled to land surface parametrization schemes, Tellus, 48(A), 708-721, doi: 10.1034/j.1600-0870.1996.t01-3-00009.x.

Lohmann, D., E. Raschke, B. Nijssen, and D. P. Lettenmaier (1998), Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model, *Hydrol. Sci. J.*, 43(1), 131-141, doi: 10.1080/02626669809492107.

Mao Y., W. T. Crow, and B. Nijssen (2018), A framework for diagnosing factors degrading the streamflow performance of a soil moisture data assimilation system, *J. Hydrometeorol.*, in review.

Massari, C., L. Brocca, A. Tarpanelli, and T. Moramarco (2015), Data Assimilation of Satellite Soil Moisture into Rainfall-Runoff Modelling: A Complex Recipe?, *Remote Sens.*, 7, 11403-11433, doi:10.3390/rs70911403.

Massari, C., S. Camici, L. Ciabatta, and L. Brocca (2018), Exploiting satellite-based surface soil moisture for flood forecasting in the Mediterranean area: State update versus rainfall correction, *Remote Sens.*, 10, 292, doi: 10.3390/rs10020292.

Maurer, E.P., A.W. Wood, J.C. Adam, D.P. Lettenmaier, and B. Nijssen (2002), A long-term hydrologically-based data set of land surface fluxes and states for the conterminous United States, *J. Clim.*, 15(22), 3237-3251, doi:10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2.

Mehra, R. K. (1971), On-line identification of linear dynamic systems with applications to Kalman filtering, *IEEE Trans. Autom. Control.*, 16(1), 12-21, doi:10.1109/TAC.1971.1099621.

O'Neill, P. E., S. Chan, E. G. Njoku, T. Jackson, and R. Bindlish (2016), SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Version 4, Boulder, Colorado USA, NASA National Snow and Ice Data Center Distributed Active Archive Center, Accessed 2018-01-18, doi:10.5067/OBBHQ5W22HME.

Pan, M., E. F. Wood, D. B. McLaughlin, and D. Entekhabi (2009), A multiscale ensemble filtering system for hydrologic data assimilation. Part I: Implementation and synthetic experiment, *J. Hydrometeorol.*, 10, 794-806, doi:0.1175/2009JHM1088.1.

Pan, M., and E. F. Wood (2009), A multiscale ensemble filtering system for hydrologic data assimilation. Part II: Application to land surface modeling with satellite rainfall forcing, *J. Hydrometeorol.*, 10, 1493-1506, doi:10.1175/2009JHM1155.1.

Pan, M., and E. F. Wood (2010), Impact of accuracy, spatial availability, and revisit time of satellite-derived surface soil moisture in a multiscale ensemble data assimilation system, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 3 (1), 49-56, doi:10.1109/JSTARS.2010.2040585.

Reichle, R. H., and R. D. Koster (2003), Assessing the impact of horizontal error correlations in background fields on soil moisture estimation, *J. Hydrometeorol.*, 4 (6), 1229-1242, doi: 10.1175/1525-7541(2003)004<1229:ATIOHE>2.0.CO;2.

Talagrand, O., R. Vautard, and B. Strauss (1997), Evaluation of probabilistic prediction systems, technical report, Eur. Cent. for Medium-Range Weather Forecast., Reading, UK.

United States Geological Survey (USGS) (2018), USGS Surface-water daily data for the nation. [Available at https://waterdata.usgs.gov/nwis/dv/?referred_module=sw.]

Wanders, N., D. Karssenberg, A. De Roo, S. M. De Jong, and M. F. P. Bierkens (2014), The suitability of remotely sensed soil moisture for improving operational flood forecasting, *Hydrol. Earth Syst. Sci.*, *18*(6), 2343-2357, doi:10.5194/hess-18-2343-2014.

Western, A. W., R. B. Grayson, and G. Blöschl (2002), Scaling of soil moisture: a hydrologic perspective, *Annu. Rev. Earth Planet. Sci.*, *30*(1), 149-180, doi:10.1146/annurev.earth.30.091201.140434.

Wilks, D. S. (2011), Statistical methods in the atmospheric sciences (3rd edition), Elsevier/Academic Press, Amsterdam; Boston.

Xia, Y. et al., NCEP/EMC (2009), NLDAS Primary Forcing Data L4 Hourly 0.125 x 0.125 degree V002, Edited by David Mocko, NASA/GSFC/HSL, Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed 2018-02-27, doi:10.5067/6J5LHHOHZHN4.

Yapo, P. O., H. V. Gupta, and S. Sorrooshian (1998), Multi-objective global optimization for hydrologic models, *J. Hydrol*. 2014, 83-97, doi: 10.1016/S0022-1694(97)00107-8.

Chapter 4. A UNIFIED DATA-DRIVEN METHOD TO DERIVE HYDROLOGIC DYNAMICS FROM GLOBAL SMAP SURFACE SOIL MOISTURE AND GPM PRECIPITATION DATA

This chapter is to be submitted to *Water Resources Research* as

Mao Y., W. T. Crow, and B. Nijssen (2018), A unified data-driven method to derive hydrologic dynamics from global SMAP surface soil moisture and GPM precipitation data, *Water Resources Research*, manuscript in preparation.

The supplemental material for this chapter is provided in appendix C.

Abstract

The new satellite-observed Soil Moisture Active Passive (SMAP) and the Global Precipitation Measurement (GPM) datasets contain rich information about land surface hydrologic processes. In this study, a unified regression method is proposed and applied to these global datasets to quantify factors governing surface soil moisture (SSM) dynamics. Two simple forms of regressors are implemented: 1) the linear regressors of SSM itself and precipitation input, and 2) the two linear regressors with an additional interaction term. Results of the coefficients fitted on the 3-year global SMAP and GPM data show that the unified regression method is able to reproduce or mimic the SSM characteristics found by several recent studies, including the SSM exponential decay rate, the fraction of precipitation retained in surface soil layer, and the effective depth of hydrologic storage. Additionally, including the interaction regressor provides an innovative approach to derive the sensitivity of infiltration/runoff partition process to antecedent SSM level without the need for streamflow observation data. The regression results from the satellite data are also compared with a model-based global dataset, providing insight into the suitability of model structure and parameterization. Compared to the satellite data, the physically-based model retains moisture longer in the top layer, shows less sensitivity of the runoff/infiltration partition process to the top-layer soil moisture, and exhibits less spatial variation in SSM dynamics. This study demonstrates that data-driven methods are

capable of deriving hydrologic dynamics purely from the SMAP data and that these derived dynamics are useful in evaluating process representations in hydrologic models.

4.1 INTRODUCTION

A new generation of earth observing satellites have been developed and become available in recent years. Particularly relevant to hydrologic science, the NASA Soil Moisture Active Passive (SMAP) mission [Entekhabi et al., 2010] and the Global Precipitation Measurement (GPM) mission [Hou et al., 2014] provide land surface soil moisture (SSM) and precipitation observations, respectively, with near global coverage. These satellite datasets have presumably higher data quality than their predecessors and contain rich information about land surface hydrologic processes. For example, the SSM evolution during no-precipitation periods is primarily a result of SSM drying processes, mainly controlled by evaporation and drainage to deeper layers. Also, the SSM response to precipitation events reflects the magnitude of land surface processes including canopy interception of rainfall and surface runoff.

Since the launch of GPM in 2014 and SMAP in 2015, multiple years of their retrievals have become available. There has been an emerging number of new studies that have attempted to extract hydrologic information from SMAP at continental to global scale, some of them combined with GPM. For example, McColl et al. [2017a] calculated global SSM distribution and SSM memory, defined as the fraction of precipitation input flux that enters the SSM storage and remains there after a time interval. McColl et al. [2017b] extracted SSM drydown periods and calculated an average SSM decay e-folding time scale. Koster et al. [2017] estimated a non-linear SSM loss function, defined as the SSM decay rate as a function of SSM level itself, and applied the loss function to perform real-time SSM nowcasting and forecasting. Akbar et al. [2018a] also estimated a non-linear SSM loss function from SMAP, and further classified the shape of the loss function across the Continental United States (CONUS). Shellito et al. [2016] and Shellito et al. [2018] calculated SSM drydown rate from SMAP, and compared it with in-situ measurements and model simulated SSM, respectively. Akbar et al. [2018b] estimated an effective storage depth from SMAP. Koster et al. [2018] inferred rainfall from SMAP and further estimated basin-scale streamflow with regression analysis on streamflow observation data. These

recent studies have demonstrated the potential for improving our process-level understanding of hydrologic dynamics via straightforward data-driven analysis of SMAP and GPM data.

One shortcoming of these studies is that they typically selected a certain hydrologic characteristic (e.g., SSM drydown or precipitation fraction retained) and a pre-assumed way of quantifying it (e.g., SSM exponential decay time scale; a multi-stage SSM loss function). While these approaches are able to extract a single fixed aspect of hydrologic information contained in SMAP, they require *a priori* assumptions about the SSM evolution or response to precipitation inputs, and they usually make subjective decisions in pre-processing the SMAP data (e.g., extracting SSM decay periods; separating rain versus no-rain periods). While these *a priori* assumptions result in directly interpretable results, they also prevent the potential opportunity to discover a full range of more complex, often interrelated factors that govern the land surface hydrologic dynamics. To overcome this shortcoming, we propose a unified regression method to the global SMAP and GPM data with the goal of deriving the dominating factors that govern SSM dynamics. We aim to place our results in the context of recent SMAP studies (mentioned above) and to provide new insights into the interaction between factors governing SSM dynamics. We focus on simple linear regressors of SSM itself and precipitation as well as their interaction, but the proposed regression method itself is very flexible and can be expanded to include any other form of governing terms such as nonlinear or categorical variables as well as additional data sources (e.g., air temperature).

In addition to providing a greater knowledge of land surface processes, deriving the governing hydrologic dynamics from satellite data is also potentially useful for informing large-scale hydrologic model development. Recent studies that assimilated SSM from SMAP to update modeled states [e.g., Mao et al., 2018a, 2018b] pointed out that data assimilation techniques, as a commonly used approach to combine modeled and observed information, only correct random errors and are insufficient to substantially improve modeling accuracy of hydrologic variables. Instead, improving structural representation of hydrologic processes in models is necessary. Since dominant hydrologic processes may be drastically different at various spatial scales, satellite SM data provides direct hydrologic information at a coarse resolution that is commensurate with the resolution of continental-scale distributed hydrologic models. This makes it possible to use the hydrologic dynamics derived from SMAP and GPM to evaluate and enhance the structure and parameterization of large-scale hydrologic models. A recent study by

Shellito et al. [2018] provided an excellent example on this topic, where they compared the SSM drying rates observed by SMAP and simulated by the Noah land surface model and discussed their different sensitivity to factors such as potential evaporation, vegetation and soil texture. In this study, we generalize the SMAP-versus-model comparison to the results of our regression model.

The following sections are organized as follows. Section 4.2 describes the details of data used, data quality control and the analysis method. Results derived from the satellite observations and model simulations are shown in Section 4.3. Section 4.4 discusses our results in the context of past literature and highlights our contributions. Section 4.5 concludes with our findings.

4.2 METHODS

4.2.1 Data and quality control

4.2.1.1. SMAP satellite SSM data

The SMAP mission was launched in March 31, 2015 and provides SSM in the top 5 cm of soil column. In this study, we combined the ascending (PM) and descending (AM) retrievals from the SMAP L3 Passive product [O'Neill et al., 2016] (data Version 4), which has a spatial resolution of 36 km and a typical revisit time ranging from 12 hours (with ascending and descending retrievals available in a single day) to 2-3 days, depending on location and time of year. Three full years of SMAP data, from April 2015 to March 2018, were used in our analysis.

A few quality control steps were applied to the raw SMAP retrievals prior to analysis. First, a few SMAP pixels with constant retrieval values, which indicates obvious quality flaw, were masked out. Then the internal quality flags provided by the SMAP mission were applied and only data points with “recommended” quality were retained. Figure 4.1 shows a map of the number of SMAP retrievals after applying quality control. Regions are masked out by the quality flag including Amazon, boreal forest regions in North America and Eurasia, eastern United States, southeastern China, and part of central Africa, western Europe, Southeast Asia, Middle East and Australia. The SMAP internal quality flags are relatively conservative and mask out a relatively large number of retrievals. Nevertheless, we applied these flags to ensure high quality

of data used in our analysis. Finally to mitigate systematic error between the AM and PM retrievals, we mapped the ascending (PM) data to the descending (AM) regime via a seasonal cumulative distribution function (CDF) matching approach, since AM retrievals are generally more accurate than PM retrievals [Chan et al., 2018].

The SMAP data were directly analyzed at their native 36 km spatial resolution. The original SMAP volumetric moisture content [m^3/m^3] was converted to a moisture depth [mm] by multiplying the nominal SMAP detecting depth of 50 mm. While the unit conversion does not affect the conclusions from the regression analysis, it facilitates interpretation of the results.

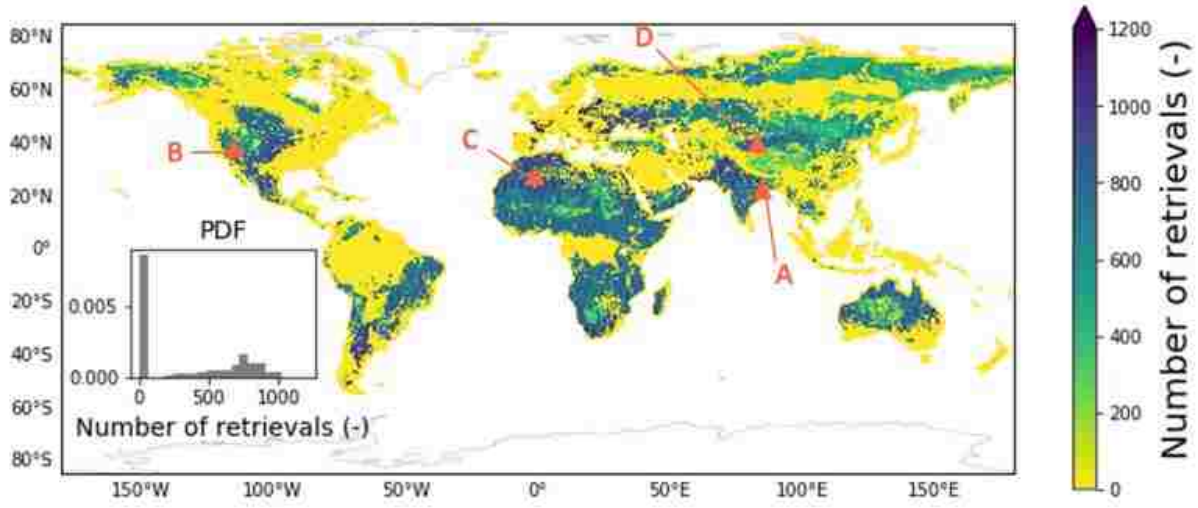

Figure 4.1. Total number of SMAP retrievals during April 2015 to March 2018 after quality control. Four example locations where time series are presented are labeled on the map (*A*: eastern India; *B*: western United States; *C*: western Sahara; *D*: Tarim Basin in western China).

4.2.1.2. GPM satellite precipitation data

The Integrated Multi-satellitE Retrievals for GPM (IMERG) Level 3 Version 05 Final Run precipitation data was used [Hoffman, 2017], which combines various satellite observations and provides 30-minute cumulative precipitation products at 0.1º spatial resolution [Huffman et al., 2015]. The Final Run product used in this study further includes monthly gauge analysis. We regridded the original 0.1º precipitation data to the 36 km SMAP resolution. We further converted the IMERG data from UTC to local solar time (LST) to match the SMAP time zone.

IMERG data at high-latitude Arctic regions (latitude > 60º) are only available during relatively short warm seasons and these pixels were excluded from our analysis. We also excluded pixels where less than 10% of the timesteps of the IMERG precipitation (12-hour accumulated) is nonzero, which are mainly located in the eastern Sahara Desert and the Arabian Peninsula. These pixels were excluded from our analysis because the infrequent precipitation leads to a low signal-to-noise ratio which often results in unstable regression fitting (see Section 4.2.2 for further discussion). These pixels also correspond to regions with very arid sandy soils where SMAP presumably has low accuracy.

4.2.1.3. Modeled SSM and precipitation data

The Global Land Data Assimilation System (GLDAS) gridded precipitation and simulated SSM data products by the Variable Infiltration Capacity (VIC) land surface model [Rodell et al., 2004] were obtained and analyzed in this study. Both precipitation and SSM data from GLDAS-VIC are at 1º spatial resolution and 3-hourly timestep. Note that the GLDAS-VIC SSM variable [mm] is the simulated soil moisture in the top 10-cm of the soils, about twice the depth of the SMAP observations. GLDAS-VIC SSM was temporally resampled to have the same time-varying data intervals as SMAP to eliminate any impact of sampling frequency on analysis results during model-observation comparison. High-latitude Arctic regions and some pixels in Sahara and Arabian Peninsula were also excluded from the regression analysis for GLDAS-VIC as for IMERG to be consistent and comparable with the SMAP/IMERG analysis. No spatial regridding was applied to the GLDAS data and the regression analysis was conducted at its native 1º resolution.

4.2.1.4. Climatic aridity index data

The distribution of the regression results over different aridity regions was examined in this study. Long-term-averaged climatology precipitation and air temperature data was obtained and calculated from GLDAS-VIC over 1979 to 2017. Potential evapotranspiration (PET) was then estimated for each pixel location by a simple temperature-based method [Malmstrom, 1969; Dingman, 2002]:

$$PET_m = 40.9 \cdot e_a^* \tag{4.1}$$

where PET_m [mm/month] is the estimated monthly climatic *PET*; e_a^* [kPa] is saturated vapor pressure for each month of year, which is calculated from the GLDAS monthly climatic air temperature, T [ºC], via [Dingman, 2002]:

$$e_a^* = 0.611 \cdot \exp(17.3 \cdot \frac{T}{237.3+T}) \quad (4.2).$$

Finally, the aridity index, $\frac{PET}{P}$ [-], was calculated for each pixel location, where *PET* [mm/year] is the annual average *PET*, and *P* is the GLDAS long-term-averaged annual precipitation.

4.2.2 Regression analysis for SSM dynamics

The goal of the regression analysis is to quantify empirically the governing processes controlling the SSM dynamics. In our case, SSM dynamics at each pixel location are represented by the SSM changing rate in a discrete form:

$$\frac{dSSM}{dt} \approx \frac{\Delta SSM}{\Delta t} \quad (4.3)$$

where *dt* and Δt [day] are the derivative and discrete form of timestep, respectively; *dSSM* and ΔSSM [mm] are derivative and discrete form of SSM change, respectively, within the timestep. $\frac{\Delta SSM}{\Delta t}$ is the dependent variable (i.e., regression target) in the regression. Given the available data (i.e., SSM and precipitation), the simplest independent variables for a multivariate regression are the linear forms of the SSM state variable itself and the precipitation flux, in which case the regression problem can be written as:

Regression Form I: $$\frac{\mathbf{\Delta SSM}}{\mathbf{\Delta t}} = \beta_1 \cdot \mathbf{SSM} + \beta_2 \cdot \frac{\mathbf{P}}{\mathbf{\Delta t}} + \boldsymbol{\varepsilon} \quad (4.4)$$

where *P* [mm] is the cumulative precipitation flux input during Δt; ε [mm/day] is the regression residual; β_1 [day^{-1}] and β_2 [-] are the coefficients to be fitted. The bold terms indicate a column vector that contains all the available data points constructed from a time series. Specifically, Equation (4.4) can be written in a more explicit format as:

$$\begin{bmatrix} \frac{SSM_1 - SSM_0}{t_1 - t_0} \\ \vdots \\ \frac{SSM_k - SSM_{k-1}}{t_k - t_{k-1}} \\ \vdots \\ \frac{SSM_n - SSM_{n-1}}{t_n - t_{n-1}} \end{bmatrix} = \beta_1 \cdot \begin{bmatrix} SSM_0 \\ \vdots \\ SSM_{k-1} \\ \vdots \\ SSM_{n-1} \end{bmatrix} + \beta_2 \cdot \begin{bmatrix} \frac{P_1}{t_1 - t_0} \\ \vdots \\ \frac{P_k}{t_k - t_{k-1}} \\ \vdots \\ \frac{P_n}{t_n - t_{n-1}} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_k \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (4.5)$$

where the subscript *k* is the timestep index, and *n* is the total number of data points. In other words, each data point in the regression is formulated as:

$$\frac{SSM_k - SSM_{k-1}}{t_k - t_{k-1}} = \beta_1 \cdot SSM_{k-1} + \beta_2 \cdot \frac{P_k}{t_k - t_{k-1}} + \varepsilon_k \quad (4.6)$$

Here we used SSM at the beginning of a timestep as the independent variable and $P_k = \sum_{t=t_{k-1}}^{t_k} P_t$.

The time interval, $\Delta t = t_k - t_{k-1}$, is not constant due to the time-varying nature of SMAP data intervals (Δt ranges from 12 hours to 5 days, see the description of data quality control below), and the regression results reflect the coefficients over averaged Δt. Overall, we do not find significant impact of the time-varying Δt on regression results (see Section C3 in Supplemental Material for detailed examination).

A direct extension of the Form I of regression (Equation (4.4)) is to add an interaction term between SSM and *P*:

Regression Form II: $$\frac{\mathbf{\Delta SSM}}{\mathbf{\Delta t}} = \gamma_1 \cdot \mathbf{SSM} + \gamma_2 \cdot \frac{\mathbf{P}}{\mathbf{\Delta t}} + \gamma_3 \cdot \mathbf{SSM} \cdot \frac{\mathbf{P}}{\mathbf{\Delta t}} + \boldsymbol{\varepsilon} \quad (4.7)$$

where γ_1 [hour^{-1}], γ_2 [-] and γ_3 [mm^{-1}] are again coefficients to be fitted. Different symbols are used to distinguish the regression coefficients from those in Regression Form I. The interaction term is the dot product of the matrix **SSM** and $\frac{\mathbf{P}}{\mathbf{\Delta t}}$, that is:

$$\mathbf{SSM}\cdot\frac{\mathbf{P}}{\mathbf{\Delta t}}=\begin{bmatrix} SSM_0\cdot\dfrac{P_1}{t_1-t_0} \\ \vdots \\ SSM_{k-1}\cdot\dfrac{P_k}{t_k-t_{k-1}} \\ \vdots \\ SSM_{n-1}\cdot\dfrac{P_n}{t_n-t_{n-1}} \end{bmatrix} \quad (4.8).$$

In this study, the regression analysis was conducted for the 3-year data at each pixel location individually, i.e., 36-km pixels for the satellite data and 1° pixels for the modeled GLDAS-VIC data. For both the satellite data (SMAP and IMERG) and the modeled data (GLDAS-VIC resampled to SMAP data interval), only adjacent SSM data points with Δt no longer than 5 days were included. To ensure statistical stability of the regression results, pixels where the total number of constructed data points in regression, n, is smaller than 100 are excluded from the analysis. In addition, columns of the independent variables are sometimes highly correlated, resulting in the collinear problem and therefore highly unstable fitted coefficients. Collinearity tends to occur for satellite data especially at locations with very few precipitation events (for example, Sahara in Africa; see Figure C1 in Supplemental Material for the pairwise scatter plot of Regression Form II variables at an example pixel in Sahara with collinearity). To mitigate this issue, we dropped the **SSM·P** interaction column (in Form II only) and/or the **P** column in the following situations: 1) drop both the **P** and the **SSM·P** columns if **P** and $\frac{\mathbf{\Delta SSM}}{\mathbf{\Delta t}}$ are not positively correlated, which indicates a very low signal-to-noise ratio in the precipitation data; 2) if any column pair has a correlation coefficient ≥ 0.98, drop the latter column. These criteria were empirically selected and adopted such that unstable regression fittings (e.g., very large fitted values for γ_3) were mitigated. The coefficients for the dropped columns were set to zero when being displayed.

As mentioned in the introduction, in this study we focus on the two regression forms with linear and interaction terms of SSM and precipitation. As the first application of this unified regression approach to discover governing hydrologic processes from the SMAP dataset, we refrained from more complicated regressors but focus on interpreting the regression results and its linkage and innovation compared with recent studies (see Sections 4.3 and 4.4 below).

However, this regression method itself is very flexible, and can be expanded to include any other regressors such as higher-order polynomial terms, other nonlinear terms, categorical terms, additional variables, and interaction between any of them. If a large number of terms are included, regularization of the coefficients needs to be applied (e.g., Lasso or Ridge regularization) to avoid overfitting. Regularization techniques can also be treated as a way of selecting the governing factors for SSM dynamics. For example, in a recent study, Brunton et al. [2016] proposed to use the Lasso regularization, which tends to shrink a subset of regression coefficients to zero, as a way of determining the governing equations of a dynamic system. No regularization was applied to the two regression forms in this study due to the small number of regressors.

Finally, to evaluate the goodness of fit of the multivariate regression, we calculated the regression R^2 from a 5-fold cross-validation (i.e., out-of-sample) experiment. While the out-of-sample experiment was applied to calculate R^2, the fitted coefficient results shown in Section 4.3 are based on the entire 3-year data. Details on model evaluation and R^2 results from both regression forms are described in Section C2 in Supplemental Material.

4.3 RESULTS

In this section, the regression results from the SMAP and IMERG data will first be presented in Section 4.3.1, focusing on the interpretation of the fitted coefficients from the regression analysis. A comparison of satellite-based and model-based results will be presented in Section 4.3.2.

4.3.1 Regression results from the SMAP and IMERG data

4.3.1.1. SSM decay rate - β_1 or γ_1

In both regression forms, the β_1 (in Regression Form I as in Equation (4.4)) or γ_1 (in Regression Form II as in Equation (4.7)) coefficient [day^{-1}] is directly related to the SSM loss function commonly defined in past studies [e.g., Koster et al., 2017; Akbar et al., 2018a] which characterizes SSM decay rate as a function of SSM level itself:

$$L(SSM) = -\frac{\Delta SSM}{\Delta t} = f(SSM) \tag{4.9}$$

In our case, we essentially assumed a linear loss function where β_1 or γ_1 is the negative slope. Since a linear loss function is equivalent to an exponential decay of SSM, another more intuitive interpretation is that $\tau = -\frac{1}{\beta_1}$ or $-\frac{1}{\gamma_1}$ [day] is the exponential SSM decay e-folding time scale [McColl et al., 2017b]. Figure 4.2 shows the converted coefficient τ [day] from the fitting of both regression forms. The τ results from the two regression forms have very similar magnitude and spatial distribution, indicating that the fitted SSM decay coefficient τ is insensitive to the additional interaction term (**SSM·P**) in the regression, and that τ is stable regardless of the regression form. Therefore, we do not distinguish in the discussion between τ converted from β_1 or γ_1. Overall, more arid regions exhibit smaller τ, in other words faster SSM decay, as summarized in Figure 4.2b and Figure 4.2d. This trend is also evident in the spatial distribution of τ in Figure 4.2a and Figure 4.2c, with smaller τ in regions such as the Sahel and southern Africa, Australia, southwestern United States, the northwestern China/Mongolia region and southwestern South America.

To further illustrate the ability of τ to characterize the SSM decay process, we back-calculated the SSM dynamics, as well as the SSM time series itself, using the fitted coefficients. The back-calculated, or "predicted", SSM time series at selected locations are shown in Figure 4.3 (left column), which illustrates the part of SSM dynamics that are captured by the fitted regression model. For all example locations ranging from humid to arid climates, the predicted SSM time series (from either regression form) is able to capture the rapid SSM decay period immediately following precipitation events. However, after the initial decay period, the SMAP retrievals tend to flatten toward a near-constant minimum baseline (likely residual soil moisture) without further precipitation input, while the predicted SSM dries out to zero. While it is possible that SMAP has an artificial and incorrect minimum detection threshold, the regression forms applied in our study likely only capture part of the SSM decay dynamics. Specifically, the regressions only include the linear form of the SSM loss function and omit the non-linear behavior of the SSM loss function in Equation (4.9). Typically, SSM decay is non-linear with a much-slower SSM decay rate when SSM is low. While it is straightforward to expand the regression method to include non-linear regressors to depict more complex SSM loss functions, we refrain from doing so since the linear SSM coefficients can be interpreted more directly. By

retaining only the linear part of the SSM loss function, we focus on the processes that are responsible for the largest soil moisture changes immediately following rainfall events.

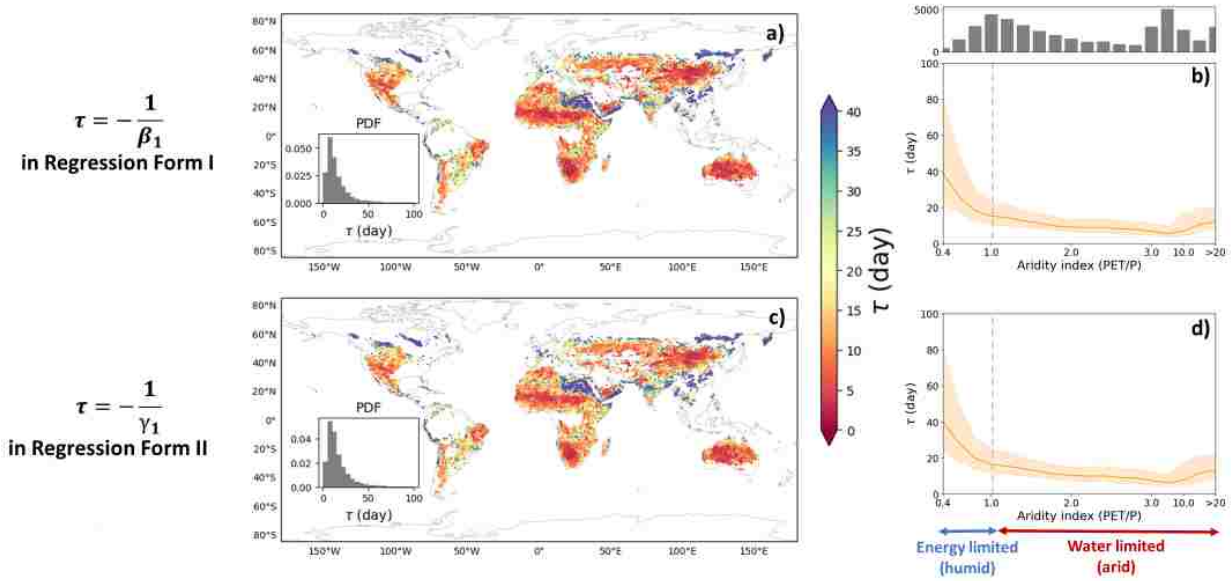

Figure 4.2. The exponential SSM decay e-folding time scale, τ [day], fitted from the SMAP/IMERG satellite data. The upper and lower panels show τ results from Regression Form I and Regression Form II, respectively. $\tau = -\frac{1}{\beta_1}$ in Regression Form I and $\tau = -\frac{1}{\gamma_1}$ in Regression Form II. The right column (Panel *b* and *d*) summarizes τ from the two regression forms as a function of the aridity index. In these panels, the solid lines are the median coefficient of all pixel locations at an aridity level, and the shaded areas show the 25th and 75th quantiles. The histogram on the top shows the pixel count for different values of the aridity index.

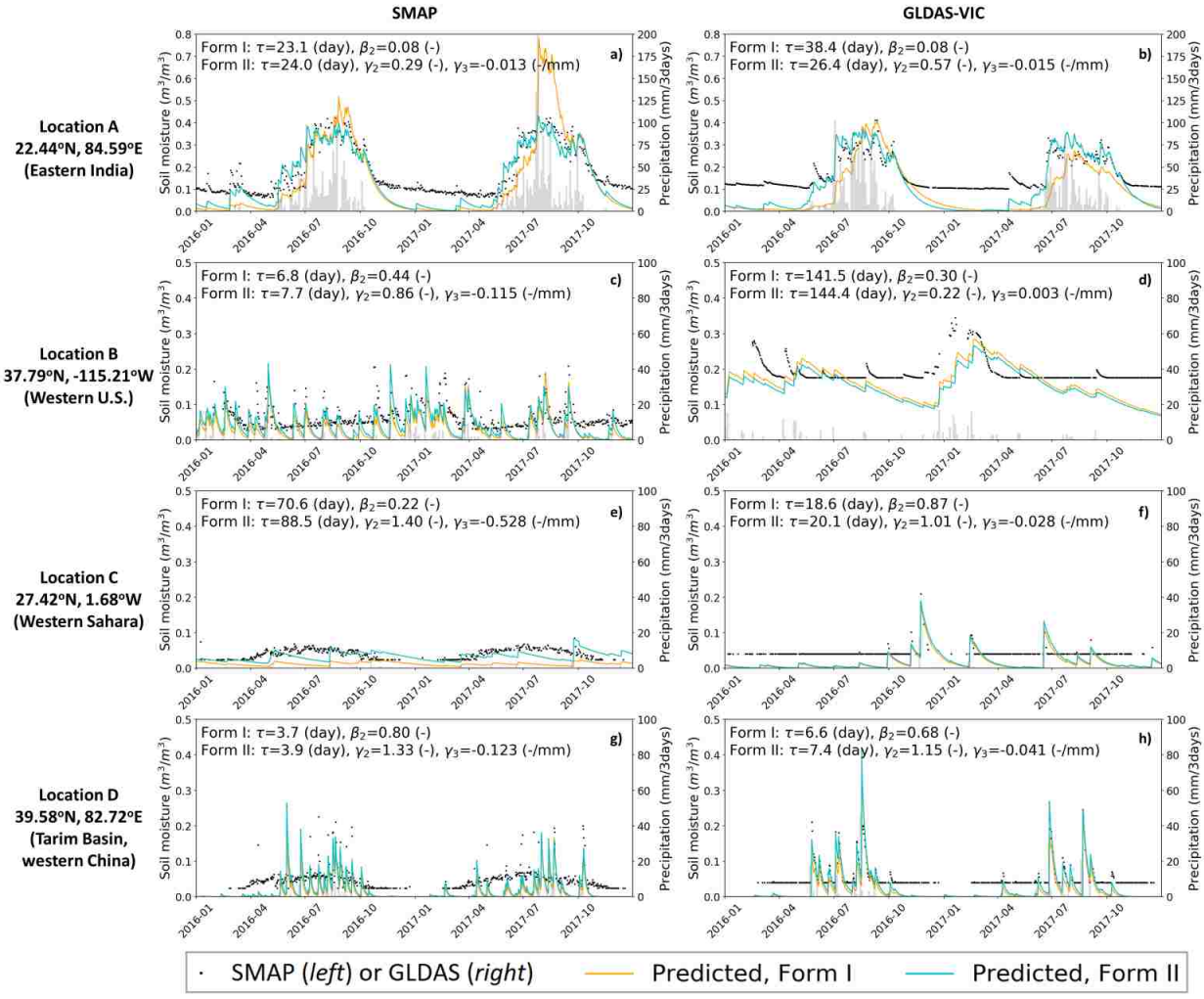

Figure 4.3. Time series of the original SMAP or GLDAS SSM (*black dots*), predicted SSM from Regression Form I (*orange lines*) and predicted SSM from Regression Form II (*green lines*) at example pixel locations. The left column shows the time series of SMAP/IMERG and the corresponding predictions, and the right column shows the time series of GLDAS-VIC and the corresponding predictions. Each row shows the time series for a single pixel location (see Figure 4.1 for the map of these locations). Precipitation data is displayed in grey bars on each subplot based on IMERG (*left*) or GLDAS (*right*) inputs. The fitted coefficients from both regression forms are labeled on each subplot.

4.3.1.2. Precipitation fraction retained – β_2

If considering the top 5 cm of soil, as detected by SMAP, as the control volume, the dimensionless coefficient β_2 in Regression Form I (Equation (4.4)) can be interpreted as the fraction of precipitation input during Δt that is added to and retained in the control volume at the end of Δt. As mentioned in the introduction, β_2 is a quantity that combines various land surface processes including canopy interception, surface runoff and water drainage to deeper soil layers, and it is not possible to separate these processes by examining β_2 alone. Figure 4.4 shows the map of the fitted β_2. More arid regions exhibit generally higher β_2 values, for example, higher β_2 values in the Sahara and southern Africa than in central Africa, higher β_2 values in western China than the rest of Asia, and higher β_2 values in the southwestern United Stated than in the central United States. This climatic trend of β_2 is also summarized in Figure 4.4b.

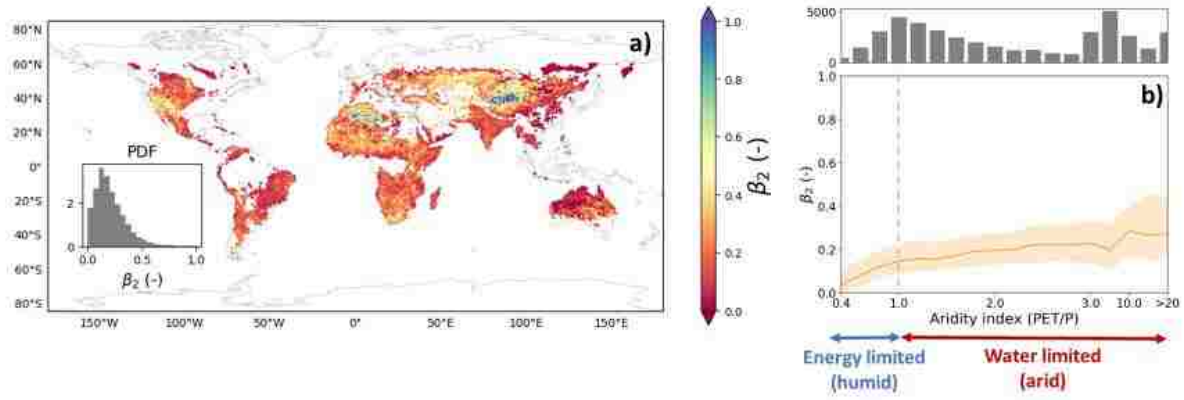

Figure 4.4. The fitted precipitation fraction retained in the top 5 cm of soil, β_2 (in Regression Form I), fitted from the SMAP/IMERG satellite data. Panel *a* shows the map of fitted β_2 and Panel *b* summarize it over aridity index (same format as Figure 4.2b and Figure 4.2d).

4.3.1.3. Sensitivity of precipitation fraction retained to SMM level - γ_3

Compared to Regression Form I, an additional interaction regressor, **SSM·P**, is included in Regression Form II (see Equation (4.7)). The combined term, $(\gamma_2 + \gamma_3 \cdot SSM)$, is analogous to the single β_2 [-] coefficient in Regression Form I which lumps all land surface processes that determine the fraction of precipitation retained in the top layer. However, the additional interaction term allows the precipitation fraction retained to vary as a function of SSM, while β_2 [-] in Regression Form I averages the fraction at all SSM levels and does not consider its

variation with SSM. The coefficient γ_3 [-/mm] can therefore be interpreted as the sensitivity of precipitation fraction retained to SSM level, or more specifically the change in the fraction per 1 mm change of SSM. Figure 4.5 (lower panels) shows that the fitted γ_3 coefficient is negative over the entire globe, meaning that precipitation fraction retained tends to be lower with higher SSM level.

Only those processes that affect the precipitation fraction retained and that vary with SSM are reflected in γ_3. For example, canopy interception affects the precipitation fraction retained but does not vary with SSM and is therefore not reflected in γ_3. In contrast, surface runoff is reflected in γ_3, because it is dependent on SSM level. For surface runoff generated via the infiltration-excess mechanism, the infiltration capacity of soil is likely smaller when surface-layer soil is wetter. For surface runoff generated via the saturation-excess mechanism, runoff is in theory only generated when soil is saturated and therefore not a function of SSM level at point scale. But over a large-scale pixel (e.g., a 36-km SMAP pixel), usually only a fraction of the pixel area is saturated and generating runoff, and this saturation area is likely to increase when the pixel-averaged SSM increases. Therefore, via either mechanism, more surface runoff tends to be generated (and thus smaller precipitation fraction retained) when SSM is higher and its sensitivity to the SSM level is reflected in the γ_3 coefficient. Finally, part of the precipitation infiltrates into the surface soil and leaves the surface layer via drainage or soil evaporation. We argue that this loss of infiltrated precipitation during Δt is reflected in γ_1 in the next timestep instead of in γ_3, since the **SSM·P** regressor uses the SSM value at the beginning of Δt (explicit numerical differentiation in Equation (4.8)). As a result, the timestep-beginning SSM does not affect the infiltrated precipitation loss during Δt. In summary, γ_3 only reflects the sensitivity of surface runoff to antecedent SSM, while canopy interception is lumped into γ_2 and the loss of infiltrated precipitation is reflected in γ_1.

To further illustrate the effect of this sensitivity on SSM dynamics, we compared the back-calculated SSM time series from the two regression forms (Figure 4.3, left column). In Figure 4.3a (time series for a pixel in eastern India), for example, the inclusion of γ_3 results in different SSM behavior especially during high-rainfall seasons. Specifically, when a rainfall event happens when the surface soil is wet (e.g., in August 2016 and July 2017), the predicted SSM from Regression Form II (*green* line) increases less than that from Regression Form I (*orange* line). Conversely, when a rainfall event happens when the surface soil is dry (e.g.,

February 2016, May 2016 and May 2017), the green line goes up more rapidly than the orange line. In other words, by including a negative γ_3 in Regression Form II, less precipitation infiltrates into the SSM storage, and therefore more surface runoff occurs, when the antecedent SSM is wet; conversely, more precipitation infiltrates and less surface runoff occurs when the antecedent SSM is dry. The failure to capture this dynamic of infiltration/runoff partition as in Regression Form I results in a generally less accurate SSM response to precipitation input, and therefore also less accurate surface runoff, especially the "flashiness" of the time series.

The spatial distribution of γ_3 as shown in Figure 4.5c and summarized in Figure 4.5d shows generally larger negative values in more arid regions, indicating higher sensitivity of infiltration/runoff partition to antecedent SSM. Although γ_3 is generally smaller in humid regions, a small error in the response of SSM to precipitation accumulates over frequent precipitation events in these regions, and the accumulated error persists due to the slower SSM decay (see Section 4.3.1.1). Therefore, even a small γ_3 difference can have a significant and longer-term effect on SSM evolution in humid regions (e.g., the time series in eastern India shown in Figure 4.3a). The R^2 of regression (Figure C2 in Supplemental Material) also show that Regression Form II has a bigger R^2 increase compared to Regression Form I in these humid regions, suggesting that the **SSM·P** term plays an important role in SSM dynamics in these regions. In contrast, arid regions with larger γ_3 values typically experience less precipitation input and more rapid SSM decays. Therefore, although the consideration of γ_3 affects the SSM response to rainfall events and thus the SSM peaks in these arid regions, the difference between with or without γ_3 fades faster over time (e.g., Figure 4.3c, Figure 4.3e and Figure 4.3g). In summary, including the **SSM·P** interaction term is important for correctly describing the SSM response to rainfall inputs in both humid and arid regions, and is essential to accurately capture the long-term SSM evolution in humid regions.

Finally, we would like to point out that the γ_2 [-] coefficient in Regression Form II represents the precipitation fraction retained when SSM is zero (see Figure 4.5, upper panels). This is a mathematical extrapolation of the precipitation fraction, and the $SSM = 0$ situation may or may not happen in reality. Since the precipitation fraction retained typically decreases with increasing SSM, γ_2 is the maximum potential precipitation fraction retained at a certain location. The coefficient γ_2 therefore reflects a property of fixed land surface conditions as a function of soil texture, vegetation cover, topography, etc., as opposed to time-varying properties that

depend on meteorological conditions. Although not the focus of this study, γ_2 can potentially be used to derive these land surface properties in future studies, and this task is only possible by including the **SSM·P** interaction term. Although β_2 and γ_2 are both regression coefficients for $\frac{\mathbf{P}}{\mathbf{\Delta t}}$ in the two regression forms, β_2 is the average precipitation fraction retained over the observed SSM dynamic range, while γ_2 is the maximum possible fraction under the hypothetical situation of $SSM = 0$. The fitted values for γ_2 (Figure 4.5, upper panels) thus show a larger magnitude and different spatial pattern than those for β_2 (Figure 4.4).

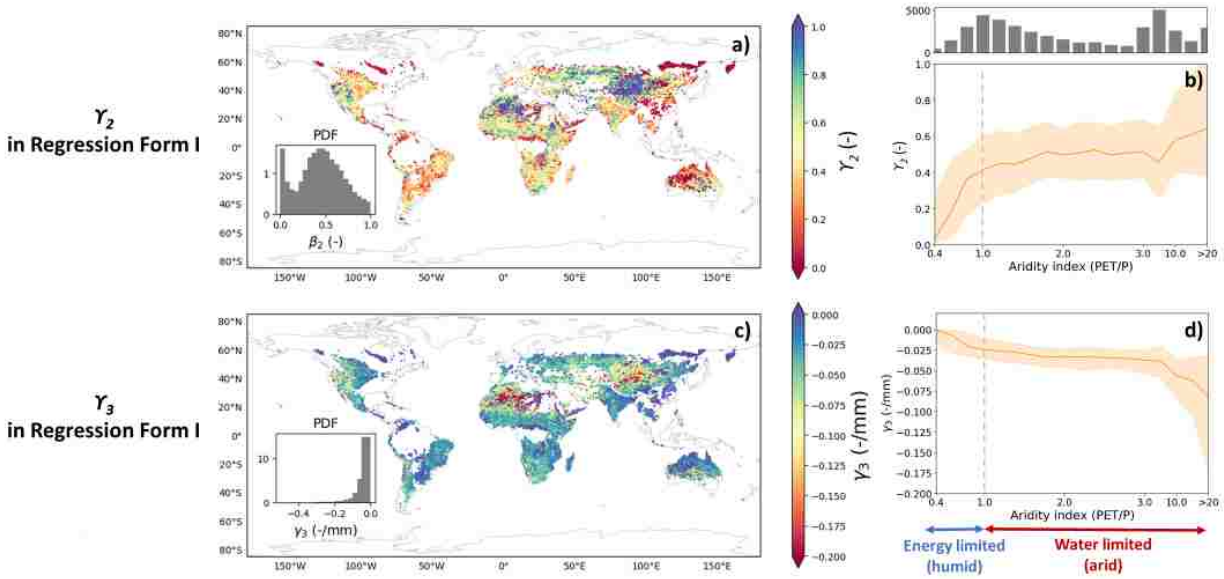

Figure 4.5. The γ_2 [-] (*upper panels*) and γ_3 [-/mm] (*lower panels*) coefficients in Regression Form II fitted from the SMAP/IMERG satellite data. The coefficient γ_2 represents the precipitation fraction retained when SSM is zero, while γ_3 represents the sensitivity of this fraction to SSM level. The right column (Panels *b* and *d*) summarize γ_2 and γ_3, respectively, as a function of the aridity index (same format as Figure 4.2b and Figure 4.2d).

4.3.2 Regression results from the modeled GLDAS-VIC data in comparison with the SMAP/IMERG data

The same regression analysis was carried out for the model-based GLDAS-VIC SSM and the corresponding GLDAS precipitation. The top soil layer in the GLDAS-VIC implementation is 10 cm, which is slightly thicker than the 5 cm detection depth of SMAP (we will discuss the implication of this difference below). As mentioned in Section 4.2.1, SSM at each 1^o GLDAS-VIC pixel was subsampled to have the same data gaps as the nearest SMAP pixel, such that the impact of sampling frequency is eliminated when comparing GLDAS-VIC with SMAP results. Based on the coefficient interpretation discussed in Section 4.3.1, here we focus on the observation/model comparison on the SSM decay time scale (τ), the overall precipitation fraction retained (β_2 in Regression Form I) and the sensitivity of this fraction to SSM level (γ_3 in Regression Form II).

4.3.2.1. SSM decay time scale τ

The GLDAS-VIC SSM decay time scale, τ, is significantly slower than that from SMAP/IMERG over most of the globe (comparing Figure 4.6a with Figure 4.2c; also summarized in Figure 4.6b). This result is consistent with a recent study by Shellito et al. [2018] who found the SSM drying rates from the Noah land surface model to be approximately double compared to that from SMAP during the initial drying period after a precipitation event. There are several possible reasons for the difference between the SMAP-observed and modeled SSM decay rates. First, the GLDAS-VIC surface soil layer is slightly deeper than the SMAP detection depth, and the larger moisture volume represented by a deeper soil layer typically decays more slowly due to slower averaged evaporation rate (although Shellito et al. [2018] claimed, via additional modeling work, that the difference in depth between SMAP and their Noah model was not the major cause for the drying rate difference). Second, Shellito et al. [2016] found that SMAP dries down faster than observed from in-situ SSM probes and concluded that the actual SMAP sensing depth may be shallower than 5 cm. Third, the VIC model structure/parameterization may be inaccurate, such as a hydraulic conductivity that is too low or insufficient vegetation water uptake, which results in SSM drainage that is too slow or too little evapotranspiration.

As summarized in Figure 4.6b, the difference between τ derived from SMAP and the modeled GLDAS-VIC is relatively small in humid regions, but τ from GLDAS-VIC tends to increase with aridity while τ from SMAP tends to decrease with aridity. This opposite climatic pattern of τ suggests that, in addition to the systematic bias of the modeled SSM decay rate by GLDAS-VIC, the model representation of SSM decay processes in GLDAS-VIC may vary incorrectly across different climatic regions. This difference in climatic pattern of τ can also be seen from the example time series in Figure 4.3. At the more humid location in eastern India, for example, the GLDAS SSM decays only slightly more slowly than the SMAP SSM, and visually the time series behave quite similarly (Figure 4.3a and Figure 4.3b). However, at the more arid location in western United States, the GLDAS-VIC SSM decays much more slowly than SMAP, resulting in a visually striking difference in the behavior of the two time series (Figure 4.3c and Figure 4.3d).

4.3.2.2. Overall precipitation fraction retained

The fitted overall β_2 [-] (in Regression Form I) from GLDAS-VIC is in general larger than that from SMAP/IMERG (comparing Figure 4.7a with Figure 4.4a). Since β_2 is a lumped quantification of total precipitation retained, the comparison result here indicates that there is systematic bias in the model representation of processes including canopy interception and surface runoff generation. As summarized in Figure 4.7b, the model-derived β_2 is larger than SMAP-derived β_2 especially in humid, energy-limited regions, suggesting inaccurate model representation of the above processes in these regions. In arid regions, the difference between model-derived and SMAP-derived β_2 is smaller.

4.3.2.3. Sensitivity of precipitation fraction retained to SSM level

The γ_3 coefficient [-/mm] derived from GLDAS-VIC (Figure 4.8a) has an overall smaller magnitude than the one derived from SMAP and its spatial distribution is more uniform (Figure 4.5c). More specifically, Figure 4.8b shows that the magnitude of the SMAP-derived γ_3 increases with aridity index, while the GLDAS-derived γ_3 remains at a low level.

As discussed in Section 4.3.1.3 above, the γ_3 coefficient reflects the sensitivity of the infiltration/runoff partition process to antecedent SSM level. In VIC, the infiltration/runoff partition process is mainly controlled by the variable infiltration curve. The infiltration curve determines the infiltration/runoff partition based on the saturation area fraction, which is estimated from the pixel-averaged soil moisture saturation level. Our satellite/model comparison

of γ_3 here suggests that the surface runoff generation process in the VIC setup in GLDAS is overall too insensitive to SSM (this conclusion is consistent with Crow et al. [2018]), especially in arid regions. From the model structure perspective, since VIC uses the saturation level of the top (l-1) layers of soil (where l is the total number of vertical layers; l = 3 in the GLDAS-VIC setup) to determine the infiltration/runoff partition, the thicker deeper layers, instead of the thin top layer, typically dominates the runoff/infiltration partition [Mao et al., 2018a]. Our γ_3 results suggest that more weight should be given to the top layer saturation level when partitioning precipitation into infiltration and runoff. From the model parameterization perspective, the variable infiltration curve parameter, *binfilt*, which determines the infiltration rate as a function of soil saturation, was set to be a global constant of 0.2 in the current version of GLDAS-VIC. Our γ_3 results suggest that a spatially varying *binfilt* parameter may result in better representation of infiltration/runoff partition.

In summary, we demonstrate the potential of using the SSM governing dynamics derived from SMAP to evaluate large-scale hydrologic model setup. Although the derived governing SSM dynamics do not necessarily allow us to derive a specific parameter or parameter value in the hydrologic model, it provides guidance to the direction of model improvement that would better represent the dominating SSM dynamics in land surface models.

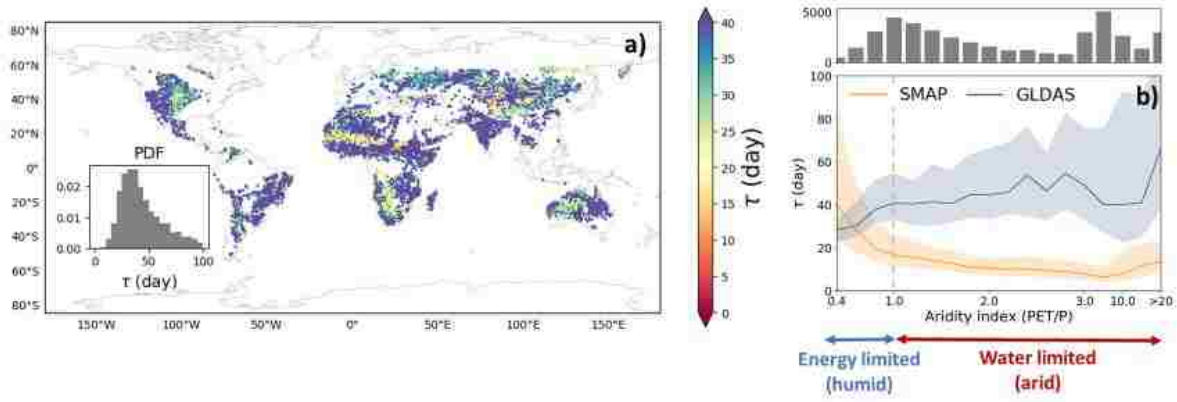

Figure 4.6. The left panel shows the exponential SSM decay e-folding time scale, τ [day], fitted for the modeled GLDAS-VIC data (Regression Form II). The right panel summarizes the GLDAS-fitted τ over aridity index (*blue*) and compared with the SMAP-fitted τ (*orange*). The format of the right panel is the same as Figure 4.2b and Figure 4.2d.

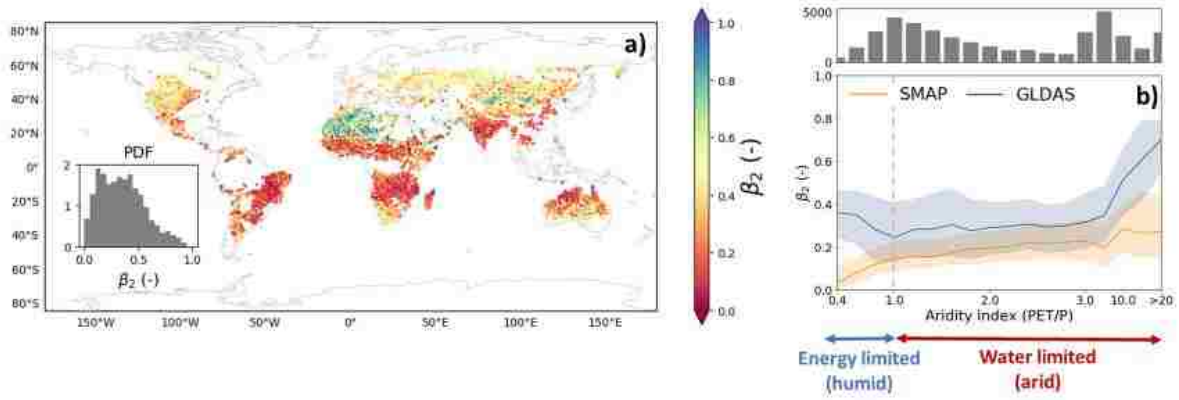

Figure 4.7. Same as Figure 4.6, but for precipitation fraction retained, β_2.

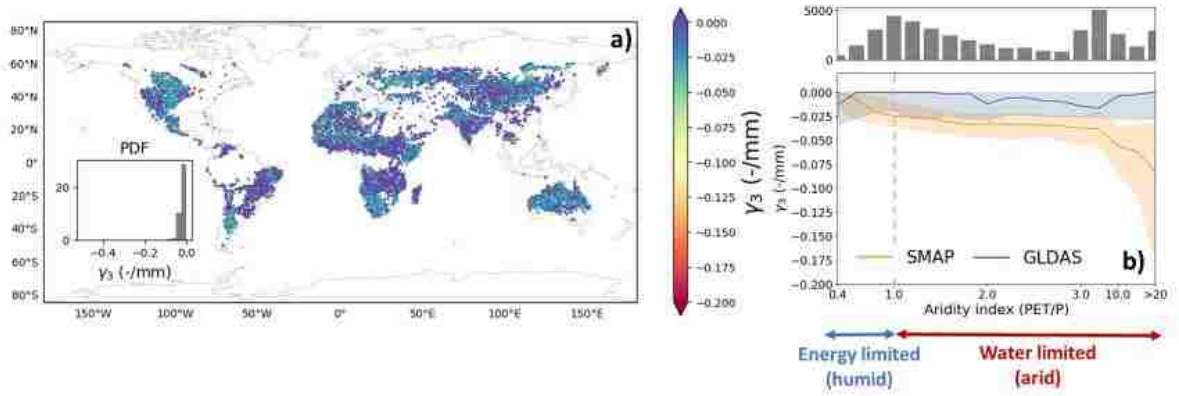

Figure 4.8. Same as Figure 4.6, but for the sensitivity of the precipitation fraction retained to SSM level, γ_3.

4.4 DISCUSSION

In Section 4.3.1 we demonstrate the ability of the unified regression method to derive multiple governing SSM dynamics from the SMAP data. In this section we further discuss our findings in the context of recent literature and highlight our contributions.

4.4.1 Comparison of the derived dynamics from SMAP with findings in literature

1) SSM decay time scale (τ)

While our calculated τ coefficient is related to several recent studies that quantified SSM drying behavior using a non-linear loss function [e.g., Shellito et al., 2016; Koster et al., 2017; Akbar et al., 2018a; Shellito et al., 2018], it is most directly related to the study by McColl et al. [2017b] where the authors extracted SSM drydown periods and calculated an averaged exponential decay time scale. Figure C7 in Supplemental Material shows that our fitted τ coefficient from SMAP reproduces the results from McColl et al. [2017b] with similar τ magnitude and spatial patterns. However, our unified regression method does not require that SSM drydown sequences be extracted first and is able to fully use the time series of data even during precipitation events.

2) Precipitation fraction retained (β_2)

The β_2 coefficient is directly related to the soil moisture memory calculated from SMAP by McColl et al. [2017a], defined as the average proportion of precipitation falling on a soil layer that is still present in the soil layer after a certain period of time. Our fitted β_2 agrees with the spatial pattern found by McColl et al. [2017a], including higher values in the Sahara, western China, the Arabian Peninsula and western United States (see Figure C8 in Supplemental Material).

Our fitted β_2 is also related to the study by Akbar et al. [2018b], who calculated an effective depth, Δz, over CONUS by calibrating the following simple water balance model against the SMAP retrievals:

$$SSM_{t+1} = SSM_t + \frac{P_t}{\Delta z} \cdot \Delta t - L(SSM) \cdot \Delta t \tag{4.10}$$

where *L(SSM)* is the SSM loss function. Comparing this water balance model with our Regression Form I (see Equations (4.4) and (4.6)), the Δz value in their study is mathematically equivalent to $\frac{1}{\beta_2}$ (except that they assumed a non-linear SSM loss function as opposed to the linear loss function used in this study). We plotted our fitted $\frac{1}{\beta_2}$ values in CONUS (Figure C9 in Supplemental Material) and observe a similar east-west gradient as Δz in Akbar et al. [2018b]. Our $\frac{1}{\beta_2}$ is generally larger than Δz from Akbar et al. [2018b] possibly due to our linear simplification of the loss function as well as different parameter optimization techniques. While Akbar et al. [2018b] argued that Δz represents the depth of an active hydrologic control volume, they did not consider important land surface processes including surface runoff and other parts of

the precipitation fluxes that never enter the land surface, such as canopy interception. In contrast, our interpretation of β_2 implicitly accounts for all these land surface processes.

4.4.2 Summary of the advantages of the unified regression method

As described above, in this study a single regression analysis is able to derive multiple SSM characteristics from SMAP simultaneously and reproduce or mimic findings from several recent stand-alone studies. This provides a unified approach to resolve various SSM dynamic terms from the SMAP data. In addition, the regression approach does not require *a priori* and somewhat subjective data pre-processing (e.g., extraction of SSM drydown periods) and can be applied directly to the original time series. Information contained in the time series data can also be extracted regardless of rain/no rain conditions.

The unified regression approach also facilitates the interpretation and distinction of multiple SSM dynamic processes. For example, Section 4.3.1 clearly shows that the SSM decay time scale, τ, represents a fundamentally different type of hydrologic process than is represented by the precipitation fraction retained, β_2. The former represents SSM loss processes while the latter represents SSM's response to precipitation input, and they correspond to two different coefficients in the regression equations. This process distinction is not clear when the individual coefficients are examined separately. For example, in the study by McColl et al. [2017b], they compared their calculated SSM exponential decay time scale τ with the soil moisture memory calculated by a previous study [McColl et al., 2017a], and discussed that the differences in the spatial patterns of the two variables were mainly due to differences in the treatment of drainage fluxes. However, we are able to point out through the unified regression approach that τ (as calculated by McColl et al. [2017b]) and soil moisture memory (as defined by McColl et al. [2017a] and equivalent to β_2) cannot be compared directly.

Finally, our regression approach is capable of resolving interrelated land surface dynamics, notably the sensitivity of the infiltration/runoff partition to the SSM level as depicted by γ_3. This coefficient reflects the coupling strength between pre-storm soil moisture and surface runoff (i.e., to what extent antecedent soil moisture affects runoff generation). Understanding this coupling is one of the key goals of land surface hydrologic science and is essential for applications such as flood forecast, drought monitoring and hydrologic data assimilation. Recently studies, for example Crow et al. [2018], have started to explore the potential for

understanding this coupling by examining the relationship between SMAP-observed pre-storm soil moisture and during-storm streamflow observations. However, estimating this coupling strength remains challenging for basins without streamflow observations. Although limited to surface soil storage and surface runoff only, the γ_3 coefficient presented here provides an innovative way to derive this coupling strength over large-scale ungauged locations without the need for streamflow observations.

4.5 CONCLUSIONS

We proposed a multivariate regression method that quantifies the governing factors that control SSM dynamics. We focus on two simple forms of regressors: the linear regressors of SSM and precipitation input (Regression Form I), and these two linear regressors with an additional interaction term (Regression Form II). This approach unifies various governing processes into a single regression analysis without the need of *a priori* data pre-processing. By examining the coefficients fitted on the 3-year global SMAP and IMERG datasets, we are able to reproduce or mimic the SSM characteristics found by several previous studies, notably the SSM exponential decay rate [McColl et al., 2017b], SSM memory defined by the fraction of precipitation retained in the top layer soil [McColl et al., 2017a], and an effective depth of hydrologic storage [Akbar et al., 2018b]. Additionally, including the **SSM·P** interaction term in the regression analysis provides a new approach to deriving the sensitivity of the infiltration/runoff partition process to antecedent SSM level in large-scale ungauged regions without the need for streamflow observations.

The analysis results derived from SMAP/IMERG are also compared with those derived from the model-based GLDAS-VIC dataset, which suggests that the GLDAS-VIC setup shows SSM decay that is too slow (through drainage and/or evaporation), and the runoff/infiltration partition shows too little sensitivity to top-layer soil moisture. This conclusion is only made possible by including the **SSM·P** interaction term in regression. These results provide guidance to the direction of model development that would better capture the dominating SSM dynamics. The satellite-model comparison also points out the inaccuracy of the spatial distribution of these land surface processes in different climatic regions.

A direct extension of the presented regression analysis would be to include additional regressors to derive more comprehensive governing processes for SSM. As illustrated in this study, adding the **SSM·P** term in addition to the two linear regressors of SSM and precipitation allows us to resolve interesting new processes. Other potential regressor terms to add include: 1) higher-order polynomial SSM terms that will help capture non-linear SSM decay behaviors (as extracted by stand-alone approaches in previous studies such as in Akbar et al. [2018a] and Koster et al. [2017]), 2) binary SSM-threshold term which will take into account the non-linear processes due to SSM saturation, and 3) additional data sources such as air temperature, which would make it possible to separate out energy-driven processes such as evaporation. Care needs to be taken to avoid overfitting the noisy data when including a larger number of regressors. Interpreting the fitted coefficients from these more complicated regressors is also more challenging and we refrained from using them in this initial study to highlight the usage and interpretation of the unified regression methodology.

Another potential direction of future research is to derive soil properties from the resolved SSM dynamics at the resolution of the SMAP measurements, which are of similar resolution as continental- to global-scale hydrologic model implementation. Deriving soil properties at this resolution circumvents the difficulty that it is not straightforward to upscale point-scale soil measurements to pixel-averaged values due to unobserved, complex soil/vegetation heterogeneity. Our comparison of the regression coefficients based on SMAP observations and on GLDAS-VIC provided insights into soil properties that determine runoff/infiltration partitioning, but additional soil property parameters can potentially be developed. It is also potentially possible for future research to investigate human impacts on SSM dynamics. Water source/sink due to human activities, such as irrigation and pumping, could be included as input in the regression model to improve the accuracy of the analysis. In regions where such data is unavailable, it may be possible to derive human activity signals from the derived SSM dynamics.

Acknowledgements

This work was supported in part by NASA Terrestrial Hydrology Program Award NNX16AC50G to the University of Washington and NASA Terrestrial Hydrology Program Award 13-THP13-0022 to the United States Department of Agriculture, Agricultural Research

Service. All the data sources used in this study have been referenced in the manuscript. The analysis code for this study is available at https://github.com/yixinmao/smap_data.

REFERENCES

Akbar, R., D. J. Short Gianotti, K. A., McColl, E. Haghighi, G. D. Salvucci, and D. Entekhabi (2018a), Estimation of landscape soil water losses from satellite observations of soil moisture, *J. Hydrometeorol.*, 19 (5), 871-889, doi:10.1175/JHM-D-17-0200.1.

Akbar, R., D. S., Gianotti, K. A., McColl, E. Haghighi, G. D. Salvucci, and D. Entekhabi (2018b), Hydrological storage length scales represented by remote sensing estimates of soil moisture and precipitation, *Water Resour. Res.*, 54, 1476-1492, doi:10.1002/2017WR021508.

Brunton, S. L., J. L. Proctor, and J. N. Kutz (2016), Discovering governing equations from data by sparse identification of nonlinear dynamical systems, *Proc. Natl. Acad. Sci.*, 113 (15), 3932-3937, doi: 10.1073/pnas.1517384113.

Chan et al. (2018), Development and assessment of the SMAP enhanced passive soil moisture product, *Remote Sens. Environ.*, 204, 931-941, doi:10.1016/j.rse.2017.08.025.

Crow, W. T., F. Chen, R. H. Reichle, Y. Xia, and Q. Liu (2018), Exploiting soil moisture, precipitation, and streamflow observations to evaluate soil moisture/runoff coupling in land surface models, *Geophys. Res. Lett.*, 45, 4869-4878, doi:10.1029/2018GL077193.

Dingman, L. S. (2002), Physical Hydrology (2nd edition), Long Grove: Waveland Press, Inc.

Entekhabi et al. (2010), The Soil Moisture Active and Passive (SMAP) Mission, *Proceedings of the IEEE*, 98(5), 704-716, doi:10.1109/JPROC.2010.2043918.

Hou, A. Y., R. K. Kakar, S. Neeck, A. A. Azarbarzin, C. D. Kummerow, M. Kojima, R. Oki, K. Nakamura, and T. Iguchi (2014), The Global Precipitation Measurement mission, *Bull. Amer. Meteor. Soc., 95(5), 701-722, doi:10.1175/BAMS-D-13-00164.1.*

Huffman, G. J., D. T. Bolvin, and E. J. Nelkin (2015), Integrated Multi-Satellite Retrievals for GPM (IMERG) Technical Documentation. Tech. Doc., NASA GSFC. [Available online at https://docserver.gesdisc.eosdis.nasa.gov/public/project/GPM/IMERG_doc.05.pdf.]

Huffman, G. (2017), GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V05, Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: 2018/10/17, doi:10.5067/GPM/IMERG/3B-HH/05.

Koster, R. D., R. H. Reichle, and S. P. P. Mahanama (2017), A data-driven approach for daily real-time estimates and forecasts of near-surface soil moisture, *J. Hydrometeorol.*, 18 (3), 837-843, doi:10.1175/JHM-D-16-0285.1.

Koster, R. D., W. T. Crow, R. H. Reichle, and S. P. Mahanama (2018), Estimating basin-scale water budgets with SMAP soil moisture data, *Water Resour. Res.*, 54, doi:10.1029/2018WR022669.

Malmstrom, V. H. (1969), A new approach to the classification of climate. *J. of Geography*, 68 (6), 351-357, doi:10.1080/00221346908981131.

Mao Y., W. T. Crow, and B. Nijssen (2018a), A framework for diagnosing factors degrading the streamflow performance of a soil moisture data assimilation system, *J. Hydrometeorol.*, in review.

Mao Y., W. T. Crow, and B. Nijssen (2018b), Dual state/rainfall correction via soil moisture assimilation for improved streamflow simulation: Evaluation of a large-scale implementation with SMAP satellite data, in preparation.

McColl, K. A., S. H. Alemohammad, R. Akbar, A. G. Konings, S. Yueh, and D. Entekhabi (2017a), The global distribution and dynamics of surface soil moisture, *Nat. Geosci.*, 10, 100-104, doi:10.1038/NGEO2868.

McColl, K. A., W. Wang, B. Peng, R. Akbar, D. J. Short Gianotti, H. Lu, M. Pan, and D. Entekhabi (2017b), Global characterization of surface soil moisture drydowns, *Geophys. Res. Lett.*, 44, 3682–3690, doi:10.1002/2017GL072819.

O'Neill, P. E., S. Chan, E. G. Njoku, T. Jackson, and R. Bindlish (2016), SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Version 4, Boulder, Colorado USA, NASA National Snow and Ice Data Center Distributed Active Archive Center, Accessed 2018-10-17, doi:10.5067/OBBHQ5W22HME.

Rodell, M., P.R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J.K. Entin, J.P. Walker, D. Lohmann, and D. Toll (2004), The Global Land Data Assimilation System, *Bull. Amer. Meteor. Soc.*, 85(3), 381-394, doi:10.1175/BAMS-85-3-381.

Shellito, P. J., et al. (2016), SMAP soil moisture drying more rapid than observed in situ following rainfall events, *Geophys. Res. Lett.*, 43, 8068-8075, doi:10.1002/2016GL069946.

Shellito, P. J., E. E. Small, and B. Livneh (2018), Controls on surface soil drying rates observed by SMAP and simulated by the Noah land surface model, *Hydrol. Earth Syst. Sci.*, 22, 1649-1663, doi:10.5194/hess-22-1649-2018.

Chapter 5. CONCLUSIONS AND FUTURE WORK RECOMMENDATIONS

5.1 CONCLUSIONS

In this dissertation, I examine various ways of extracting hydrologic information from the SMAP surface soil moisture data and assess its ability to improve hydrologic modeling. This works is motivated by the four main research questions described in Chapter 1: 1) To what extent can the SMAP surface soil moisture data improve streamflow simulation via state updating data assimilation techniques, and why? 2) To what extent can the SMAP surface soil moisture data correct the GPM precipitation estimates, and why? 3) To what extend can the SMAP surface soil moisture data improve streamflow simulation via a dual state/rainfall correction system? And 4) How can we extract soil moisture dynamics from the SMAP data and use this information to evaluate process representation in hydrologic models?

Chapter 2 concludes that state updating alone is not sufficient to substantially improve large-scale streamflow simulations. The main reasons include that, at least in our case study using the Variable Infiltration Capacity (VIC) model, fast-response surface runoff is not primarily coupled with antecedent states, and slow-response subsurface runoff is controlled by the deeper-layer soil moisture that is not well corrected via the assimilation of surface measurements (Question 1). Chapter 3 further includes the rainfall correction scheme that uses SMAP to correct the GPM rainfall estimates and finds only a slight rainfall correction, mainly because of the improved baseline quality of the new-generation GPM satellite product (Question 2). Combining the corrected rainfall with updated soil moisture states, the dual state/rainfall correction system leads to slight to moderate improvement in simulated streamflow (Question 3). Chapter 2 and Chapter 3 together point out that, while soil moisture data assimilation (DA) techniques are able to nudge the hydrologic variables (i.e., antecedent soil moisture states, the rainfall estimates and the subsequent streamflow simulations) toward the correct direction, these methods only correct the zero-mean random errors in a hydrologic simulation system, but not the often larger systematic error. Therefore, substantial improvement of streamflow simulation is not possible via DA techniques alone, but will rely on reduction of systematic errors (Question 3).

These findings from Chapter 2 and Chapter 3 serve as a strong motivation for the work conducted in Chapter 4, where the surface soil moisture dynamics are extracted from SMAP and are compared with a model-based global soil moisture dataset to inform its structure/parameterization inaccuracy. To do so, a unified multivariate regression method is proposed to extract the governing factors for surface soil moisture dynamics from data. This unified regression method is able to mimic several SMAP characteristics (such as soil moisture decay rate and soil moisture memory) extracted separately from recent studies, and it additionally provides a new approach to derive the dependency of the infiltration/runoff partition process on antecedent surface soil moisture level without the need for streamflow observation data (Question 4). The SMAP/model comparison results suggest that the hydrologic model examined retains moisture longer in the top soil layer, shows less sensitivity of the infiltration/runoff partition process to the top-layer moisture, and exhibits less spatial variation in surface soil moisture dynamics (Question 4).

5.2 FUTURE WORK RECOMMENDATIONS

The work in this dissertation demonstrates that extracting information from the new-generation satellite data products (e.g., SMAP) is promising for better understanding land surface hydrologic processes, as well as for informing hydrologic models and model implementations. Although the more traditional soil moisture DA approaches that use satellite soil moisture estimates are able to provide some improvement to the simulated hydrologic variables (e.g., soil moisture states, rainfall estimates, streamflow), we may be approaching the limit of using such methods to correct the random error components in a hydrologic modeling system. Therefore, it would be more beneficial to focus future research effort on extracting hydrologic behaviors from satellite data and using them to derive improved model structure/parameterization. While Chapter 4 in this dissertation gives an initial illustration toward this research direction, additional data-driven approaches need to be developed to more directly link the findings from satellite data with hydrologic modeling setup. Furthermore, the findings from this dissertation also stress the benefit and potential of continuing to upgrade the earth-observing satellites themselves, such that an even more accurate depiction of the land surface processes would be possible.

APPENDIX A

A framework for diagnosing factors degrading the streamflow performance of a soil moisture data assimilation system – Supplemental Material

This appendix includes the supplemental materials from chapter 2. This material has been accepted for publication in its current form in the *Journal of Hydrometeorology*.

Mao Y., W. T. Crow, and B. Nijssen (2018), A framework for diagnosing factors degrading the streamflow performance of a soil moisture data assimilation system, *Journal of Hydrometeorology*, accepted, doi:10.1175/JHM-D-18-0115.1.

A1. Mathematical details of ensemble Kalman filter (EnKF)

The ensemble Kalman filter (EnKF) method is one of the most commonly used data assimilation (DA) techniques in hydrology. EnKF was first introduced by Evensen [1994] and has subsequently been applied to a large number of land DA applications [e.g., Crow and Ryu, 2009; Chen et al., 2014; Massari et al., 2015]. It represents model error by an ensemble of model run replicates, which are combined with measurements sequentially to update model states. Specifically, the EnKF method is based on a propagation model and a measurement model:

$$x_{k+1} = f(x_k, u_k) + \omega_k \quad \text{(A1)}$$

$$\tilde{y}_k = Hx_k + v \quad \text{(A2)}$$

where subscript k is a discrete time index; x is a column vector of model states to update (the column vector length is the total number of state variables to update); u is model meteorological forcing; $f()$ is a land surface model that propagates states to the next timestep; ω lumps together modeling errors during propagation from various sources including forcing data error, model structure error and parameterization error; $\tilde{y}$ is measurement data, in our context surface SM measurements; H is an observation operator that relates model states x to measurements $\tilde{y}$; and v is measurement error.

In a standard EnKF, an ensemble size of N model replicates is propagated and updated sequentially over time in the following way:

1) An ensemble of initial model states is first generated by perturbing the initial deterministic model states to represent initial state error;

2) For each ensemble member, the land surface model is run until the next measurement time with perturbed meteorological forcing to represent forcing error. Model states are directly perturbed as well to represent random errors from model structure and parameterization;

3) Once an observation time is reached, the Kalman gain *K* is calculated as:

$$K_k = P_k H^T \cdot (H P_k H^T + R)^{-1} \tag{A3}$$

where *R* is the measurement error variance, and the forecast state error covariance matrix P_k is estimated by sampling across the propagated ensemble states:

$$P_k = \frac{1}{N-1} \sum_{j=1}^{N} (\hat{x}_k^{-(j)} - \overline{\hat{x}}_k^{-})(\hat{x}_k^{-(j)} - \overline{\hat{x}}_k^{-})^T \tag{A4}$$

where $\hat{x}_k^{-(j)}$ is the propagated state vector at time *k* for the *j*th ensemble member, and $\overline{\hat{x}}_k^{-}$ is the mean of $\hat{x}_k^{-(j)}$ across all ensemble members;

4) Following the calculation of *K*, each ensemble member of states is individually updated as:

$$\hat{x}_k^{+(j)} = \hat{x}_k^{-(j)} + K_k \cdot (\tilde{y}_k + v_k^{(j)} - \hat{y}_k^{-(j)}) \tag{A5}$$

where $\hat{y}_k^{-(j)}$ is the simulated measurement at time *k* for the *j*th ensemble member, i.e., $\hat{y}_k^{-(j)} = H\,\hat{\mathrm{x}}_k^{-(j)}$; $v_k^{(j)}$ is random noise added to represent measurement error whose error statistic is consistent with *R* in Equation (A3).

A2. Mathematical details of the evaluation metrics used in the study

1) Percent error reduction (PER)

PER is defined as the percent reduction in the root-mean-squared error (RMSE) compared to the open-loop baseline:

$$PER = \left[1 - \frac{RMSE_a}{RMSE_{open}}\right] \times 100 \tag{A6}$$

where the subscripts *open* and *a* denote the open-loop and DA analysis runs, respectively.

2) Kling-Gupta efficiency (KGE)

The Kling-Gupta efficiency (KGE) [Gupta et al. 2009] combines the performance in terms of correlation, variance and bias:

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \tag{A7}$$

where r is the correlation coefficient between simulated and observed streamflow; α is the ratio of their standard deviations; and β is the ratio of their means. KGE ranges from negative infinity to 1 with values closer to 1 indicating better performance.

3) Percent continuous rank probability score reduction (PSR)

Continuous rank probability score (CRPS) measures the deviation of the cumulative distribution function (CDF) of an ensemble from that of a reference (observation in the real-data case or truth in the synthetic case) [Hersbach, 2000]. If we assume the reference has zero uncertainty, then its CDF (denoted by $F_t^r(s)$ where t denotes timestep and r denotes observation) is a unit step function:

$$F_t^o(\mathrm{s}) = \begin{cases} 0, & \mathrm{s} < \mathrm{y}_t^r \\ 1, & s \geq \mathrm{y}_t^r \end{cases} \tag{A8}$$

where y_t^r is the reference variable value and s is a random variable. CRPS is then calculated as the temporal mean of the CDF deviation from the reference:

$$CRPS = \frac{1}{n}\sum_{t=1}^{n}\int_{-\infty}^{\infty}[F_t^a(\mathrm{s}) - F_t^r(\mathrm{s})]^2\, ds \tag{A9}$$

where $F_t^a(s)$ denotes the CDF of an analysis ensemble at time t, and n is the total length of the time series. In practice, the continuous CDF of an analysis, $F_t^a(s)$, is empirically estimated by the finite ensemble. Note that CRPS penalizes both a deviation of the ensemble mean from the observation and a large ensemble spread. A smaller (i.e., closer-to-zero) CRPS value indicates better ensemble performance.

Analogous to PER, the percent CRPS reduction (PSR) quantifies the percent reduction in CRPS of a DA analysis compared to the open-loop baseline (note that here the baseline is the open-loop ensemble instead of the deterministic open-loop run):

$$PSR = \left[1 - \frac{CRPS_a}{CRPS_{open}}\right] \times 100 \tag{A10}$$

4) Normalized ensemble skill (NENSK)

NENSK measures the ensemble-mean error normalized by ensemble spread:

$$NENSK = \frac{ENSK}{ENSP} \tag{A11}$$

where the ensemble skill (ENSK) and ensemble spread (ENSP) are calculated as:

$$ENSK = \frac{1}{n}\sum_{t=1}^{n}(\bar{y}_t^a - y_t^o)^2 \quad \text{(A12)}$$

$$ENSP = \frac{1}{n}\sum_{t=1}^{n}(y_t^{a,(j)} - \bar{y}_t^a)^2 \quad \text{(A13)}$$

where $\bar{y}_t^a$ denotes the ensemble-mean value at timestep *t*, and $y_t^{a,(j)}$ denotes the value of the *j*th ensemble member. In an ideal situation where an ensemble is a correct representation of analysis uncertainty, the observed or true condition looks like one of the realizations of the ensemble [Anderson, 1996; Wilks, 2011] and therefore NENSK should be approximately one [Talagrand et al., 1997; Wilks, 2011] (NENSK > 1 indicates an under-dispersed ensemble while NENSK < 1 indicates an over-dispersed ensemble). This metric has been used to verify ensemble hydrologic variables [e.g., De Lannoy et al., 2006; Brocca et al., 2012; Alvarez-Garreton et al., 2014].

REFERENCES

Alvarez-Garreton, C., D. Ryu, A. W. Western, W. T. Crow, and D. E. Robertson (2014), The impacts of assimilating satellite soil moisture into a rainfall-runoff model in a semi-arid catchment, *J. Hydrol.*, 519, 2763-2774, doi:10.1016/j.jhydrol.2014.07.041.

Anderson, J. L. (1996), A method for producing and evaluating probabilistic forecasts from ensemble model integrations, *J. Clim*, 9, 1518-1530, doi: 10.1175/1520-0442(1996)009<1518:AMFPAE>2.0.CO;2.

Brocca, L., T. Moramarco, F. Melone, W. Wagner, S. Hasenauer, and S. Hahn (2012), Assimilation of surface-and root-zone ASCAT soil moisture products into rainfall–runoff modeling, *IEEE Trans. Geosci. Remote Sens.*, 50(7), 2542-2555, doi: 10.1109/TGRS.2011.2177468.

Chen, F., W. T. Crow, and D. Ryu (2014), Dual forcing and state correction via soil moisture assimilation for improved rainfall–runoff modeling, *J. Hydrometeorol.*, *15*(5), 1832–1848, doi:10.1175/JHM-D-14-0002.1.

Crow, W. T., and D. Ryu (2009), A new data assimilation approach for improving hydrologic prediction using remotely-sensed soil moisture retrievals, *Hydrol. Earth Syst. Sci.*, 12(1-16), doi:10.5194/hess-13-1-2009.

De Lannoy, G. J. M., P. R. Houser, V. R. N. Pauwels, and N. E. C. Verhoest (2006), Assessment of model uncertainty for soil moisture through ensemble verification, *J. Geophys. Res.*, 111, D10101, doi: 10.1029/2005JD006367.

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*(C5), 10143-10162, doi:10.1029/94JC00572.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez (2009), Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, 377, 80-91, doi: 10.1016/j.jhydrol.2009.08.003.

Hersbach, H. (2000), Decomposition of the continuous ranked probability score for ensemble prediction systems, *Weather Forecast.*, 15(5), 559-570, doi: 10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Massari, C., L. Brocca, A. Tarpanelli, and T. Moramarco (2015), Data Assimilation of Satellite Soil Moisture into Rainfall-Runoff Modelling: A Complex Recipe?, *Remote Sens.*, 7, 11403-11433, doi:10.3390/rs70911403.

Talagrand, O., R. Vautard, and B. Strauss (1997), Evaluation of probabilistic prediction systems, technical report, Eur. Cent. for Medium-Range Weather Forecast., Reading, UK.

Wilks, D. S. (2011), Statistical methods in the atmospheric sciences (3rd edition), Elsevier/Academic Press, Amsterdam; Boston.

APPENDIX B

Dual state/rainfall correction via soil moisture assimilation for improved streamflow simulation: Evaluation of a large-scale implementation with SMAP satellite data – Supplemental Material

This appendix includes the supplemental materials from chapter 3.

B1. The ensemble Kalman smoother (EnKS) version of the Soil Moisture Analysis Rainfall Tool (SMART)

The Soil Moisture Analysis Rainfall Tool (SMART) is a rainfall correction scheme developed and updated by Crow et al. [2009; 2011] and Chen et al. [2012]. It is based on sequential assimilation of soil moisture (SM) measurements into a simple Antecedent Precipitation Index (API) model to obtain SM increments, and then linearly relates these increments to rainfall accumulation errors. In the study we extended the ensemble Kalman filter (EnKF) version of SMART developed by Crow et al. [2011] to an ensemble Kalman smoother (EnKS) version with probabilistic rainfall estimates.

Following Crow et al. [2009; 2011], the API model is used to capture the response of moisture storage (represented by the *API* state) to rainfall input:

$$API_t = \gamma API_{t-1} + P_t \tag{B1}$$

where t is a timestep index; P is the original uncorrected precipitation observation and γ is a loss coefficient (dimensionless) that accounts for storage loss through evaporation, drainage, etc. In the ensemble version of SMART [Crow et al., 2011], Equation (B1) is converted to:

$$API_t^{(j)} = \gamma API_{t-1}^{(j)} + \eta_t^{(j)} P_t + \omega_t^{(j)} \tag{B2}$$

where the superscript (j) denotes the jth ensemble member; η is multiplicative noise with mean 1 added to the observed precipitation to represent random precipitation forcing error; and ω is zero-mean Gaussian noise to represent random API model structure and parameterization error. The API state can be related directly to SM content via rescaling [Crow et al., 2009]. The rescaled SM measurement, θ, can therefore be assimilated to update the *API* states via the standard EnKS technique both at the measurement timestep and during the data gap before the measurement timestep. Mathematically, if two adjacent measurements come in at time k and

time m with $m - k \geq 1$, then the measurement at time m is used to calculate the gain K and API increment δ for each timestep i at timestep m as well as during the gap (i.e., $k < i \leq m$):

$$K_i = \frac{T_{im}}{T_m + R_m} \tag{B3}$$

and

$$\delta_i^{(j)} = API_i^{+(j)} - API_i^{-(j)} = K_i \cdot (\theta_m + \kappa_m^{(j)} - API_m^{-(j)}) \tag{B4}$$

where K is the Kalman gain; T_{im} is the covariance matrix between *API* states at time i and m; R is the measurement error variance for the rescaled SM measurements; the superscript *(j)* denotes the jth ensemble member; the superscripts "-" and "+" denote *API* states before and after an update, respectively; κ is zero-mean Gaussian noise added to represent the random SM measurement error. T_{im} is calculated as:

$$T_{im} = \frac{1}{M-1} \sum_{j=1}^{M} (API_i^{-(j)} - \overline{API}_i^{-}) \cdot (API_m^{-(j)} - \overline{API}_m^{-}) \tag{B5}$$

where M is the ensemble size; $\overline{API}_t^{-}$ is the ensemble-mean *API* states before update.

The SMART algorithm then uses ensemble-mean *API* increment δ to estimate the rainfall correction amount via a simple linear relation. We extended this relation to produce an ensemble of corrected rainfall time series (instead of the single rainfall estimates in past studies) where each ensemble member of the perturbed rainfall time series is corrected by the corresponding member of δ.:

$$[P_{corr}^{(j)}]_l = [\eta^{(j)} P^{(j)}]_l + \lambda [\delta^{(j)}]_l \tag{B6}$$

where "[ ]" denotes temporally aggregated P or δ (in the SMART study in this paper, this window was set to the 3-hour native SMART timestep without aggregation); l is the new time index for the aggregated windows; λ is a scaling factor that can either be calibrated or set to a prescribed constant. Finally, negative P_{corr} resulted from Equation (B6) are reset to zero, and the final corrected precipitation time series is (multiplicatively) rescaled to be unbiased over the entire simulation period toward the original precipitation observation time series.

B2. Investigation of cross-correlation of errors in the dual system

B2.1. Background and methods

It is well known that correlated errors in different parts of a Kalman filter result in sub-optimal filter outputs. Therefore, in the original paper detailing the dual state/rainfall correction system, Crow and Ryu [2009] advised that the corrected rainfall (informed by the SM measurements) should not be fed back into the state EnKF correction scheme into which the same SM measurements are assimilated. Instead, corrected rainfall and states should be combined via an offline model simulation (see Figure 3.1 and Section 3.2.4.3 in the main manuscript). Later studies that applied the dual correction system all followed this general guideline [e.g., Chen et al., 2014; Alvarez-Garreton et al., 2016]. However, although this guideline helps avoid first-order error correlation in the system, it does not completely eliminate the possibility of error cross-correlation. Specifically, the corrected rainfall and the updated states are informed by the same SM measurement, thus they potentially inherit the same error from the SM measurement. When fusing the two schemes together, such inherited error could potentially be amplified, degrading streamflow performance or cause a probabilistic estimate (based on an implicit assumption of independent errors) to be biased or have inaccurate uncertainty spread. In other words, it is possible that the current system still suffers from some second-order issue of overusing the information of SM measurements. Massari et al. [2018] intentionally avoided combining the state update scheme and the rainfall correction scheme in their study due to this legitimate concern.

To further investigate this issue, we designed a set of synthetic experiments and applied in an arbitrary small domain within the Arkansas-Red (a box around the Little Arkansas subbasin, see Table 3.3 and Figure 3.2 in the main manuscript for its location). Synthetic measurements, instead of the real SMAP measurements, were generated and assimilated into the dual correction system so that we have complete control over all the error statistics and correlation, which is impossible in a real-data case. Specifically, a single perturbed VIC realization (with perturbed forcing and states) was treated as the synthetic "truth". Synthetic measurement can then be generated at daily interval by degrading the true surface-layer SM by adding random measurement errors. Precipitation perturbation was assumed to be temporally auto-correlated (first-order autoregressive noise with parameter $\phi = 0.9$), and all the other error

assumptions and dual correction setup were consistent with those described in Section 3.2.4 in the main manuscript.

We generated two sets of synthetic measurements based on the same truth with the same measurement error statistics but mutually independent realizations of errors. Then, two scenarios of dual correction were designed and carried out (see Figure B1 for illustration):

Scenario 1: the same set of synthetic SM measurement were assimilated into both the state update and the rainfall correction schemes. This scenario mimics the issue in the real-data dual system with error cross-correlation in the two schemes and potentially degraded streamflow;

Scenario 2: two sets of synthetic SM measurements (with mutually independent errors) were assimilated into the two schemes separately. This scenario completely avoids the issue of error cross-correlation.

The final runoff performance from the dual correction system were evaluated toward the truth, and the runoff performance from the two scenarios was compared. Differences in the performance of the two scenarios would indicate degradation caused by error cross-correlation. For these synthetic experiments, runoff was evaluated locally at each grid cell without routing, since we know the true condition locally.

B2.2. Results

Deterministic and probabilistic results from the two scenarios were compared in Figures B2 and B3. Clearly, runoff results show only very little difference between the two scenarios in terms of both PER and NENSK (see Section 3.2.5 in the main manuscript for details of the two metrics). This is true for both the total runoff and the fast- and slow-response runoff components separately. This suggests that the streamflow performance is not noticeably degraded by assimilating the same SM retrievals to both the state update and rainfall correction schemes. Although the cross-correlated error theoretically exists in the system, they are not big enough to cause problematic streamflow results. In other words, we are not over-using the information contained in SM retrievals in the system. This is true both from a deterministic sense and in terms of probabilistic representation. We also experimented the case where the synthetic measurements were assumed to have temporally auto-correlated errors instead of white errors, which in theory creates bigger risk of degradation in the subsequent streamflow, but drew similar conclusions as above (results not shown).

The synthetic results in this section validates that we can safely assimilate the SMAP retrievals into both schemes of the dual correction system without significantly degrading the final streamflow performance.

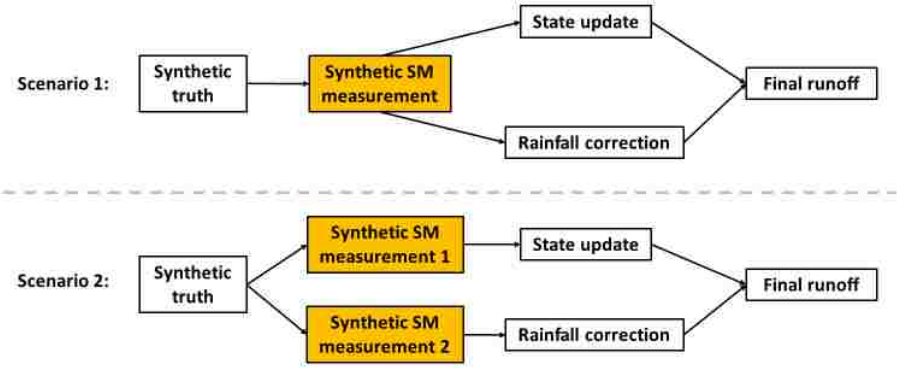

Figure B1. Illustration of the synthetic experiments for investigating error cross-correlation.

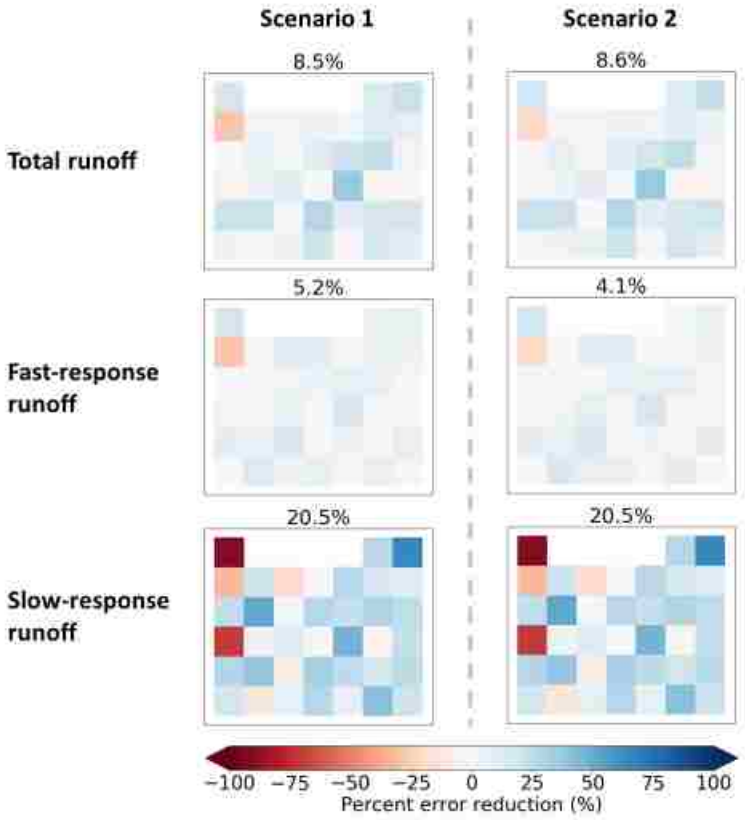

Figure B2. Percent RMSE reduction (PER) of synthetic daily runoff results from the error cross-correlation experiment. Blue color indicates runoff improvement after dual correction while red color indicates degraded runoff. The two columns show the results from the two assimilation scenarios described in Section B2. The three rows show results of total runoff, fast-response runoff and slow-response runoff, respectively. The number on top of each subplot indicates the domain-median PER.

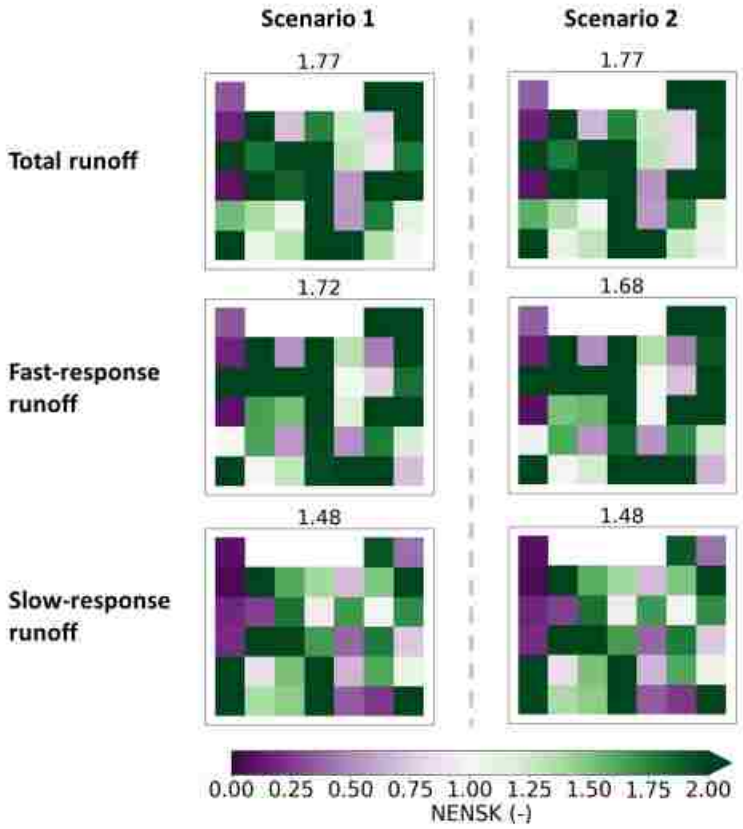

Figure B3. Same as Figure B2 but for NENSK. Lighter color (either green or purple) indicates closer-to-one (thus better) NENSK.

REFERENCES

Alvarez-Garreton, C., D. Ryu, A. W. Western, W. T. Crow, C.-H. Su, and D. R. Robertson (2016), Dual assimilation of satellite soil moisture to improve streamflow prediction in

data-scarce catchments, *Water Resour. Res.*, 52(7), 5357-5375, doi:10.1002/2015WR018429.

Chen F., W. T. Crow, and T. R. H. Holmes (2012), Improving long-term, retrospective precipitation datasets using satellite-based surface soil moisture retrievals and the Soil Moisture Analysis Rainfall Tool, *J. Appl. Remote Sens.*, 6(1), 063604, doi: 10.1117/1.JRS.6.063604.

Chen, F., W. T. Crow, and D. Ryu (2014), Dual forcing and state correction via soil moisture assimilation for improved rainfall–runoff modeling, *J. Hydrometeorol.*, *15*(5), 1832–1848, doi:10.1175/JHM-D-14-0002.1.

Crow, W. T., and D. Ryu (2009), A new data assimilation approach for improving hydrologic prediction using remotely-sensed soil moisture retrievals, *Hydrol. Earth Syst. Sci.*, 12(1-16), doi:10.5194/hess-13-1-2009.

Crow W. T., G. J. Huffman, R. Bindlish, and T. J. Jackson (2009), Improving satellite-based rainfall accumulation estimates using spaceborne surface soil moisture retrievals, *J. Hydrometeorol.*, 10, 199-212, doi: 10.1175/2008JHM986.1.

Crow, W. T., M. J. van den Berg, G. J. Huffman, and T. Pellarin (2011), Correcting rainfall using satellite-based surface soil moisture retrievals: The Soil Moisture Analysis Rainfall Tool (SMART), *Water Resour. Res.*, 47, W08521, doi:10.1029/ 2011WR010576.

Massari, C., S. Camici, L. Ciabatta, and L. Brocca (2018), Exploiting satellite-based surface soil moisture for flood forecasting in the Mediterranean area: State update versus rainfall correction, *Remote Sens.*, 10, 292, doi: 10.3390/rs10020292.

APPENDIX C

A unified data-driven method to derive hydrologic dynamics from global SMAP surface soil moisture and GPM precipitation data - Supplemental Material

This appendix includes the supplemental materials from chapter 4.

C1. Collinearity between regressors

In a multivariate regression analysis, the problem of collinearity occurs when columns of the regressors are highly correlated. Collinearity causes unstable results in fitted coefficients, making the coefficients uninterpretable. Section 4.2.2 in the main manuscript discusses the collinearity problem in the context of the unified regression method proposed in the study. Figure C1 here demonstrates the pairwise scatter plot of Regression Form II at an example pixel in the Sahara Desert where collinearity occurs.

C2. Regression model evaluation

C2.1. Methodology

To evaluate the goodness of fit of the multivariate regression, we calculated the regression R^2 from a 5-fold cross-validation (i.e., out-of-sample) experiment. Specifically, for each pixel the constructed data points (including the $\frac{\Delta SSM}{\Delta t}$ and the corresponding independent variables) were randomly shuffled and divided into 5 subsets of identical size. To ensure that model evaluation is performed on an independent set of data used in fitting, each of the 5 subsets was left out in turn, and the regression was performed on the rest of the data while R^2 was calculated on the left-out subset. The mean R^2 of the 5-fold cross-validation was calculated and reported at each pixel. This cross-validation model evaluation strategy was applied to both regression forms and both the satellite analysis and the model-based analysis. Note that the cross-validation was only applied to calculate R^2, while the fitted coefficient results shown in the paper are based on the entire 3-year data.

C2.2. Results from the SMAP/IMERG satellite data

The cross-validated R^2 was calculated and shown in Figure C2 for both regression forms. R^2 in the majority of the measured globe is significantly greater than zero for both regression

forms, indicating that the simple regressors account for at least part of the surface soil moisture (SSM) dynamics. R^2 is generally larger in humid regions and smaller in arid regions, which is consistent with the predicted time series shown in Figure 4.3 in the main manuscript with visually more accurate back-calculated SSM at an example humid location (Figure 4.3a, eastern India) than at example arid locations (Figure 4.3c, Figure 4.3e and Figure 4.3g). Although R^2 is not close to 1 anywhere around the globe, we emphasize that the goal of the regression study is to quantify the governing factors for SSM dynamics, rather than accurately simulate the detailed behaviors. While only linear regressors are included in this study, the resulting R^2 shows that the simplified representation of governing factors is able to capture a significant part of SSM behavior.

R^2 is around zero in the eastern part of the Sahara Desert. Precipitation is very low in this region, providing little forcing that drives SSM change. As a result, the signal-to-noise ratio of $\frac{\Delta SSM}{\Delta t}$ observed by SMAP is very low, resulting in approximately zero R^2. Similarly, R^2 is relatively lower (although above zero) in other low precipitation regions (e.g, the majority of Australia, western Sahara) than in high precipitation regions (e.g., Sahel and southern Africa, central and western United States, India, central South America).

In general, Regression Form II displays higher out-of-sample R^2 than Form I (Figure C2) in most regions, especially the high-precipitation regions as mentioned above, suggesting that adding the **SSM·P** interaction term as a regressor improves the goodness of fit of describing SSM dynamics. In other words, **SSM·P** is a non-negligible contributor to SSM dynamics in these regions.

C2.3. Results from the GLDAS-VIC modeled data

R^2 of GLDAS-based regression is shown in Figure C3. It exhibits significantly larger R^2 than that of SMAP/IMERG over most of the globe (Figure C2), since there is no measurement error in either SSM or precipitation input in the modeled case. Similar to the satellite data case, Regression Form II achieves a better goodness of fit due to the additional interaction term **SSM·P**.

C3. Impact of SMAP data intervals on regression results

In the paper, the modeled GLDAS-VIC SSM data was subsampled to have the same time-varying data gaps as SMAP retrievals to ensure a fair comparison between satellite data and

modeled data. Here we examined the effect of the irregular SMAP intervals (typically ranging from 12 hours to 2-3 days) on fitted regression coefficients utilizing the GLDAS-VIC modeled data for which we have control on the SSM data interval.

Figures C4, C5 and C6 show the SSM decay time scale τ, the overall precipitation fraction retained (i.e., β_2 in Regression Form I), and the sensitivity of the precipitation fraction retained to SSM level (i.e., γ_3 in Regression Form II), respectively, derived from the GLDAS-VIC data with 12-hour interval. The temporally-uniform 12-hour interval was used since it is the lower limit of SMAP retrieval interval. Comparing these three figures to the corresponding figures used in the paper from GLDAS-VIC data subsampled to the SMAP interval (Figure 4.6, Figure 4.7 and Figure 4.8 in the main manuscript), we observe similar magnitude and spatial distribution SSM decay time scale τ with the two intervals. β_2 from the 12-hourly data is different than that from the SMAP-frequency data in some regions, for example the Sahara Desert exhibits higher β_2 values based on the 12-hourly data and the Tibet Plateau exhibits lower β_2 values based on the 12-hourly data, but the difference is not visually obvious in most other regions. Finally, the γ_3 coefficient also displays similar magnitude and spatial distribution from the 12-hourly SSM data compared to the SMAP-frequency data.

In summary, although small impact exists (especially on β_2), overall the 12-hourly to 2- to 3-day irregular SMAP data interval does not have significant effect on the resulting fitted coefficients from the regression analysis.

C4. Comparison of the regression results with literature

In the main manuscript, a few fitted coefficients from the SMAP regression analysis, including the SSM decay time scale, τ, and the precipitation fraction retained, β_2, are compared with related findings from recent studies. Figures C7, C8 and C9 show the maps of these fitted coefficients (or their converted forms) in a colormap that is directly comparable with figures in literature. Specifically, Figure C7 shows our fitted τ in a similar colormap as Figure 2 in McColl et al. [2017b] for their exponential decay e-folding time scale extracted from SMAP. Figure C8 shows our fitted β_2 in a similar colormap as Figure 2b in McColl et al. [2017a] for their soil moisture memory. Figure C9 shows $\frac{1}{\beta_2}$ in the Continental United States (CONUS) in a similar colormap as Figure 6 in Akbar et al. [2018b] for their effective depth Δz. All these figures and

their comparison with literature are discussed in more detail in Section 4.4 in the main manuscript.

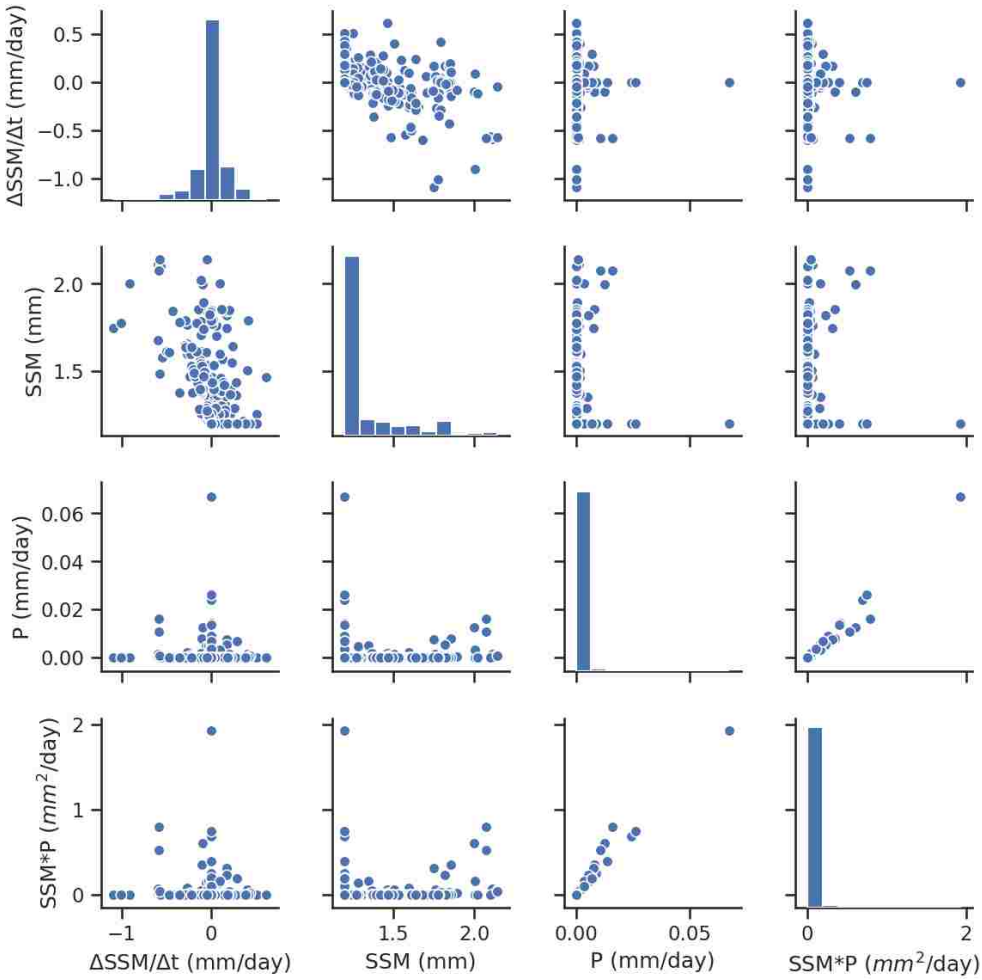

Figure C1. Pairwise scatter plot of the dependent variable and regressors constructed in Regression Form II from SMAP/IMERG data at an example pixel (23.98^{o}N, 23.71^{o}E) in Sahara with collinearity.

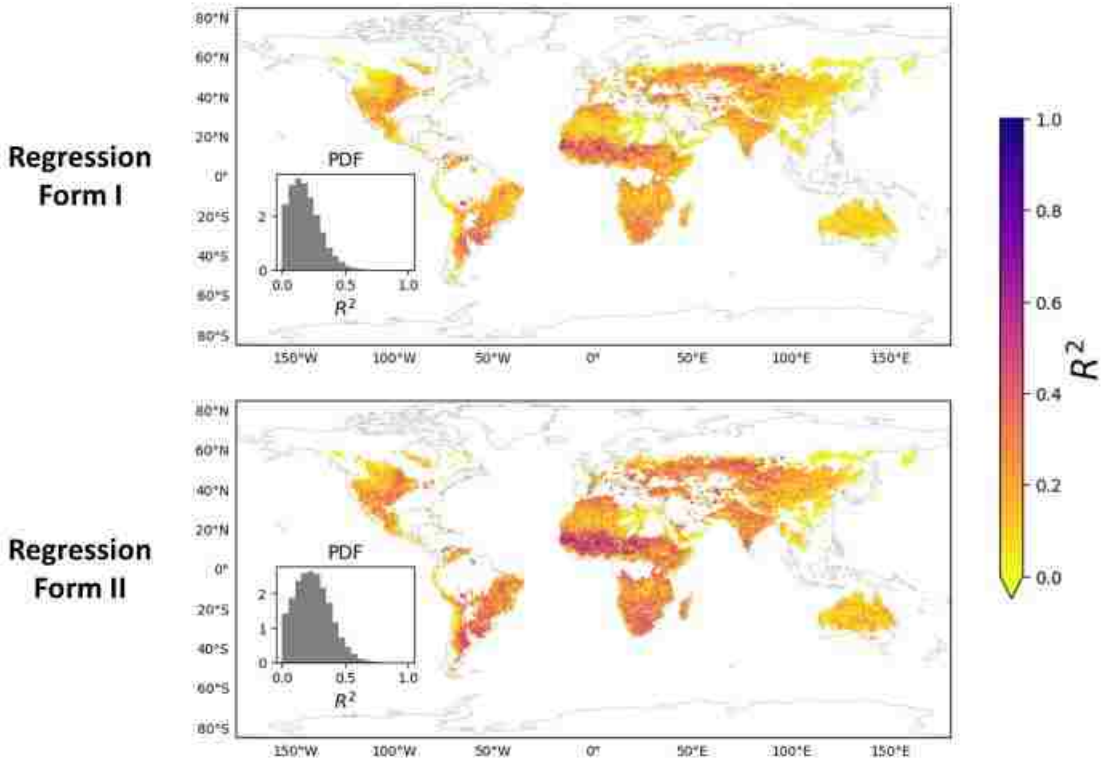

Figure C2. Average R^2 of the 5-fold cross-validation experiment for the regression of SMAP and IMERG satellite data. The upper and lower panels show the R^2 maps of the Regression Form I and Form II, respectively.

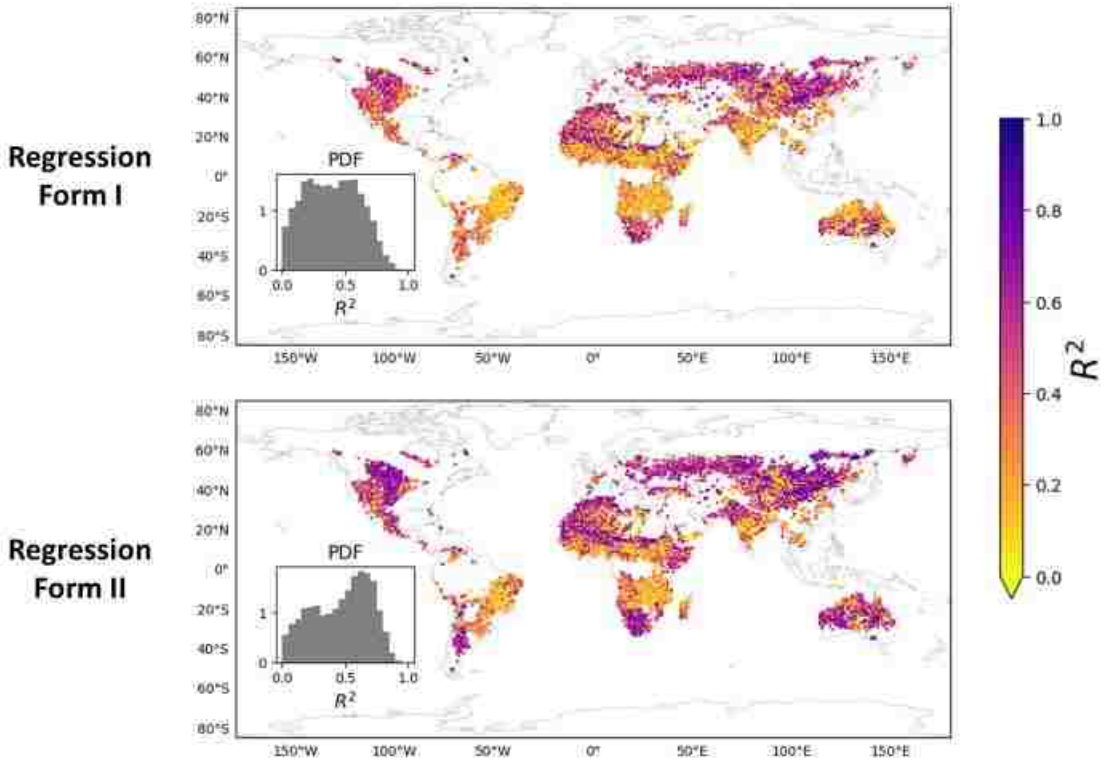

Figure C3. Same as Figure S2 but for modeled GLDAS-VIC data.

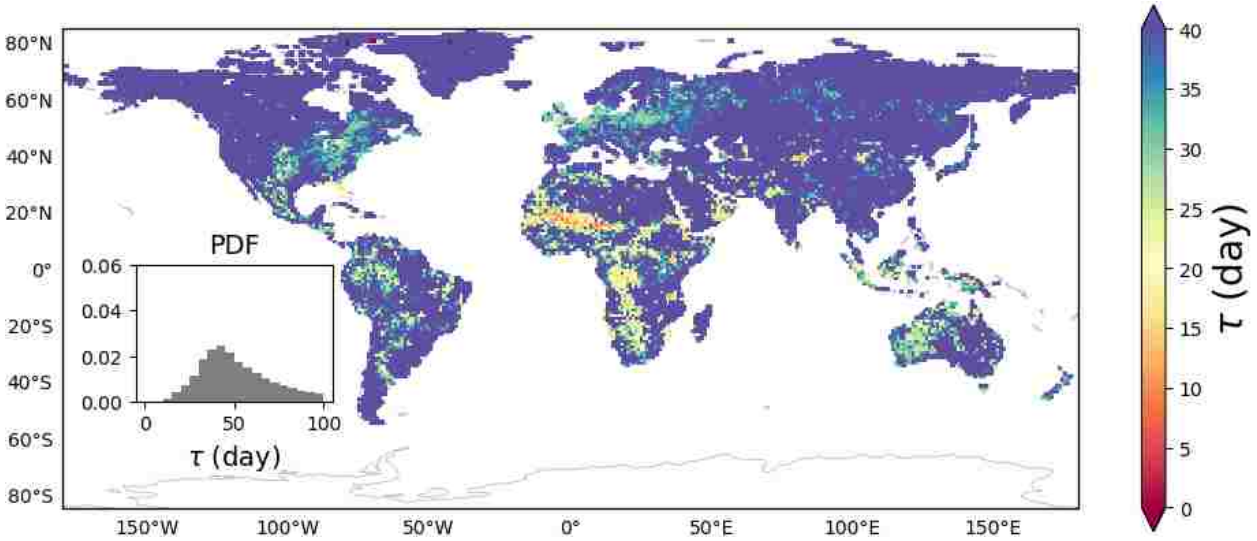

Figure C4. The exponential SSM decay e-folding time scale, τ [day], fitted for the modeled GLDAS-VIC data with 12-hour data interval (Regression Form II).

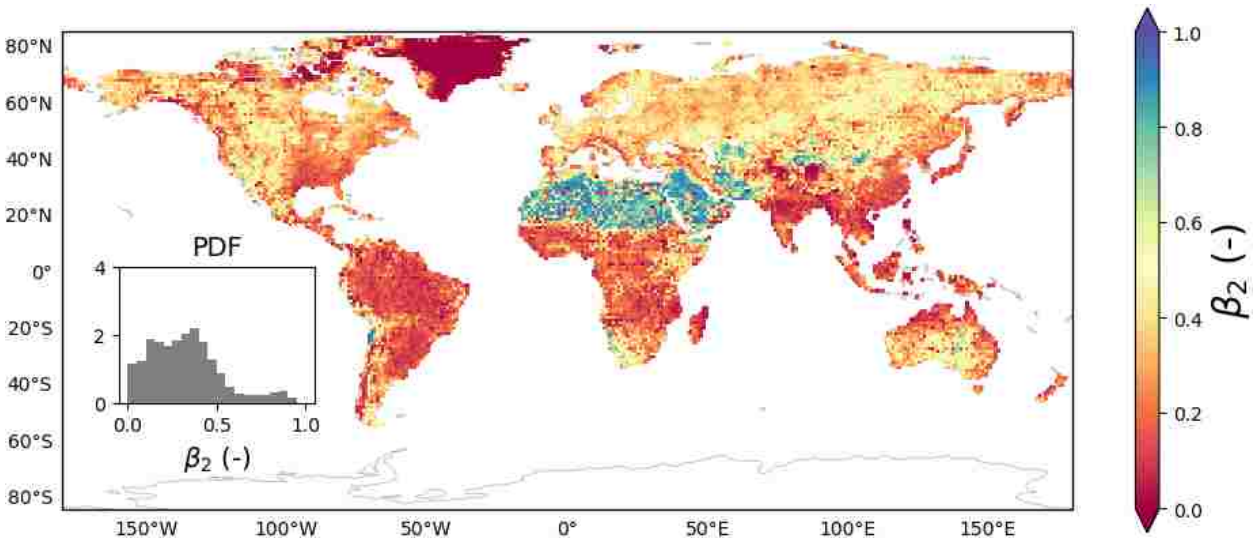

Figure C5. The fitted precipitation fraction retained in the top soil layer, β_2 (in Regression Form I), fitted from the modeled GLDAS-VIC data with 12-hour data interval.

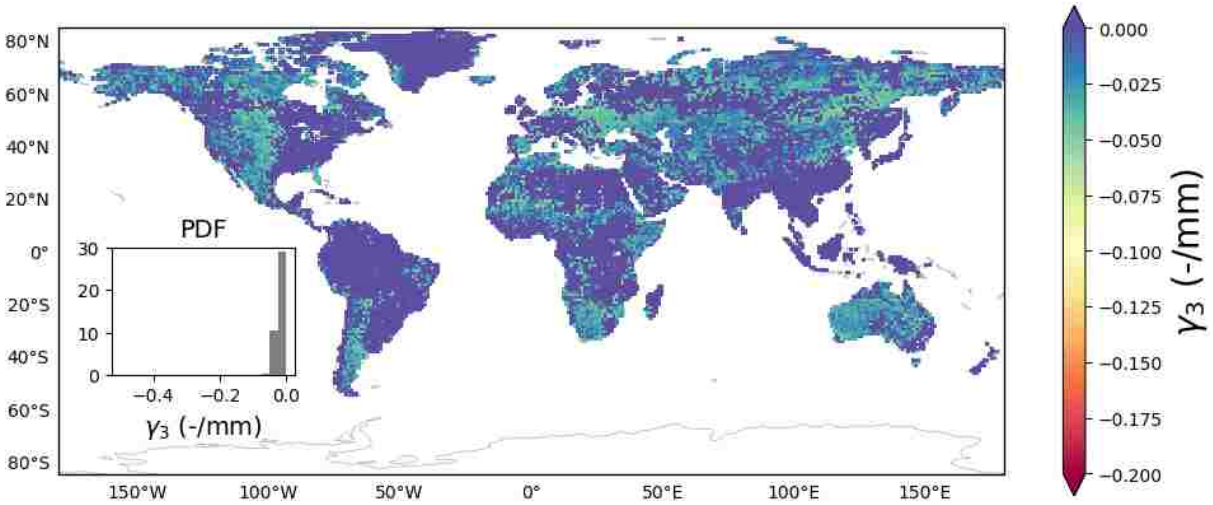

Figure C6. The γ_3 [-/mm] coefficient in Regression Form II, representing the sensitivity of the precipitation fraction retained to SSM level, fitted from the modeled GLDAS-VIC data with 12-hour data interval.

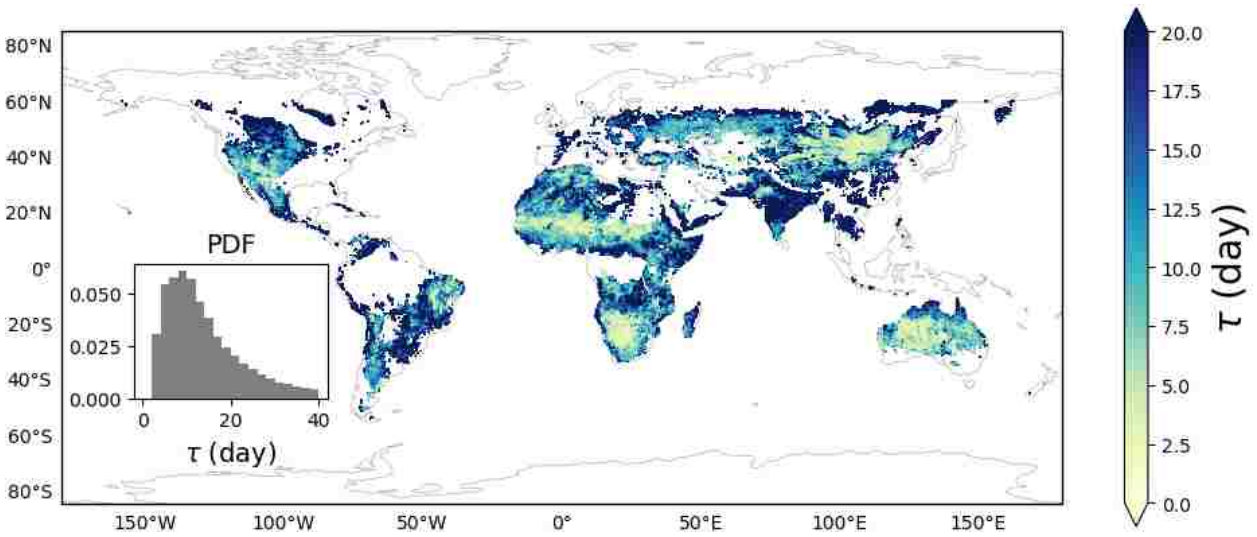

Figure C7. The τ coefficient fitted in this study (from Regression Form II), plotted in a similar colormap as Figure 2 in McColl et al. [2017b] for their SMAP exponential decay time scale.

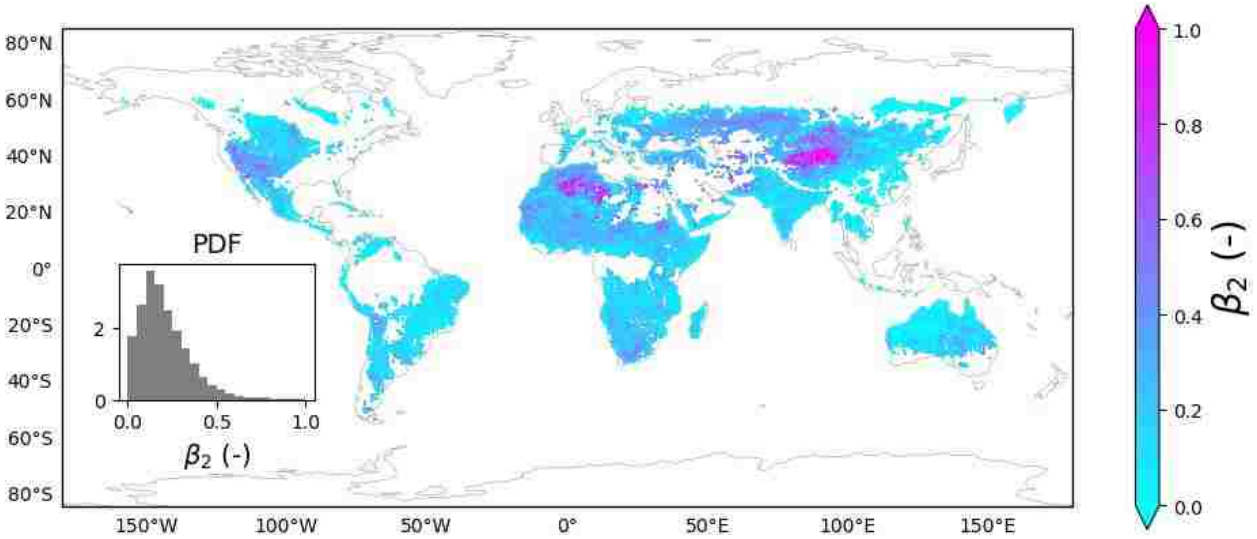

Figure C8. The β_2 coefficient fitted in this study (from Regression Form I) plotted in a similar colormap as Figure 2b in McColl et al. [2017a] for their soil moisture memory.

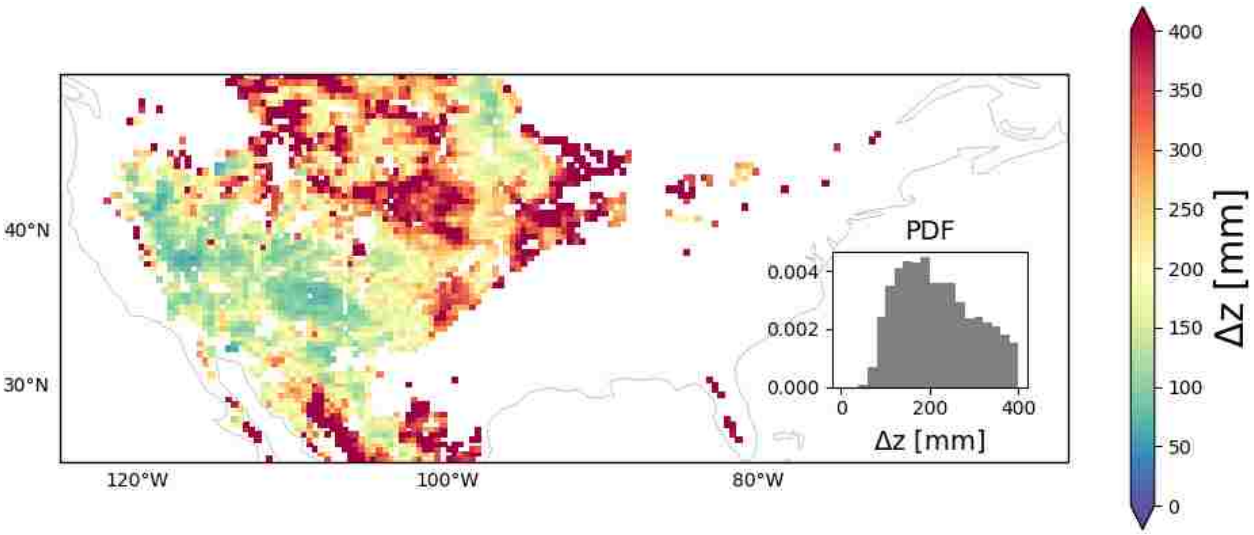

Figure C9. The $\frac{1}{\beta_2}$ coefficient fitted in this study (from Regression Form I) in the Continental United States (CONUS) plotted in a similar colormap as Figure 6 in Akbar et al. [2018b] for their Δz.

REFERENCES

Akbar, R., D. S., Gianotti, K. A., McColl, E. Haghighi, G. D. Salvucci, and D. Entekhabi (2018b), Hydrological storage length scales represented by remote sensing estimates of soil moisture and precipitation, *Water Resour. Res.*, 54, 1476-1492, doi:10.1002/2017WR021508.

McColl, K. A., S. H. Alemohammad, R. Akbar, A. G. Konings, S. Yueh, and D. Entekhabi (2017a), The global distribution and dynamics of surface soil moisture, *Nat. Geosci.*, 10, 100-104, doi:10.1038/NGEO2868.

McColl, K. A., W. Wang, B. Peng, R. Akbar, D. J. Short Gianotti, H. Lu, M. Pan, and D. Entekhabi (2017b), Global characterization of surface soil moisture drydowns, *Geophys. Res. Lett.*, 44, 3682–3690, doi:10.1002/2017GL072819.

VITA

Yixin Mao

EDUCATION

Ph.D., 2018: Civil and Environmental Engineering, University of Washington, Seattle, WA. Dissertation title: Extracting hydrologic information from the Soil Moisture Active Passive (SMAP) satellite data for improved hydrologic modeling (Advisor: Bart Nijssen)

B.E., 2013: Energy and Resources Engineering, Peking University, Beijing, China. Thesis title: Response of surface and subsurface water resources to human activities and climate change in the Luanhe Plain, China (Advisor: Yi Zheng)

PUBLICATIONS

Mao, Y., W. T. Crow, and B. Nijssen (2018), A framework for diagnosing factors degrading the streamflow performance of a soil moisture data assimilation system, *Journal of Hydrometeorology*, accepted, doi:10.1175/JHM-D-18-0115.1.

Cao, Q., E. A. Clark, **Y. Mao**, and D. P. Lettenmaier (2018), Trends and interannual variability in terrestrial water storage over the eastern United States, 2003-2016, *Water Resources Research*, in review.

Hamman, J. J., B. Nijssen, T. J. Bohn, D. R. Gergel, and **Y. Mao** (2018), The Variable Infiltration Capacity Model, Version 5 (VIC-5): Infrastructure improvements for new applications and reproducibility, *Geoscintific Model Development*, 11, 3481-3496, doi: 10.5194/gmd-11-3481-2018.

Niemeyer, R., Y. Cheng, **Y. Mao**, J. Yearsley, and B. Nijssen (2018), A thermally-stratified reservoir module for large-scale distributed stream temperature models with application

in the Tennessee River Basin, *Water Resources Research*, 54, doi: 10.1029/2018WR022615.

Feng D., Y. Zheng, **Y. Mao**, A. Zhang, B. Wu, J. Li, Y. Tian, and X. Wu (2018), An integrated hydrological modeling approach for detection and attribution of climatic and human impacts on coastal water resources, *Journal of Hydrology*, 557, 305-320, doi:10.1016/j.jhydrol.2017.12.041.

Mizukami N., M. P. Clark, K. Sampson, B. Nijssen, **Y. Mao**, H. McMillan, R. J. Viger, S. L. Markstrom, L. E. Hay, R. Woods, J. R. Arnold and L. D. Brekke (2016), mizuRoute version 1: a river network routing tool for a continental domain water resources applications, *Geoscientific Model Development*, 9, 2223-2238, doi:10.5194/gmd-9-2223-2016.

Mao Y., B. Nijssen and D. P. Lettenmaier (2015), Is climate change implicated in the 2013–2014 California drought? A hydrologic perspective, *Geophysical Research Letters*, 42(8), 2805-2813, doi:10.1002/2015GL063456.

Henn B., Q. Cao, D. P. Lettenmaier, C. S. Magirl, C. Mass, J. B. Bower, M. St Laurent, **Y. Mao** and S. Perica (2015), Hydroclimatic conditions preceding the March 2014 Oso landslide, *Journal of Hydrometeorology*, 16 (3), 1243-1249, doi:10.1175/JHM-D-15-0008.1.

SELECTED PRESENTATIONS

Mao Y. (Feb 2018), Python as a tool to build large-scale geoscience systems - An example on a hydrologic data assimilation system. Oral presentation, *Python in the Geosciences seminar series*, University of Washington, Seattle, WA.

Mao Y., W. T. Crow and B. Nijssen (Dec 2017), A 3-step framework for understanding the added value of surface soil moisture measurements for large-scale runoff prediction via data assimilation. Oral presentation, *2017 AGU Fall Meeting*, New Orleans, LA.

Mao Y., W. T. Crow and B. Nijssen (Dec 2016), Dual state/rainfall correction via soil moisture assimilation for improved hydrologic prediction - a synthetic study using the VIC Model in the Arkansas-Red River basin. Poster session, 2016 *AGU Fall Meeting*, San Francisco, CA.

Chegwidden O., B .Nijssen, **Y. Mao** and D. E. Rupp (Dec 2016), Hydrologic climate change impacts in the Columbia River Basin and their sensitivity to methodological choices. Oral presentation, 2016 *AGU Fall Meeting*, San Francisco, CA.

Niemeyer R. J., Y. Cheng, **Y. Mao**, J. R. Yearsley and B. Nijssen (Dec 2016), Incorporating a simple two-layer reservoir into a coupled land surface and river routing model to improve river temperature simulations in the Tennessee River Basin. Poster session, *2016 AGU Fall Meeting*, San Francisco, CA.

Cheng Y., Niemeyer R. J., **Y. Mao**, J. R. Yearsley and B. Nijssen (Dec 2016), Climate change impacts on river temperature in the southeastern United States: a case study of the Tennessee River basin. Poster session, *2016 AGU Fall Meeting*, San Francisco, CA.

Mao Y., T. Zhou, J. R. Yearsley and B. Nijssen (Dec 2015), Future climate impacts on streamflow and stream temperature in the Tennessee River basin. Poster session, *2015 AGU Fall Meeting*, San Francisco, CA.

Mao Y., E. Clark, M. Xiao, B. Nijssen and D. P. Lettenmaier (Apr 2015), Did climate change cause the 2013-2014 California drought? Oral presentation, *2015 Hydrophiles Water Research Symposium*, Corvallis, OR.

Mao Y., E. Clark, M. Xiao, B. Nijssen and D. P. Lettenmaier (Dec 2014), Did climate change cause the 2012-2014 California drought? Oral presentation, *2014 AGU Fall Meeting*, San Francisco, CA.

Mao Y., E. Clark, M. Xiao, B. Nijssen and D. P. Lettenmaier (Oct 2014), The 2014 California drought in an historical context. Oral presentation, *the 39th Annual Climate Diagnostics and Prediction Workshop*, St. Louis, MO.

Mao Y., E. Clark, M. Xiao, B. Nijssen and D. P. Lettenmaier (May 2014), Analysis of the current drought in California, a historical context. Poster session, *the 4th Annual Hydrophiles Water Symposium*, Corvallis, OR.