nature geoscience

ARTICLES

PUBLISHED ONLINE: 16 JANUARY 2017 | DOI: 10.1038/NGEO2868

The global distribution and dynamics of surface soil moisture

Kaighin A. McColl[1,2], Seyed Hamed Alemohammad[1], Ruzbeh Akbar[1], Alexandra G. Konings[1,3], Simon Yueh[4] and Dara Entekhabi[1,5]*

Surface soil moisture has a direct impact on food security, human health and ecosystem function. It also plays a key role in the climate system, and the development and persistence of extreme weather events such as droughts, floods and heatwaves. However, sparse and uneven observations have made it difficult to quantify the global distribution and dynamics of surface soil moisture. Here we introduce a metric of soil moisture memory and use a full year of global observations from NASA's Soil Moisture Active Passive mission to show that surface soil moisture—a storage believed to make up less than 0.001% of the global freshwater budget by volume, and equivalent to an, on average, 8-mm thin layer of water covering all land surfaces—plays a significant role in the water cycle. Specifically, we find that surface soil moisture retains a median 14% of precipitation falling on land after three days. Furthermore, the retained fraction of the surface soil moisture storage after three days is highest over arid regions, and in regions where drainage to groundwater storage is lowest. We conclude that lower groundwater storage in these regions is due not only to lower precipitation, but also to the complex partitioning of the water cycle by the surface soil moisture storage layer at the land surface.

Arguably, one of the most fundamental tasks of hydrologic science is to quantify the distribution of water across global storages (which we term the 'water budget'), and the rates at which water cycles between them (the 'water cycle'). The water cycle is often quantified in terms of a storage's average 'residence time' (the time water spends in the storage, on average, between entering and exiting) or 'memory' (broadly, the time taken for the storage to dissipate a positive or negative anomaly). Many previous studies have provided estimates of these quantities, for storages including the atmosphere, groundwater, glaciers and oceans (for example, ref. 1). At the interface of the lithosphere (which has residence times on the order of months to tens of thousands of years) and atmosphere (which has residence times on the order of days), the surface soil layer (defined here as ∼5-cm deep) displays complex moisture dynamics governed by an enormous range of timescales.

Quantifying the magnitude and dynamics of the surface soil moisture (SSM) storage is essential for many practical reasons. Soil moisture plays an important role in soil microbial respiration[2], biogeochemical cycles[3], streamflow[4], crop yield[5], dust generation[6], and disease transmission[7]. Soil moisture in deeper soil layers can be a more relevant controlling variable for some processes, and can become decoupled from SSM in dry conditions[8]. However, in many cases, SSM is well correlated with soil moisture in deeper layers and little information is lost by focusing exclusively on SSM[9]. The residence times of soil moisture are also important for the prediction of heatwaves, droughts, floods and thunderstorms. This is because soil moisture has considerable memory compared with the atmosphere. An atmospheric anomaly (such as a thunderstorm) will dissipate rapidly (hours), whereas the resulting soil moisture anomaly will take much longer to dissipate (days–months). This anomaly can then modulate or trigger subsequent atmospheric anomalies, suggesting that it may have substantial utility in atmospheric forecasts at seasonal timescales[10].

Many previous studies have estimated soil moisture memory, using models or point-scale observations (for example, refs 11–15). Some studies explicitly estimate a soil moisture memory timescale (for example, ref. 11). Others estimate it implicitly, by estimating the time series autocorrelation (for example, ref. 12) (see Methods: Relation of $F_P(f)$ to other soil moisture memory metrics). Given that soil moisture varies considerably across models[16], it is particularly important to obtain estimates from observations. However, point estimates of soil moisture are too sparse, and coverage is too uneven, to provide a global picture of soil moisture memory. Furthermore, soil moisture memory estimates are dependent on the temporal sampling frequency of the soil moisture observations. Previous large-scale studies estimated memory from approximately monthly observations, missing substantial short-term variability. This missing variability can lead to overestimation of the memory timescale (see Methods: Overestimation of soil moisture memory due to finite sampling frequency in previous studies).

Spaceborne measurements of surface soil moisture

Global observations of SSM are now available from Earth-orbiting satellites. Building on previous missions[17–19], National Aeronautics and Space Administration's (NASA's) Soil Moisture Active Passive (SMAP) satellite mission[20] launched in January 2015. SMAP measures L-band microwave radiation emitted and scattered from the land surface, using a radiometer and radar, respectively. The measured radiation bears the signature of SSM, allowing its retrieval after accounting for the effects of other factors, such as vegetation, surface temperature and surface roughness. The

[1]Department of Civil and Environmental Engineering, MIT, Cambridge, Massachusetts 02139, USA. [2]Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA. [3]Department of Earth System Science, Stanford University, Stanford, California 94305, USA. [4]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA. [5]Department of Earth, Atmospheric and Planetary Sciences, MIT, Cambridge, Massachusetts 02139, USA. *e-mail: darae@mit.edu

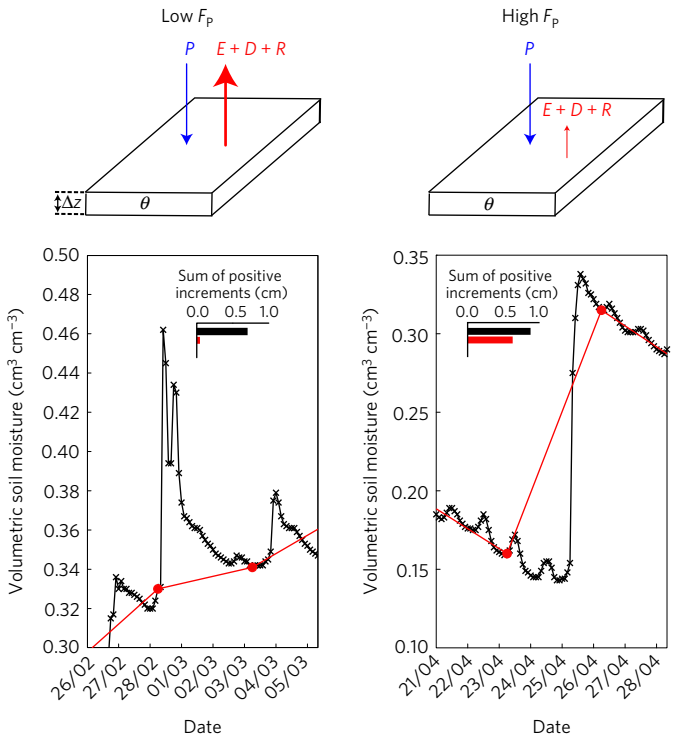

Figure 1 | The stored precipitation fraction. Top left and right: the stored precipitation fraction $F_P(f)$ is a dimensionless measure of the degree to which a soil layer (of depth Δz) retains precipitation inputs (P) over a given timescale $1/f$, given losses due to evapotranspiration (E), drainage (D) and runoff (R). Bottom left and right: two example soil moisture time series from *in situ* observations[30], sampled at two different sampling frequencies — $f = 12\ d^{-1}$ (black, crosses) and $f = 1/3\ d^{-1}$ (red, circles). Inset: the sum of positive increments in the soil water time series, for the two different sampling frequencies.

radar and radiometer measure SSM in different ways, and provide complementary estimates. The radar measured SSM at 3-km spatial resolution, and collected observations for 11 weeks before ceasing operations due to an instrument anomaly. The SMAP radiometer measures SSM at approximately 40-km resolution, and continues to provide high-quality SSM retrievals. In addition to providing global coverage, SMAP's sampling frequency (nominally, $f = 1/3\ d^{-1}$) is substantially higher than that used in any other large-scale, observations-based estimate of soil moisture memory.

In this study, we use the first full annual cycle of SMAP SSM observations to quantify the global distribution and dynamics of SSM. To do this, we introduce a new measure of soil moisture dynamics, the stored precipitation fraction ($F_P(f)$), defined as the average proportion of precipitation falling on a soil layer that is still present in the soil layer after $1/f$ days (Fig. 1). Given a soil moisture time series discretely sampled (with sampling frequency f) over a finite time period, it is calculated as the sum of positive soil water increments, normalized by the total precipitation falling in the time period (see Methods).

Figure 2 shows the global distribution of mean SSM and $F_P(f)$ observed by SMAP, where the nominal sampling frequency for SMAP is $f = 1/3\ d^{-1}$. The global distribution of mean daily precipitation from the Global Precipitation Measurement (GPM) mission is also shown (see Methods). Mean SSM is highest in the tropics (substantially masked out in Fig. 2 due to significant vegetation cover), over the eastern United States, and across northern Europe and Russia; and lowest in deserts (such as those in central Asia, Australia and the Sahara) and in semi-arid regions (such as large parts of the western US). The probability density function (PDF) of SSM peaks around 0.1, but exhibits a heavy tail. These results are broadly consistent with previous global remote-sensing studies (for example, refs 21,22). The SSM dynamics are also quantified in Fig. 2b. Like SSM and mean daily precipitation, the PDF of $F_P(1/3)$ is also heavy-tailed. $F_P(1/3)$ is consistently low in the tropics (again, substantially masked out in Fig. 2 due to significant vegetation cover), where precipitation is often very intense, and both evapotranspiration and drainage fluxes are generally fast on average. This includes southeast Asia and southern China, India and northwestern Australia. It is also low across the eastern United States. In these regions, the terrestrial water cycle at the land surface overturns rapidly, with the vast majority of inflows from precipitation leaving the surface soil layer within three days.

$F_P(1/3)$ is highest in mid-latitudes. In particular, it is high in northern Africa, parts of the Middle East, central Asia and northern China. It is also high in the western United States. In these regions, the water cycle at the interface of the atmosphere and the land surface overturns at a slower rate. On average, the global mean storage of SSM is equivalent to only a thin layer of water over all continents with a depth of merely 8 mm. Yet, on average, 14% of precipitation falling on land remains in this layer after three days. The SSM storage is therefore only a minor component of the global water budget, but plays a relatively much more significant role in the global water cycle. It should be noted that the effective sampling frequency of the SMAP observations differs slightly from the nominal sampling frequency of $1/3\ d^{-1}$ in some regions. However, the effect of these deviations is relatively small (see Methods: Effective SMAP sampling frequency). Measurement noise also induces an estimation bias in $F_P(f)$, which while typically small, can be on the order of 10^{-1} in some dry regions. Other soil moisture memory metrics also suffer from estimation biases (see Methods: Estimation bias induced by measurement noise). The relatively large observed fraction of precipitation remaining in the SSM storage after three days is consistent with a recent study that found satellite-observed SSM can be used to obtain skilful estimates of antecedent precipitation[23].

Determinants of surface soil moisture memory

$F_P(1/3)$ decreases monotonically with increasing mean SSM, albeit with substantial variance at low SSM values (Fig. 3). As SSM increases, both drainage and runoff increase significantly. Regions of low $F_P(1/3)$ are broadly located in regions where SSM is high, on average, and where groundwater recharge[24] and groundwater storage[25] are both largest. In other words, in regions where drainage to groundwater storage is largest, the surface soil layer retains a lower fraction of incoming precipitation; presumably, more of the precipitation rapidly drains to deeper groundwater storages. For drier soils however, hydraulic conductivity (and therefore drainage to groundwater storages) is reduced, as is saturation excess runoff, resulting in a higher fraction of incoming precipitation being retained in the surface soil layer (that is, higher $F_P(1/3)$). These results are in contrast to a previous model-based study, which found soil moisture memory (defined as the 27-day-lagged autocorrelation of the modelled total soil moisture, vertically integrated across the soil profile) is highest in regions with intermediate amounts of soil moisture[12]. While $F_P(1/3)$ appears to be lowest at low soil sand fractions, there is substantial variability, and there do not appear to be clear relations between $F_P(1/3)$, and soil sand and clay fractions (Fig. 3). We also note that estimation biases may contribute to the larger $F_P(1/3)$ in drier regions (see Methods: Estimation bias induced by measurement noise). Overall, these results demonstrate that, speaking broadly at global scales, lower groundwater storage in drier regions is due not only to lower precipitation, but also to greater partitioning of the water cycle by the SSM storage. Hence, while SSM is only a small component

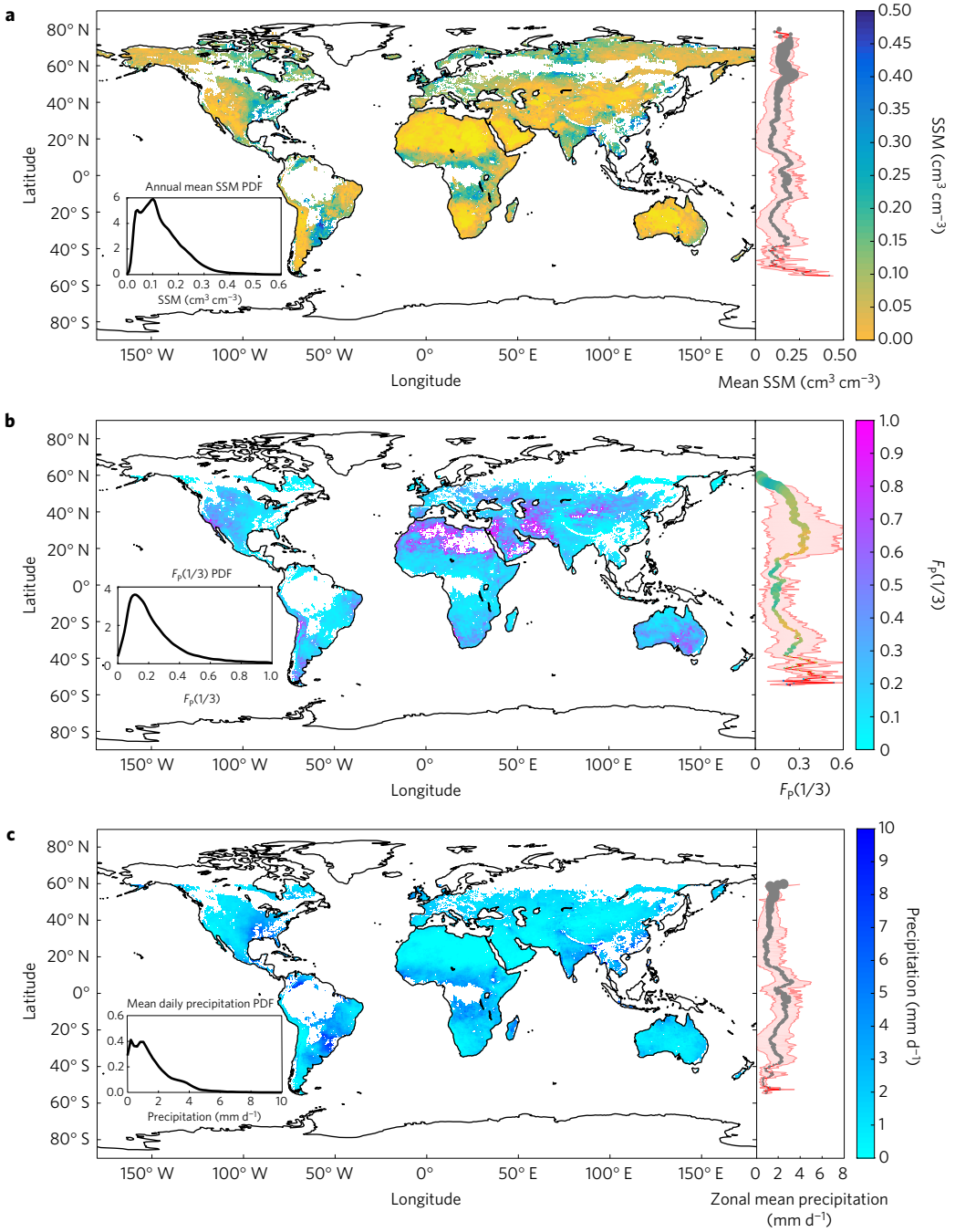

Figure 2 | Global distribution and memory of surface soil moisture. a, Global map of annual mean SSM (1 April 2015–31 March 2016), with PDF (inset) and zonal mean (right panel). Marker sizes in all zonal plots are proportional to zonal land area; shaded region shows ±1 standard deviation. White regions in map are missing or masked (see Methods). **b**, The same as in **a**, except for $F_P(1/3)$, and the marker colours in the zonal plot indicate the zonal mean SSM. **c**, The same as in **a**, for mean daily precipitation.

of the global water budget (making up less than 0.001% of the global freshwater budget by volume[1]), it plays a substantial role in partitioning water between storages.

These analyses provide the first global estimates—derived from observations rather than models—of both the magnitude and dynamics of the SSM storage. In most regions, $F_P(1/3)$ is less than or equal to 14%, and decreases with increasing SSM. While this is large relative to the volume stored in SSM, many previous studies estimate a soil moisture memory timescale on the order of several months, which seems inconsistent with the results seen here. This result is further confirmed by repeating the analysis on SMAP observations thinned to lower sampling frequencies (Supplementary Figs 5–8): in particular, estimated $F_P(1/30)$ rarely exceeds 5% (Supplementary Fig. 8), and the normalized difference $(F_P(1/3) - F_P(1/30))/(F_P(1/3))$ is close to one in most parts of the world (Supplementary Fig. 9). While a memory timescale on the order of months may be accurate for deeper soil layers (~1-m), to our knowledge, the only observational study looking at shallower layers (10 cm) at several mid-latitude sites also yielded an estimate on the order of months[11]. It seems more likely, therefore, that this difference is largely due to the higher-frequency SMAP observations used in this study, which resolve a greater fraction of the soil

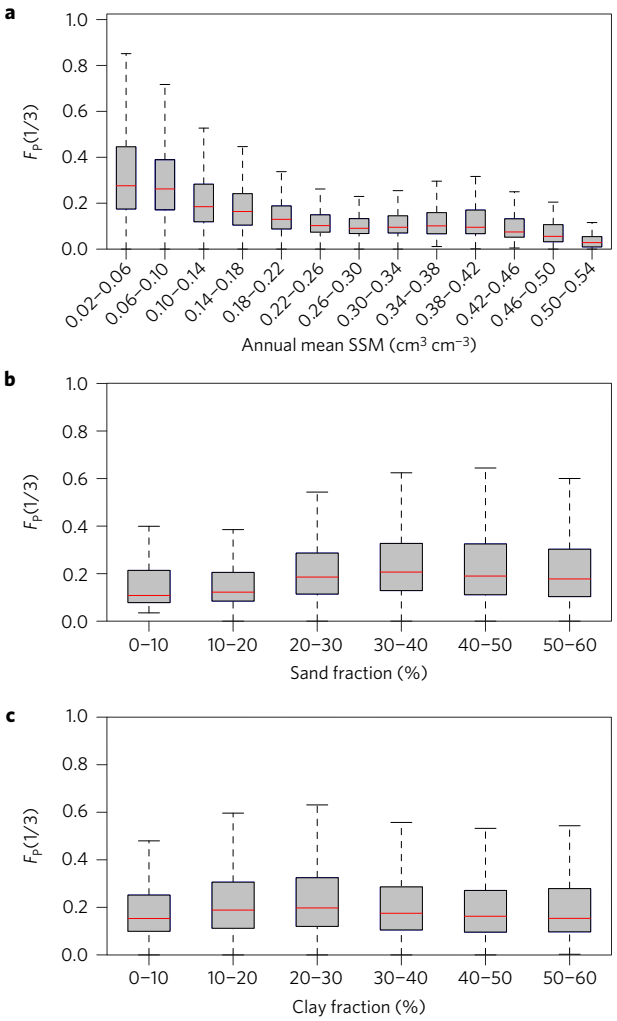

Figure 3 | Global relations between stored precipitation fraction and soil moisture content and texture. a, Global relation between $F_P(1/3)$ and annual mean SSM, estimated using one year of observations (1 April 2015–31 March 2016). Boxplots show the median (red horizontal line), 25th and 75th percentiles (top and bottom of the grey shaded box, respectively), and maximum and minimum observed values (edges of the top and bottom whiskers, respectively). **b**, Global relation between $F_P(1/3)$ and sand fraction. **c**, Global relation between $F_P(1/3)$ and clay fraction.

moisture temporal variability, and therefore allow more accurate estimates of its memory. The additional variability resolved by higher-frequency observations will, by definition, lead to a shorter memory timescale. In addition, we focus on positive anomalies typically associated with faster-dissipating rainfall pulses.

Implications for land–atmosphere interactions

Beyond the fundamental importance of characterizing the magnitude and response timescales of Earth's water storages, a key application of these results is in identifying regions with strong land–atmosphere coupling—recognizing that we focus here only on the SSM-controlled part of this coupling and cannot assess the fraction of stored precipitation in regions with dense vegetation coverage. Significant soil moisture memory is a necessary condition for land–atmosphere feedbacks[26]. However, soil moisture–precipitation feedbacks can occur at both high or low soil moisture values, and are driven by different mechanisms[27]. This analysis focuses on the memory of wet SSM anomalies, which typically have a stronger influence on precipitation forecast skill than dry anomalies[28]. Our results are consistent with a recent study that found significant positive soil moisture–precipitation feedbacks in the western United States[29], where we also find relatively high $F_P(1/3)$. These results may therefore have particularly important implications for short-term weather forecasting of extreme precipitation events and floods globally, in regions where these are strongly controlled by SSM rather than root-zone soil moisture or other processes. Furthermore, since the stored precipitation fraction can be readily estimated from land surface model outputs (for models containing a discrete surface soil moisture layer), comparing SMAP-estimated and model-estimated $F_P(f)$ will provide a useful test of model fidelity to global water cycle dynamics.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of this paper.

Received 27 September 2016; accepted 2 December 2016; published online 16 January 2017

References

1. Oki, T., Entekhabi, D. & Harrold, T. I. in *The State of the Planet: Frontiers and Challenges in Geophysics* Vol. 19 (eds Stephan, R., Spark, J. & Hawkesworth, C. J.) 225–237 (Geophysical Monograph 150, American Geophysical Union, 2004).
2. Manzoni, S., Schimel, J. P. & Porporato, A. Responses of soil microbial communities to water stress: results from a meta-analysis. *Ecology* **93,** 930–938 (2012).
3. D'Odorico, P., Laio, F., Porporato, A. & Rodriguez-Iturbe, I. Hydrologic controls on soil carbon and nitrogen cycles. II. A case study. *Adv. Water Resour.* **26,** 59–70 (2003).
4. Botter, G., Peratoner, F., Porporato, A., Rodriguez-Iturbe, I. & Rinaldo, A. Signatures of large-scale soil moisture dynamics on streamflow statistics across US climate regimes. *Wat. Resour. Res.* **43,** W11413 (2007).
5. Rosenzweig, C., Tubiello, F. N., Goldberg, R., Mills, E. & Bloomfield, J. Increased crop damage in the US from excess precipitation under climate change. *Glob. Environ. Change* **12,** 197–202 (2002).
6. Fécan, F., Marticorena, B. & Bergametti, G. Parametrization of the increase of the aeolian erosion threshold wind friction velocity due to soil moisture for arid and semi-arid areas. *Ann. Geophys.* **17,** 149–157 (1999).
7. Bomblies, A. & Eltahir, E. A. B. Assessment of the impact of climate shifts on malaria transmission in the Sahel. *EcoHealth* **6,** 426–437 (2010).
8. Hirschi, M., Mueller, B., Dorigo, W. & Seneviratne, S. I. Using remotely sensed soil moisture for land–atmosphere coupling diagnostics: the role of surface vs. root-zone soil moisture variability. *Remote Sens. Environ.* **154,** 246–252 (2014).
9. Qiu, J., Crow, W. T. & Nearing, G. S. The impact of vertical measurement depth on the information content of soil moisture for latent heat flux estimation. *J. Hydrometeorol.* **19,** 2419–2430 (2016).
10. Entekhabi, D., Rodriguez-Iturbe, I. & Bras, R. L. Variability in large-scale water balance with land surface-atmosphere interaction. *J. Clim.* **5,** 798–813 (1992).
11. Entin, J. K. *et al*. Temporal and spatial scales of observed soil moisture variations in the extratropics. *J. Geophys. Res.* **105,** 11865–11877 (2000).
12. Seneviratne, S. I. *et al*. Soil moisture memory in AGCM simulations: analysis of global land–atmosphere coupling experiment (GLACE) data. *J. Hydrometeorol.* **7,** 1090–1112 (2006).
13. Katul, G. G. *et al*. On the spectrum of soil moisture from hourly to interannual scales. *Wat. Resour. Res.* **43,** W05428 (2007).
14. Orth, R. & Seneviratne, S. I. Analysis of soil moisture memory from observations in Europe. *J. Geophys. Res.* **117,** D15115 (2012).
15. Koster, R. D. & Suarez, M. J. Soil moisture memory in climate models. *J. Hydrometeorol.* **2,** 558–570 (2001).
16. Koster, R. D. *et al*. On the nature of soil moisture in land surface models. *J. Clim.* **22,** 4322–4335 (2009).
17. Kerr, Y. H. *et al*. The SMOS mission: new tool for monitoring key elements of the global water cycle. *Proc. IEEE* **98,** 666–687 (2010).
18. Njoku, E. G., Jackson, T. J., Lakshmi, V., Chan, T. K. & Nghiem, S. V. Soil moisture retrieval from AMSR-E. *IEEE Trans. Geosci. Remote Sensing* **41,** 215–229 (2003).
19. Figa-Saldaña, J. *et al*. The advanced scatterometer (ASCAT) on the meteorological operational (MetOp) platform: a follow on for European wind scatterometers. *Can. J. Remote Sensing* **28,** 404–412 (2002).

20. Entekhabi, D. *et al*. The Soil Moisture Active Passive (SMAP) mission. *Proc. IEEE* **98,** 704–716 (2010).
21. Albergel, C. *et al*. Monitoring multi-decadal satellite earth observation of soil moisture products through land surface reanalyses. *Remote Sens. Environ.* **138,** 77–89 (2013).
22. McColl, K. A., Entekhabi, D. & Piles, M. Uncertainty analysis of soil moisture and vegetation indices using Aquarius scatterometer observations. *IEEE Trans. Geosci. Remote Sensing* **52,** 4259–4272 (2014).
23. Koster, R. D., Brocca, L., Crow, W. T., Burgin, M. S. & De Lannoy, G. J. M. Precipitation estimation using L-band and C-band soil moisture retrievals: precipitation estimation from soil moisture retrievals. *Wat. Resour. Res.* **52,** 7213–7225 (2016).
24. Döll, P. & Fiedler, K. Global-scale modeling of groundwater recharge. *Hydrol. Earth Syst. Sci.* **12,** 863–885 (2008).
25. Gleeson, T., Befus, K. M., Jasechko, S., Luijendijk, E. & Cardenas, M. B. The global volume and distribution of modern groundwater. *Nat. Geosci.* **9,** 161–167 (2015).
26. Koster, R. D. & Suarez, M. J. Impact of land surface initialization on seasonal precipitation and temperature prediction. *J. Hydrometeorol.* **4,** 408–423 (2003).
27. Gentine, P., Holtslag, A. A. M., D'Andrea, F. & Ek, M. Surface and atmospheric controls on the onset of moist convection over land. *J. Hydrometeorol.* **14,** 1443–1462 (2013).
28. Koster, R. D. *et al*. The second phase of the global land–atmosphere coupling experiment: soil moisture contributions to subseasonal forecast skill. *J. Hydrometeorol.* **12,** 805–822 (2011).
29. Tuttle, S. & Salvucci, G. Empirical evidence of contrasting soil moisture-precipitation feedbacks across the United States. *Science* **352,** 825–828 (2016).
30. Cuenca, R. H., Hagimoto, Y. & Moghaddam, M. Three-and-a-half decades of progress in monitoring soils and soil hydraulic properties. *Proc. Environ. Sci.* **19,** 384–393 (2013).

Acknowledgements

K.A.M. is funded by a National Science Foundation Graduate Research Fellowship and a Ziff Environmental Fellowship from Harvard University's Center for the Environment. The parts of this work performed by the Massachusetts Institute of Technology and by the Jet Propulsion Laboratory, California Institute of Technology were conducted under contracts with the National Aeronautics and Space Administration. The authors thank S. Seneviratne for comments on earlier drafts of the manuscript.

Author contributions

K.A.M. wrote the manuscript. R.A., K.A.M. and S.H.A. conducted analyses and produced figures. D.E. conceived and led the project, and developed the 'stored precipitation fraction' in discussions with K.A.M., S.H.A. and A.G.K. S.Y. contributed to interpretation of the results. All authors discussed and edited drafts of the manuscript.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.E.

Competing financial interests

The authors declare no competing financial interests.

Methods

Data sets. SSM estimates are obtained from the NASA Soil Moisture Active Passive (SMAP) mission[20]. Launched in January 2015, SMAP currently produces radiometer-only global SSM values at approximately 40-km spatial resolution (−3 dB half-power radiometer beamwidth but posted at 36 km) with an approximate three-day revisit cycle (see Supplementary Fig. 4 for a global map of the effective revisit time). The SMAP penetration depth is approximately 5 cm (ref. 31) (but varies subtly with soil moisture content). A full year of global SSM maps spanning 1 April 2015–31 March 2016 are used in this study[32]. Regions where vegetation water content is greater than 5 kg m^{-2}, where the soil is frozen, or where soil moisture retrievals are substantially contaminated by radio frequency interference or the presence of small water bodies are excluded from the analysis. While observations from other soil moisture satellite missions could also be used, differences in revisit times—and therefore, sampling frequencies—complicate the comparison. Therefore, for simplicity, we use only SMAP observations in this study. Preliminary validation studies of SMAP radiometer soil moisture retrievals demonstrate that they are, on average, within the mission target accuracy of 0.04 m^3 m^{-3} unbiased root-mean-square difference[33].

Precipitation estimates are obtained from the NASA Global Precipitation Mission (GPM), at a spatial resolution of 0.1°. Observations are available in the latitudinal range ±60° (the coverage of the GPM orbit). Where available, either half-hourly Final-Run[34] (1 April 2015–31 December 2015) or Late-Run (1 January 2016–1 April 2016) Integrated Multi-satellitE Retrievals for GPM (IMERG) are used to cover the same temporal period as that of the SMAP SSM observations. GPM precipitation data are re-gridded to the EASE 2.0 grid projection. Regions where annual precipitation is zero, or where estimated $F_P(f) > 1$, are excluded from the analysis.

Soil clay and sand fraction data are obtained from a global composite of soil texture data sets prepared for the SMAP mission[35].

The results in this study are estimated from one year of data, which is a large enough sample to broadly characterize the global climatology, while recognizing that in some regions, the results will deviate from the climatological mean.

Definition of stored precipitation fraction. Surface soil moisture (SSM) is defined as the volumetric water content of a soil sample $\theta = V_w/(V_s + V_w + V_a)$, where V_w is the volume of water, V_s is the volume of solids and V_a is the volume of air. It is bounded above by the volumetric soil porosity, $\phi = (V_w + V_a)/(V_s + V_w + V_a)$. The water mass balance equation for a surface layer of soil spanning a depth of Δz (m) is

$$\Delta z \frac{d\theta}{dt} = P - E - R - D$$

where P is precipitation rate (m d^{-1}), E is evapotranspiration (including interception losses) (m d^{-1}), R is surface and subsurface lateral runoff (m d^{-1}) and D is drainage into deeper layers (m d^{-1}). P and E are always positive. R can be positive (runoff) or negative (run-on). Similarly, D can be positive (drainage to deeper layers) or negative (wetting from capillary rise).

We start with a SSM time series consisting of discrete samples θ_i with sample index i. For simplicity, we assume θ_i has a constant sampling frequency f (d^{-1}) over a sampling period of length T (d). The stored precipitation fraction $F_P(f)$ is defined as

$$F_P(f) = \frac{\Delta z \sum_{i=1}^{fT} \Delta\theta_{i+}}{\int_0^T P(t)dt} \quad (1)$$

where

$$\Delta\theta_{i+} = \begin{cases} \Delta\theta_i, & \text{if } \Delta\theta_i > 0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where $\Delta\theta_i = \theta_i - \theta_{i-1}$, and $\int_0^T P(t)dt$ is the accumulated precipitation in the time period (m).

We assume the SSM increment $\Delta\theta$ is positive only if precipitation occurs ($P > 0$). This will not be true in regions where substantial soil wetting occurs due to capillary rise from deeper layers (negative D) or run-on (negative R). However, at the large spatial scales observed by SMAP (approximately 40 km), very few pixels are expected to show positive $\Delta\theta$ in the absence of P due to these mechanisms. With this assumption in place, the stored precipitation fraction measures the proportion of the precipitation input flux that enters the SSM storage and remains there after a time interval $1/f$. For SMAP, $f = 1/3$ d^{-1}, although the effective sampling frequency can differ slightly in some regions due to missing data (lower f) and satellite half-orbit overlap in the extra-tropics (higher f). See Supplementary Fig. 4. Over the time interval $1/f$, some of the precipitation input will be lost to E, R and D. The magnitudes of these losses depend on weather, geology and the time interval f itself.

If $E + D + R$ is, on average, high relative to P over the timescale $1/f$ (Fig. 1, top left), then $F_P(f)$ will be low (∼0). If $E + D + R$ is, on average, low relative to P over the timescale $1/f$ (Fig. 1, top right), then $F_P(f)$ will be high (∼1). As the sampling frequency decreases, $F_P(f)$ will decrease, since the variability resolved by the time series decreases (Fig. 1, bottom left and right).

Relation of $F_P(f)$ to other soil moisture memory metrics. Soil moisture memory is loosely defined as the time taken by the soil to 'forget' an anomaly (caused, for instance, by atypically heavy rainfall, or atypically dry conditions[15]). It has been estimated using both models[12,15,36–38] and point-scale observations[11,13,14,39–43]. Many metrics exist for quantifying soil moisture memory.

Soil moisture memory metrics may typically differ from one another in three respects. First, the metric is often defined as either a timescale over which soil moisture anomalies are largely dissipated; or as a degree of correlation between anomalies over a fixed time lag (or equivalently, a fixed sampling frequency). Second, the soil moisture anomaly is defined as a deviation from a reference state, which may be an annual mean, monthly mean, time-varying trend, or other value. Third, the metric may consider only positive anomalies (that is, wetting events), consider only negative anomalies (that is, drying events) or blend both positive and negative anomalies without distinguishing between their sign. Several relevant metrics, including the stored precipitation fraction, are compared with respect to these three properties in Supplementary Table 1 and discussed further below.

Several metrics are based on the soil moisture time series' autocorrelation. The autocorrelation $\rho(t)$ is the correlation between any two points in the time series, separated by a time interval t. Typically, it is assumed that the time series is stationary, so $\rho(t)$ does not vary in time. By definition, the mean soil moisture state is removed when estimating $\rho(t)$, so it is implicitly a measure of correlation between anomalies, where here anomalies are defined as deviations from a mean soil moisture state. The autocorrelation ignores the sign of the anomaly, so any soil moisture memory metric based on $\rho(t)$ cannot distinguish between positive and negative anomalies. Two examples of soil moisture memory timescales that may be estimated from $\rho(t)$ are the e-folding[11,36,41] and integral timescales[13,39] (Supplementary Table 1). Both timescales increase with increasing $\rho(t)$. Rather than estimating an anomaly dissipation timescale, some authors use an estimate of $\rho(t)$ itself, evaluated at a fixed value of t (for example, one month), as a soil moisture memory metric[12,15]. This can be generalized to estimating $\rho(t)$ for a range of values of t, in which case $\rho(t)$ is often Fourier-transformed to obtain the soil moisture variance spectrum $E_s(f)$, where f is frequency, corresponding to an inverse time lag[13,40].

While these metrics have enjoyed considerable success, we highlight two weaknesses. First, by defining an anomaly as a deviation from a mean soil moisture state, the anomaly becomes dependent on the estimate of the mean state; and estimates of the mean state can be highly uncertain. For instance, for monthly soil moisture observations, the monthly mean soil moisture state might be estimated from its historical average. However, in many cases, the historical record will be too short to do this without introducing significant sampling error. This problem is further exacerbated when using daily or hourly observations, where the increased soil moisture variability demands an even longer historical record to precisely estimate the mean soil moisture state at a daily or hourly timestep. Therefore, in practice, soil moisture memory metrics based on autocorrelation will be non-trivially dependent on the estimated mean soil moisture state[39]. This is particularly true when applied to observations with a sub-monthly temporal resolution. Second, the sign of the soil moisture anomaly provides physically meaningful information that is ignored by autocorrelation-based metrics. At large spatial scales, positive spikes in soil moisture are almost always caused by rapid, essentially stochastic precipitation events (with some rare exceptions). In contrast, negative anomalies are caused by slower processes (such as evapotranspiration), which are quasi-deterministic[39]. It is therefore useful to quantify the anomaly dissipation timescales of the rapid and slow processes separately.

To avoid these problems, metrics that move beyond the time series autocorrelation may be useful. One such metric is the mean persistence time[39], defined as the mean time the soil moisture time series spends above or below a fixed soil moisture threshold (Supplementary Table 1). The soil moisture threshold can be any value of interest, for example, wilting point soil moisture. Anomalies are then defined as deviations from this fixed threshold. A key benefit of this approach is that positive and negative anomalies are considered separately. However, the mean persistence time is still dependent on the choice of fixed threshold.

In contrast to previous metrics, the stored precipitation fraction $F_P(f)$ is not dependent on the choice of soil moisture reference state used to define anomalies. This is because anomalies are defined relative to the soil moisture state prior to the last precipitation event. Furthermore, positive and negative increments are considered separately. The stored precipitation fraction is a frequency-dependent measure of soil moisture variability, so it is natural to compare it to the spectrum of soil moisture variance $E_s(f)$. More precisely, the sum of positive increments (used in estimating $F_P(f)$) is a measure of the variability of soil moisture at frequencies between $1/T$ and f, so it is analogous to the integral of the soil moisture variance spectrum, integrated between $1/T$ and f. If $\Delta\theta_i$ was a Gaussian random variable, then $F_P(f)$ would be proportional to the soil moisture standard deviation. However,

the distribution of soil moisture increments is typically skewed and, therefore, non-Gaussian. This is a signature of the physically meaningful differences between positive and negative soil moisture anomalies.

Overestimation of soil moisture memory due to finite sampling frequency in previous studies. Previous studies typically used soil moisture time series with an approximately monthly sampling frequency. There are some exceptions[11,13,14,39,40], but the difficulty of maintaining long-term, high-frequency soil moisture observations means these are constrained to a handful of locations. Here, we show that a classical estimate of soil moisture memory from monthly observations will be typically overestimated.

Most previous estimates of soil moisture memory are based on an estimate of the time series autocorrelation $\rho(t)$, defined as

$$\rho(t) = \frac{R(t)}{R(0)}$$

where $R(t) = \mathrm{Cov}(\theta_i, \theta_{i+t})$ is the autocovariance of the soil moisture time series θ_i. Given soil moisture observations with a sampling frequency f d^{-1}, the smallest lag resolvable by the time series is $t = 1/f$. However, substantial autocovariance may exist at shorter time lags $t < 1/f$. Furthermore, an estimate of $R(0)$ is required to estimate $\rho(t)$. Previous studies typically estimate $R(t)$ by fitting a curve to observations for $t \geq 1/f$, and then use the fitted curve to project onto the interval $0 \leq t < 1/f$. However, this method is heuristic and often substantially underestimates $R(0)$, particularly in drier areas, resulting in a positively biased estimate of $\rho(t)$. Therefore, any soil moisture memory timescale τ estimated from $\rho(t)$ will also be overestimated. An example of this is illustrated in Supplementary Fig. 1. Acknowledging this problem, the authors of ref. 11 distinguished between two different soil moisture timescales: the (longer) 'meteorological scale', controlled by precipitation and evaporative demand, and the (shorter) 'land surface scale', controlled by soil characteristics, topography and vegetation. They estimated the 'meteorological' timescale from observations, and treated the 'land surface' timescale as white noise.

Previous estimates of soil moisture memory are, therefore, dependent on the sampling frequency f and are likely to be substantially overestimated, especially when estimated from monthly observations. While the stored precipitation fraction is also sampling frequency-dependent, it explicitly includes this dependence in its definition and interpretation.

Estimation bias induced by measurement noise. Like other soil moisture memory metrics, in the presence of measurement noise, $F_P(f)$ is a biased estimator. In this section, we characterize the bias induced by measurement noise on estimates of the stored precipitation fraction. We focus the analysis on random errors since fixed biases will be automatically subtracted out in the process of estimating temporal increments. Due to measurement noise from various sources, we observe dimensionless volumetric water content $\hat{\theta}(t) = \theta(t) + \varepsilon(t)$ over some time interval $[0, T]$ days. The noise $\varepsilon(t)$ is modelled as a mean-zero Gaussian random variable, while the true soil moisture $\theta(t)$ is treated as deterministic. We assume the noise is uncorrelated in time and has constant variance σ^2 $(-)$.

For simplicity, start by considering the distribution of a single soil moisture increment $\Delta\hat{\theta}_i = \hat{\theta}_i - \hat{\theta}_{i-1} = \Delta\theta_i + e_i$, where $e_i = \varepsilon_i - \varepsilon_{i-1}$ is a mean-zero Gaussian random variable with a constant variance of $2\sigma^2$. Now threshold $\Delta\hat{\theta}_i$ so that any negative values are forced to be zero. The PDF of the thresholded increment $\Delta\hat{\theta}_{i+}$ is

$$f(\Delta\hat{\theta}_{i+}|\Delta\theta_i,\sigma) = \begin{cases} \dfrac{1}{2\sigma\sqrt{\pi}} \exp\left(\dfrac{-(\Delta\hat{\theta}_{i+} - \Delta\theta_i)^2}{4\sigma^2}\right), & \text{if } \Delta\theta_{i+} > 0 \\ \Phi\left(0|\Delta\theta_i, 2\sigma^2\right) + \dfrac{1}{2\sigma\sqrt{\pi}} \exp\left(\dfrac{-\Delta\theta_i^2}{4\sigma^2}\right), & \text{if } \Delta\theta_{i+} = 0 \\ 0, & \text{if } \Delta\theta_{i+} < 0 \end{cases}$$

where $\Phi(0|\Delta\theta_i, 2\sigma^2)$ is the cumulative density function for a normal random variable with mean $\Delta\theta_i$ and variance $2\sigma^2$, evaluated at zero. Two example PDFs are shown in Supplementary Fig. 2. When $\Delta\theta_i \ll 0$, almost all the density is at zero, reflecting the unlikeliness of random noise causing the increment to become positive if it is highly negative. When $\Delta\theta_i \gg 0$, the distribution is essentially Gaussian and unaffected by thresholding.

Since $F_P(f)$ is just a linear combination of $\Delta\theta_{i+}$, the noise bias in $F_P(f)$ can be estimated from the individual biases in $\Delta\theta_{i+}$. The expected value of $\Delta\theta_{i+}$ is

$$E\left(\Delta\hat{\theta}_{i+}\right) = \int_{-\infty}^{\infty} x f(x)\mathrm{d}x = \int_0^{\infty} \frac{x}{2\sigma\sqrt{\pi}} \exp\left(\frac{-(x - \Delta\theta_i)^2}{4\sigma^2}\right) \mathrm{d}x$$

Solving the integral, we obtain

$$E(\Delta\hat{\theta}_{i+}) = \frac{\sigma}{\sqrt{\pi}} \exp\left(-\Delta\theta_i^2/4\sigma^2\right) + \frac{\Delta\theta_i}{2}\left(1 + \mathrm{erf}\left(\Delta\theta_i/2\sigma\right)\right) \quad (3)$$

Using equations (2) and (3), the noise bias in $\Delta\hat{\theta}_{i+}$ is

$$\begin{aligned} \delta_i &= E(\Delta\hat{\theta}_{i+}) - \Delta\theta_{i+} \\ &= \frac{\sigma}{\sqrt{\pi}} \exp\left(-|\Delta\theta_i|^2/4\sigma^2\right) - \frac{|\Delta\theta_i|}{2}\left(1 + \mathrm{erf}\left(-|\Delta\theta_i|/2\sigma\right)\right) \end{aligned} \quad (4)$$

The bias in $F_P(f)$ is

$$E\left(F_P(f)\right) - \overline{F_P(f)} = \frac{\Delta z \sum_{i=1}^{fT} E(\Delta\hat{\theta}_{i+})}{P} - \frac{\Delta z \sum_{i=1}^{fT} \Delta\theta_{i+}}{P} = \frac{\Delta z \sum_{i=1}^{fT} \delta_i}{P} \quad (5)$$

which can be evaluated using equation (4), for a general case with prescribed soil moisture increments and measurement noise.

The bias in a single increment (δ_i) is plotted in Supplementary Fig. 3, for varying $\Delta\theta_i$ and σ. The bias is largest when the soil moisture increment is small, and increases with increasing σ. For a fixed σ, it drops rapidly with increasing/decreasing $\Delta\theta_i$ away from zero.

Further insight can be gained by examining a special case. For the case where the true soil moisture is constant in time, $\Delta\theta_i = 0$. Therefore, from equations (4) and (5),

$$E\left(F_P(f)\right) - \overline{F_P(f)} = \frac{\Delta z \sum_{i=1}^{fT} \frac{\sigma}{\sqrt{\pi}}}{P} = \frac{\Delta z f T \sigma}{P\sqrt{\pi}} = \frac{\Delta z f \sigma}{\bar{P}\sqrt{\pi}}$$

where $\bar{P}$ is the average daily precipitation. The bias therefore increases linearly with sampling frequency f, sampling depth Δz, and noise standard deviation σ. It will be most pronounced in dry regions with low $\bar{P}$. The $\Delta\theta_i = 0$ case is a worst-case scenario since, for a given σ, δ_i is maximized when $\Delta\theta_i = 0$. Hence, this is an upper bound on the bias in $F_P(f)$ due to measurement noise. A typical value for this upper bound when applied to SMAP data over a semi-arid region (using $\bar{P} = 1$ mm d^{-1}, $\Delta z = 50$ mm, $f = 1/3$ d^{-1} and $\sigma = 0.04$ $(-)$) is $E\left(F_P(f)\right) - \overline{F_P(f)} \sim 0.4$. We emphasize that this is a loose upper bound on the bias, rather than an estimate of the bias itself. This upper bound is unlikely to be reached, since it would require a perfectly constant soil moisture time series. Furthermore, as evident from Supplementary Fig. 3, the bias drops rapidly as the soil moisture increment moves away from zero (either increasing or decreasing). We note that estimation biases induced by measurement noise are not unique to the stored precipitation fraction. For instance, the estimated autocorrelation will also be biased in the presence of measurement noise[44,45], meaning other metrics used in previous studies would face similar challenges when applied to satellite observations.

In summary, the bias is most significant in dry regions with little precipitation, but is generally expected to be low in most areas for SMAP observations. In this study, rather than attempting to remove the bias (which would require modelling of the observation noise and the true soil moisture time series globally, introducing substantial new sources of error into the analysis), we simply note that it is generally small but may occasionally be on the order of 10^{-1} in some dry areas.

Effective SMAP sampling frequency. SMAP sometimes does not retrieve soil moisture, for instance, if the soil is frozen, if there is significant radiofrequency interference, or if vegetation water content exceeds 5 kg m^{-2}. Hence, in some regions, the effective SMAP sampling frequency will be lower than the nominal value determined by the SMAP polar orbit and swath coverage ($f = 1/3$ d^{-1}). In others, it is higher than the nominal sampling frequency, due to overlap in the satellite descending half-orbits. In this case, the SMAP observations are undersampled to ensure $f \leq 1/3$ d^{-1}. The effective SMAP sampling frequency is shown in Supplementary Fig. 4.

In cases where the effective sampling frequency is lower than the nominal value, we expect $F_P(f)$ may be systematically lower, since it is a function of f. All else being equal, observed soil moisture variability will be lower at lower sampling frequencies (Fig. 1). Decreasing f will also decrease the estimation bias in $F_P(f)$, which decreases with f. Comparing Supplementary Fig. 4 with Fig. 2 confirms that prominent regions where f is considerably lower than the nominal value—such as the Tibetan Plateau, northeastern China and Russia, and patches of the US Midwest—correspond to lower stored precipitation fractions. However, overall, the correlation between sampling frequency and $F_P(f)$ is weak, with a linear relation between the two variables explaining less than 10% of the observed variance.

Data availability. The data used in this study are publicly available. SMAP SSM observations are available from http://dx.doi.org/10.5067/7EW92T1NI4M4. The Final-Run and Late-Run Integrated Multi-satellitE Retrievals for GPM (IMERG) are available from http://dx.doi.org/10.5067/GPM/IMERG/HH/3B and http://disc.gsfc.nasa.gov/datacollection/GPM_3IMERGHHL_V03.html, respectively. Soil texture data sets are available on request from N. Das (nndas@jpl.nasa.gov).

References

31. Monerris, A. *et al*. *IEEE MicroRad 2006* 171–175 (IEEE, 2006).
32. O' Neill, P. E., Chan, S., Njoku, E. G., Jackson, T. & Bindlish, R. *SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Version 2* (NASA National Snow and Ice Data Center Distributed Active Archive Center, 2016).
33. Chan, S. K. *et al*. Assessment of the SMAP passive soil moisture product. *IEEE Trans. Geosci. Remote Sensing* **54,** 4994–5007 (2016).
34. Huffman, G. *GPM Level 3 IMERG Half Hourly 0.1 × 0.1 Degree Precipitation, version 03* (Goddard Space Flight Center Distributed Active Archive Center (GSFC DAAC), 2015).
35. Das, N. *SMAP Ancillary Data Report: Soil Attributes* (Jet Propulsion Laboratory, California Institute of Technology, accessed September 2016, 2013); http://smap.jpl.nasa.gov/files/smap2/044_soil_attrib.pdf
36. Delworth, T. L. & Manabe, S. The influence of potential evaporation on the variabilities of simulated soil wetness and climate. *J. Clim.* **1,** 523–547 (1988).
37. Delworth, T. & Manabe, S. The influence of soil wetness on near-surface atmospheric variability. *J. Clim.* **2,** 1447–1462 (1989).
38. Wang, A., Zeng, X., Shen, S. S. P., Zeng, Q.-C. & Dickinson, R. E. Time Scales of Land Surface Hydrology. *J. Hydrometeorol.* **7,** 868–879 (2006).
39. Ghannam, K. *et al*. Persistence and memory timescales in root-zone soil moisture dynamics. *Wat. Resour. Res.* **52,** 1427–1445 (2016).
40. Nakai, T. *et al*. Radiative and precipitation controls on root zone soil moisture spectra. *Geophys. Res. Lett.* **41,** 7546–7554 (2014).
41. Vinnikov, K. Y. & Yeserkepova, I. B. Soil moisture: empirical data and model results. *J. Clim.* **4,** 66–79 (1991).
42. Vinnikov, K. Y., Robock, A., Speranskaya, N. A. & Schlosser, C. A. Scales of temporal and spatial variability of midlatitude soil moisture. *J. Geophys. Res.* **101,** 7163–7174 (1996).
43. Wu, W., Geller, M. A. & Dickinson, R. E. The response of soil moisture to long-term variability of precipitation. *J. Hydrometeorol.* **3,** 604–613 (2002).
44. McColl, K. A. *et al*. Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target. *Geophys. Res. Lett.* **41,** 2014GL061322 (2014).
45. Crow, W. T. *et al*. Robust estimates of soil moisture and latent heat flux coupling strength obtained from triple collocation. *Geophys. Res. Lett.* **42,** 2015GL065929 (2015).