

The Seasonal Cycle of Surface Soil Moisture

MASON O. STAHL^a AND KAIGHIN A. MCCOLL^{b,c}

^a *Department of Geosciences, Union College, Schenectady, New York*

^b *Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts*

^c *School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts*

(Manuscript received 2 October 2021, in final form 15 March 2022)

ABSTRACT: The seasonal cycle contributes substantially to soil moisture temporal variability in many parts of the world, with important implications for seasonal forecasting relevant to agriculture and the health of humans and ecosystems. There is considerable spatial variability in the seasonal cycle of soil moisture, yet a lack of global observations has hindered the development of parsimonious theories explaining that variability. Here, we use 6 years of global satellite observations to describe and explain the seasonal cycle of surface soil moisture globally. An unsupervised clustering algorithm is used to identify five distinct seasonal cycle regimes. Each seasonal cycle regime typically arises in both hemispheres, on multiple continents, and across substantially different local climates. To explain this spatial variability, we then show that the observed seasonal cycle regimes are reproduced very well by a simple but physically based water balance model, which only uses precipitation and downwelling surface shortwave radiation as inputs, and includes no free parameters. Surprisingly, no information on vegetation or land cover is required. To our knowledge, this is the first characterization of the seasonal cycle of surface soil moisture based on global observations.

KEYWORDS: Atmosphere-land interaction; Hydrologic cycle; Hydrology; Soil moisture; Remote sensing; Seasonal cycle; Clustering

1. Introduction

Over oceans, sea surface temperature is the most important surface state variable controlling fluxes of water and energy from the surface into the lower atmosphere. Over land, both surface temperature and surface soil moisture (SSM) are critical. Variability in both quantities is imprinted on near-surface weather and climate over land. The largest source of temporal variability in surface temperature is the seasonal cycle (also referred to as the “annual cycle”). Many previous studies have characterized the seasonal cycle in surface or near-surface temperatures using simple energy balance models to explain spatial variability (e.g., Stine et al. 2009; Dwyer et al. 2012; Stine and Huybers 2012).

Perhaps surprisingly, there has yet to be a comprehensive analysis of spatial variability in the SSM seasonal cycle at global scales. There are at least three plausible reasons for this. First, observations of soil moisture have lacked the density and coverage of near-surface temperature measurements, both spatially and temporally. Second, in models, soil moisture is primarily designed to be an index that reproduces accurate surface fluxes of heat and moisture (Koster et al. 2009), resulting in substantial additive and multiplicative biases in modeled soil moisture (e.g., Reichle et al. 2004). Such biases are much less pronounced in modeled near-surface temperatures. Third, for engineering purposes, interannual deviations from the seasonal mean are arguably more important than the mean state itself. For example, infrastructure can be designed to comfortably withstand predictably dry summers and wet winters, but cannot be redesigned to withstand an unpredictably wet or dry year. Many studies focus on subseasonal variability by removing the seasonal cycle, by 1) considering seasonal anomalies (e.g.,

Orth and Seneviratne 2012) or daily to weekly temporal fluctuations (e.g., McColl et al. 2017a,b, 2019a; Feldman et al. 2018), 2) focusing on one season, usually the northern summer (e.g., Koster and Suarez 2001; Vargas Zeppetello et al. 2020a,b), or 3) comparing model outputs to those from control integrations with soil moisture prescribed to follow a fixed, climatological seasonal cycle (e.g., Koster et al. 2006; Berg et al. 2015). Subseasonal variability is clearly important, but from the perspective of basic climate science the seasonal cycle of SSM is arguably just as important as the seasonal cycle of surface temperature.

Beyond its fundamental importance to the climate system, the seasonal cycle of SSM directly constrains many other important processes. In particular, seasonal SSM dynamics are observed to be a critical controlling factor on seasonal flooding across much of Africa (Tramblay et al. 2021), Australia (Wasko et al. 2020), and Europe (Berghuijs et al. 2019); seasonal nitric oxide emissions from soils in California (Homyak and Sickman 2014); and vegetation patterns in the western United States (Rickard 1967; Daubenmire 1968). Furthermore, seasonal cycles in soil moisture have similarly been observed to influence the mobilization of toxic metals from soils and sediments (Van den Berg et al. 1998; Connolly et al. 2022)—notably driving seasonal patterns in arsenic release in rice paddy soils (Roberts et al. 2011) and uptake into rice plants (Li et al. 2009) as well as the overall arsenic mass balance for these soils (Roberts et al. 2010).

In this study, we use global observations of SSM to describe and explain its seasonal cycle. Our focus is on satellite observations rather than model simulations, given the uniquely severe limitations of simulated soil moisture (e.g., Koster et al. 2009). Many studies document the seasonal cycle of soil water storage at specific sites or regions (e.g., Hollinger and Isard 1994; Illston et al. 2004; Nandintsetseg and Shinoda 2011;

Corresponding author: Mason Stahl, stahlm@union.edu

DOI: 10.1175/JCLI-D-21-0780.1

© 2022 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy \(www.ametsoc.org/PUBSReuseLicenses\)](#).

Fernández-Long et al. 2021), or propose models of the seasonal cycle of soil water storage (e.g., Milly 1994a,b; Laio et al. 2002; Katul et al. 2007; Feng et al. 2012, 2015), which are typically evaluated at a handful of specific sites or regions. While these studies have yielded valuable insights, it is difficult to generalize results from any one particular region due to soil moisture's complex dependence on site-specific land surface properties. Our global analysis permits a broad overview and the possibility of identifying a small number of generalizable controls. Furthermore, it has been common to characterize the seasonal cycle in terms of harmonics in previous studies of other variables such as temperature (e.g., Stine and Huybers 2012; Stine et al. 2009) or precipitation (e.g., Kirkyala and Hameed 1989). However, the seasonal cycle of soil moisture is often poorly represented by one or two harmonics, and adding higher-order harmonics clouds interpretation. Therefore, in this study, we use alternative techniques that are parsimonious, like previous analyses based on harmonics, but provide greater flexibility in the functional form fit to the soil moisture observations. To interpret the eight parameters fit per pixel, we use an unsupervised clustering algorithm to partition the results into five regimes. While the most common regime is one with no clear seasonal cycle in SSM, many regions do exhibit a pronounced seasonal cycle. To better understand the physical basis for each regime, we train a random forest classifier on global observations of meteorological drivers of surface soil moisture, and find that most of the observed variability can be explained by information on precipitation and downwelling surface shortwave radiation alone. Based on this insight, we derive a parsimonious conceptual model of soil saturation based on established physical principles. The model only uses information on precipitation and downwelling surface shortwave radiation as inputs, has zero free parameters, and captures the first-order properties of the SSM seasonal cycle in each regime remarkably well.

This manuscript is structured as follows. In section 2, we characterize seasonal SSM dynamics based on global observations and use unsupervised machine learning to identify areas with similar SSM dynamics. We then empirically identify the most important climate drivers that are responsible for the emergence of the different SSM seasonal cycle regimes. We compare how these drivers differ between SSM regimes and across the same regimes that emerge in different regions of the globe. To explain the results, the conceptual model is introduced in section 3. We discuss the environmental significance of our findings in section 4 and the study's limitations in section 5. Section 6 summarizes the study's main conclusions, and speculates on potentially interesting lines of future research.

2. Data analysis

a. Soil moisture and climatic data acquisition

We use daily 36-km spatial resolution NASA Soil Moisture Active Passive (SMAP) soil moisture data (SPL3SMP version 7) from 31 March 2015 to 28 April 2021 in our analysis (Entekhabi et al. 2010). Pixels with retrieval quality flags of 0, 1, 8, or 9 were retained for analysis. This includes pixels with dense vegetation, which are likely subject to larger

retrieval errors, although perhaps to a lesser extent than previously thought (Colliander et al. 2020). We chose to retain these pixels in the main analysis to maximize spatial coverage, but repeated the entire analysis using only pixels with the highest quality retrievals (i.e., retrieval_qual_flag = 0 or 8). Our results were qualitatively the same, though the spatial coverage was more limited (appendix B). Since SMAP has been extensively validated in previous studies (Chan et al. 2016; Colliander et al. 2017; Chen et al. 2018), we do not further validate the soil moisture retrievals here.

SMAP nominally measures soil moisture in the top 5-cm soil layer, rather than deeper layers [~ 1 m, often referred to as "root zone" soil moisture (RZSM)] that may be more relevant to plants. Hence, our study focuses on SSM rather than RZSM. However, SSM is typically well correlated with RZSM in many regions (Ford et al. 2014; Akbar et al. 2018; Short Gianotti et al. 2019), meaning it can be regarded as a proxy of RZSM under many conditions. We also distinguish between soil moisture [θ , the ratio of the volume of water to the unit volume of soil ($\text{m}^3 \text{m}^{-3}$)], and soil saturation [$s = (\theta - \theta_w) / (\theta_{fc} - \theta_w)$ (%), where θ_w ($\text{m}^3 \text{m}^{-3}$) is the wilting point soil moisture, an approximate lower bound, and θ_{fc} ($\text{m}^3 \text{m}^{-3}$) is the field capacity, an approximate upper bound for soil water storage]. To convert between the two quantities, we use gridded estimates of θ_{fc} and θ_w obtained from the HiHydroSoil (v2.0) database (Simons et al. 2020).

Globally gridded monthly normals (1981–2010) for air temperature, precipitation, and downwelling surface shortwave radiation, and potential evapotranspiration (PET) were obtained from the TerraClimate dataset, which has a spatial resolution of 4 km (Abatzoglou et al. 2018). TerraClimate PET is only used in a supplementary analysis provided in appendix A and appendix B. Seasonality metrics for the overall level of seasonality and period of seasonal concentration for both precipitation and downwelling surface shortwave radiation were computed using the method of Markham (1970). In this approach, mean monthly values are treated as vector quantities, where the magnitude is the monthly mean value (i.e., monthly mean precipitation or downwelling surface shortwave radiation) and the direction is the month of the year expressed as an angle (e.g., January = 15° , February = 45° , ...). Adding the monthly vectors for a climate variable (e.g., precipitation) provides a measure of the variable's seasonality, with the resultant vector's magnitude representing the level of seasonality and its direction representing the period of seasonal concentration. Two indices are used to summarize this information:

$$\text{SI} = \frac{(V_x^2 + V_y^2)^{1/2}}{V_{\text{tot}}} \quad \text{and}$$

$$\text{SC} = \tan^{-1}\left(\frac{V_y}{V_x}\right) \frac{180}{\pi} + \begin{cases} 180, & \text{when } V_x \leq 0 \\ 0, & \text{when } V_x > 0 \text{ and } V_y > 0, \\ 360, & \text{when } V_x > 0 \text{ and } V_y \leq 0 \end{cases}$$

where V_i is the monthly value of precipitation or downwelling surface shortwave radiation for month i , θ_i is the angle for

month i , $V_x = \sum_{i=1}^{12} V_i \cos(\theta_i)$ is the x direction component of the sum of the monthly vector, $V_y = \sum_{i=1}^{12} V_i \sin(\theta_i)$ is the y direction component of the sum of the monthly vectors, and $V_{\text{tot}} = \sum_{i=1}^{12} V_i$. SI is the seasonality index, which indicates the level of seasonality: SI = 0 indicates that values are uniformly distributed throughout the year, whereas SI = 1 indicates that values are concentrated within a single month. SC is the seasonal concentration, which is the resulting angular direction (in units of degrees) of the annual vector, and this direction indicates the month in which monthly values are most concentrated.

For the random forest modeling (section 2c) and the conceptual soil moisture modeling (section 3) some data aggregation was required to ensure consistent spatial and temporal resolution across the underlying dataset. Both required the gridded TerraClimate data at the same spatial resolution as the SMAP data. Thus the 4-km-spatial-resolution TerraClimate data were aggregated by taking the mean of pixels within each 36-km grid cell of the SMAP grid. Raster data processing was done in the R programming language (R Core Team 2020) using the Raster package (Hijmans 2020). The conceptual soil moisture model generates monthly soil moisture values from the input of monthly climate normals. Thus when comparing the output from the conceptual model to the SMAP observations, which are subweekly (see section 3), we computed mean monthly soil moisture values from the SMAP data.

b. Seasonal soil moisture regime classification

The daily SMAP data were used to compute weekly composite soil moisture values for each pixel. For a given week and pixel, this involved taking all of the daily values for that week (e.g., week 1 = 1–7 January, week 2 = 8–15 January, ...), across all years of record (2015–21) and computing the weekly mean of the soil moisture measurements. Thus pixels with a complete soil moisture record have 52 weekly composite values. For pixels where gaps in the weekly composite record existed due to frozen soils, we gap filled the weekly composite data to generate a complete record if the following conditions were met: 1) the pixel had at least 10 complete weeks of data and 2) only a single contiguous gap existed in the weekly composite record. Gap filling was done by linear interpolation. After gap filling, pixels with a complete weekly composite record (i.e., 52 weeks) were classified according to their seasonal surface soil moisture dynamics. The results of our analyses are qualitatively insensitive to gap filling (appendix B), except that the spatial coverage in regions with colder winters is much greater when gap filling is performed.

The primary focus of this study is on the seasonal dynamics of soil moisture, rather than its mean state. Therefore, to allow for better comparison between pixels that have different mean soil moisture values, we first subtracted each pixel's mean annual soil moisture value from its weekly composite values. In addition, the weekly composite record for pixels in the Southern Hemisphere were phase-shifted by 26 weeks to align seasons between the Northern and Southern Hemispheres and thus allow for a more intuitive comparison across hemispheres.

TABLE 1. Random forest soil moisture regime model prediction errors. Variables were added to the model in order of their predictive power (Fig. 3). Thus, the single-variable model uses only the most predictive variable. The two-variable model uses the top two predictors, the three-variable model the top three, and so on. Descriptions of the climate variables and the calculation of seasonal concentration (SC) and seasonality index (SI) are described in section 2a. The model with all seven predictors is referred to as the “full model” in the text, figures, and tables.

Variables used in model	Model prediction error (%)
Annual precipitation	15.55%
Downwelling surface shortwave radiation SC	
Precipitation SI	
Downwelling surface shortwave radiation SI	
Downwelling surface shortwave radiation mean	
Mean annual air temperature	
Precipitation SC	
Annual precipitation	16.26%
Downwelling surface shortwave radiation SC	
Precipitation SI	
Downwelling surface shortwave radiation SI	
Downwelling surface shortwave radiation mean	
Mean annual air temperature	
Annual precipitation	17.80%
Downwelling surface shortwave radiation SC	
Precipitation SI	
Downwelling surface shortwave radiation SI	
Downwelling surface shortwave radiation mean	
Annual precipitation	20.94%
Downwelling surface shortwave radiation SC	
Precipitation SI	
Downwelling surface shortwave radiation SI	
Annual precipitation	28.19%
Downwelling surface shortwave radiation SC	
Precipitation SI	
Annual precipitation	41.28%
Downwelling surface shortwave radiation SC	
Annual precipitation	61.94%

Once the data were rescaled we projected the discrete (i.e., weekly) observations to continuous functions using functional data analysis (FDA) (Wang et al. 2016), following the approach of Brunner et al. (2020), who successfully used FDA to characterize the seasonal dynamics of streamflow across the United States. We did this by fitting eight B-spline basis functions of order 4 to each pixel's rescaled weekly composite data. This resulted in eight coefficients per representative soil moisture pixel.

We then used cluster analysis to summarize this information globally. The pixels were clustered using the k -means algorithm (Hartigan and Wong 1979) on the computed spline coefficients. This unsupervised learning approach allowed for the identification of soil moisture regimes based

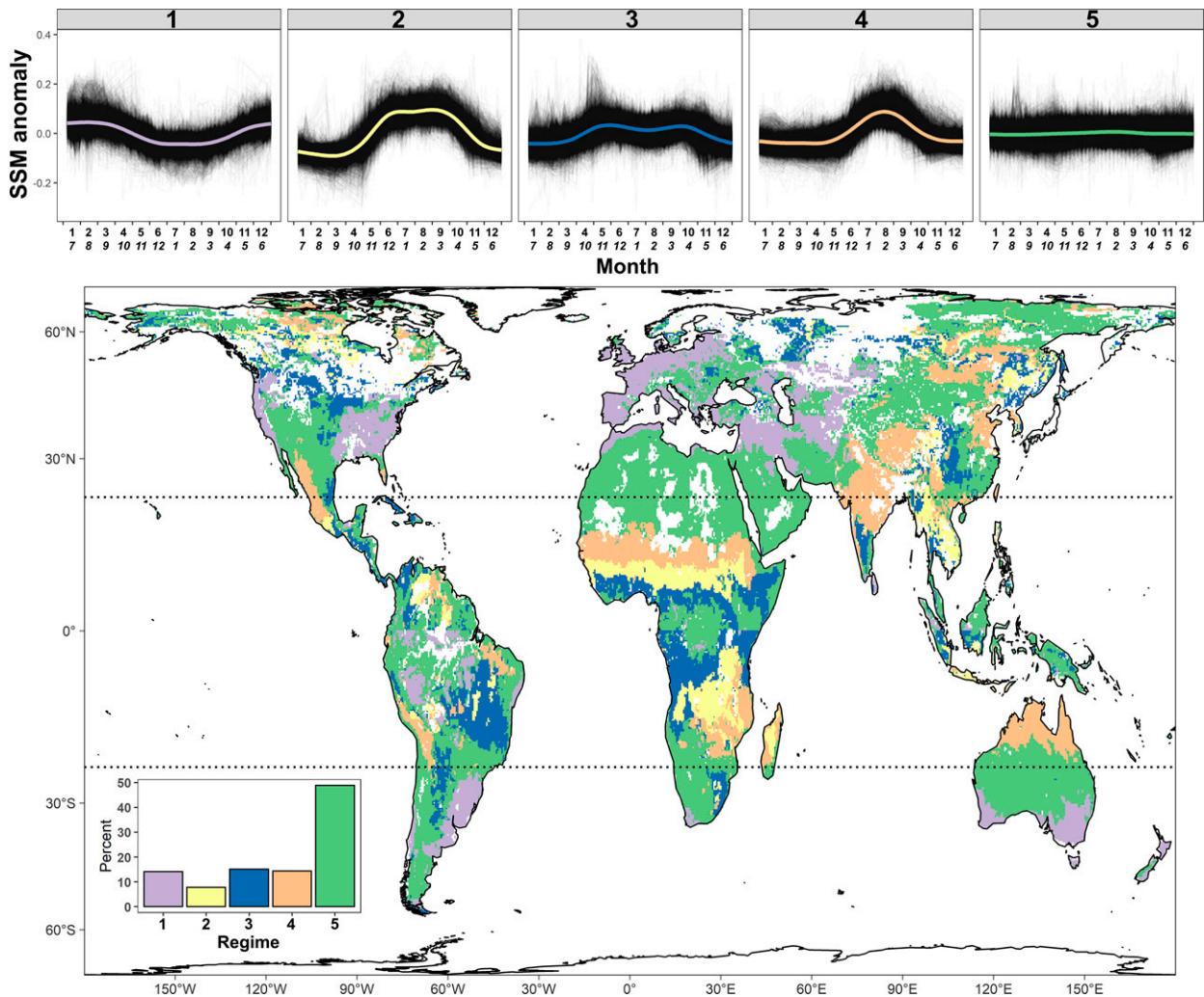


FIG. 1. (top) SSM seasonal cycle by regime, with the seasonal cycle for each individual pixel shown as gray lines and the representative LOESS smoothed seasonal cycle shown as colored lines. The upper row of month values on the horizontal axis applies to the Northern Hemisphere and the lower row applies to the Southern Hemisphere (data in the Southern Hemisphere were phase-shifted by 6 months; see section 2b). (bottom) Spatial distribution of SSM seasonal cycle regimes. White pixels represent missing data. Horizontal dashed lines represent the bounds of the tropics. Inset: percentage of total land area (excluding land with missing values) per regime.

on the underlying data. The number of clusters to use in the analysis is a somewhat subjective choice. To illustrate the tradeoffs in making this choice, consider two extreme cases for a cluster analysis of seasonal cycles for N pixels. At one extreme, choosing to use only one cluster would result in each pixel being assigned to the same cluster. This certainly reduces the dimension of the data, but the resulting cluster would not be particularly meaningful: many of the seasonal cycles assigned to the cluster would deviate substantially from the cluster's mean seasonal cycle. At the other extreme, choosing to use N clusters would result in each pixel being assigned to its own distinct cluster. In this case, each seasonal cycle assigned to a given cluster would perfectly match the cluster's mean seasonal cycle. However, the cluster analysis would be useless, since it does not reduce the dimension of the data. In general, the major trade-off in

choosing the number of clusters to use is between accuracy (the degree to which each seasonal cycle assigned to a cluster matches the cluster's mean seasonal cycle) and simplicity (the fewer clusters, the better). The marginal benefit of adding one more cluster often diminishes beyond a certain threshold. This intuition is formalized by the “elbow method” (Boehmke and Greenwell 2019) for choosing the number of clusters. We adopted that approach here. This resulted in five clusters (which we refer to as “regimes” for the rest of the study). Every pixel that met the filtering requirements detailed above was assigned to exactly one regime.

c. Identification of climatic determinants of soil moisture regimes

To identify the climatic drivers that are likely controlling the emergence of the soil moisture regimes identified through

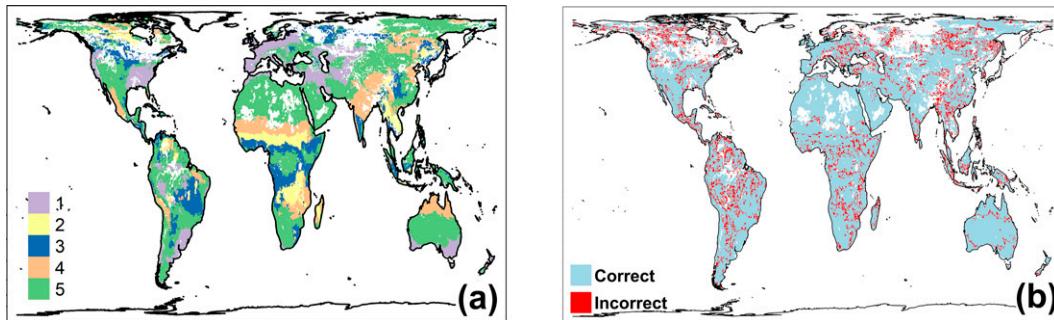


FIG. 2. (a) Random forest model (full model; see Table 1)-predicted regimes. (b) Model-predicted error map. Blue pixels are correct predictions and red pixels are incorrect predictions. White pixels represent missing data.

k-means clustering, we developed a predictive model to determine the soil moisture regime based solely on climate data. This involved training a random forest model on the soil moisture regimes determined by *k*-means clustering and using the air temperature, precipitation, and downwelling surface shortwave radiation data as predictor variables (Table 1). Briefly, a random forest model is a machine learning algorithm that can be used for classification and regression (James et al. 2021) and these models have become widely used in the geosciences (Bergen et al. 2019). For classification tasks random forests operate by constructing an ensemble of decision trees based on the model training data—observations with a known class (target) and known attributes (predictor variables)—and the trained model can then be used to predict the class of further observations based on their attributes (predictor variables) (James et al. 2021). Furthermore, random forest models allow for the straightforward interpretation of which predictor variables are most important in the classification task the model was trained for (Boehmke and Greenwell 2019; Bergen et al. 2019). Since the primary goal of the random forest model was to gain a clearer understanding of the controls on the seasonal cycle of SSM, we used as predictor variables seasonality metrics (SI and SC, described in section 2a) and annual totals as opposed to the full monthly time series for precipitation and downwelling surface shortwave radiation. This approach reduces the dimensionality of a monthly time series from 12 to 3 dimensions (SI, SC, and annual total) and thus improves the interpretability of variable importance. A 50–50 training–test split was used in the development of the model (i.e., 50% of the pixels were randomly selected and used to train the model and the remaining 50% were used to test the model). Model performance was evaluated using the results from the out-of-bag (OOB) sample. The relative importance of predictors was interpreted using the permutation importance. The permutation importance is determined by randomly permuting the value of a single predictor variable, while keeping all other predictor variables unchanged, and comparing the predictions for the permuted case to the predictions for the unpermuted case (Strobl et al. 2007). The importance of a given predictor is assessed by how much the model accuracy decreases as a result of permuting that variable. Thus, the greater the importance of a given predictor, the greater the decrease in model accuracy as a result of permuting that variable.

d. Seasonal soil moisture dynamics and their controls

Observed surface soil moisture is clustered into five distinct seasonal cycle regimes (Fig. 1):

- Regime 1: Muted minimum from May to October in the Northern Hemisphere and from November to April in the Southern Hemisphere. This regime is found in the Middle East, central Asia, western and northern Europe, much of the eastern United States, and along the U.S. West Coast in the Northern Hemisphere. In the Southern Hemisphere it is found in some areas of central and southeastern South America and southern Australia.
- Regime 2: Pronounced long-duration peak from May to September in the Northern Hemisphere and from November to March in the Southern Hemisphere. In the Northern Hemisphere it is observed in an east–west belt between the Sahel and the Congo rainforest as well as across much of Southeast Asia. In the Southern Hemisphere this regime is observed in southern central Africa, northern Madagascar, and small pockets in Brazil.

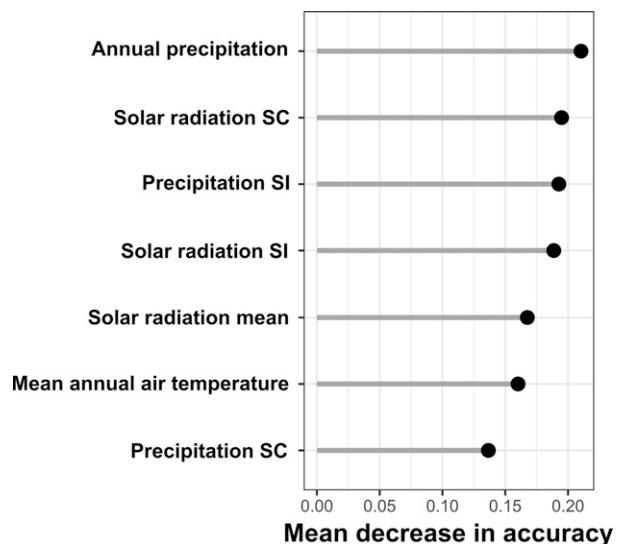


FIG. 3. Random forest model variable importance. The variable importance is computed by the permutation accuracy importance method (see section 2c for details).

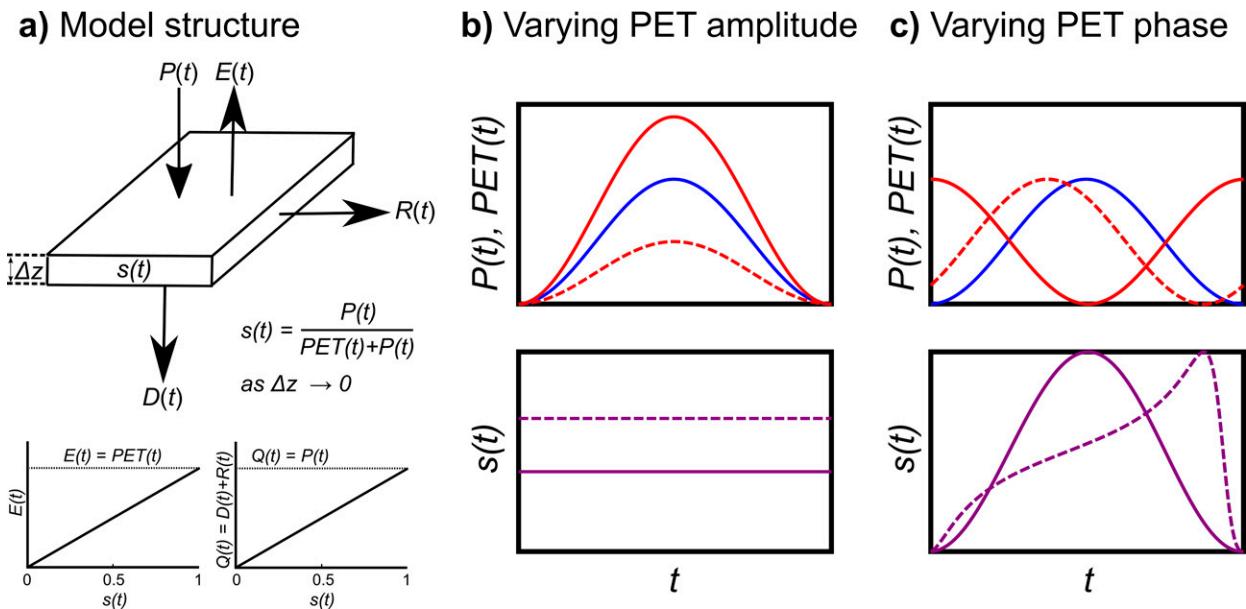


FIG. 4. (a) Conceptual model of soil saturation $s(t)$, given in Eq. (2): $P(t)$ is precipitation at time t , $E(t)$ is evapotranspiration, $R(t)$ is runoff, $D(t)$ is drainage, and Δz is the thickness of the soil layer. (b) Example model output for a case in which $P(t)$ and $PET(t)$ are in phase, and the $PET(t)$ amplitude is varied. The modeled $s(t)$ (solid and dashed purple lines) does not exhibit a seasonal cycle, even when the amplitude of PET is greater than (red solid line) or less than (red dashed line) the amplitude of P (blue solid line). (c) Example model output for a case in which $P(t)$ and $PET(t)$ have the same amplitude, and the $PET(t)$ phase is varied.

- Regime 3: Muted long-duration peak from May to October in the Northern Hemisphere and from November to April in the Southern Hemisphere. It is commonly found in the tropics of Africa, and large areas of eastern Brazil and central Argentina. It is also found in parts of Central America and the central United States, as well as southern India and central China.
- Regime 4: Pronounced short-duration peak from July to September in the Northern Hemisphere and from January to March in the Southern Hemisphere. This regime is found in an east–west band across the northern Sahel, much of India, the central Tibetan Plateau, northern Mongolia, northeast China, and in western Mexico. In the Southern Hemisphere it is found in scattered pockets throughout South America, southeast Africa, and northern Australia.
- Regime 5: No seasonal cycle, found widely throughout the Southern and Northern Hemispheres. This regime is mainly found in either arid regions (including the Sahara, Australian desert, southern Africa, and the south-central United States) or tropical rain forests (including the Amazon and Congo).

The emergence of a given soil moisture regime is largely explained by the seasonal cycles of precipitation and downwelling surface shortwave radiation. A random forest model using precipitation [annual total, seasonal timing (SC), and level of seasonality (SI)] and downwelling surface shortwave radiation [annual total, seasonal timing (SC), and level of seasonality (SI)] and air temperature (mean annual) data as predictors accurately predicts the soil moisture regimes that were determined by k -means clustering (Fig. 2; Table 1; see also

Tables A1 and A2). The most important explanatory variables in the random forest model were 1) total annual precipitation, 2) timing of peak seasonal downwelling surface shortwave radiation, 3) level of seasonality of precipitation, and 4) level of seasonality of downwelling surface shortwave radiation (Fig. 3). The explanatory power of the top four most important predictors (annual precipitation, downwelling surface shortwave radiation SC, precipitation SI, and downwelling surface shortwave radiation SI) is highlighted by 1) the relatively strong performance (prediction error = 20.9%) of the random forest model using only these four variables and 2) the dramatic improvement in model performance between the single-variable random forest model and the four-variable random forest model (Table 1). Further, although more modest, gains in model performance occur as additional predictor variables are added to the random forest model, with the full model having a prediction error of 15.5% (Table 1). Notably, regions without a seasonal cycle in surface soil moisture (regime 5) were largely identifiable due to low precipitation alone, while regions with a seasonal cycle in soil moisture are distinguished based on the magnitudes and seasonal cycles of precipitation and downwelling surface shortwave radiation.

3. Conceptual model

Why is the SSM seasonal cycle mainly determined by precipitation and downwelling surface shortwave radiation? To answer that question, we introduce a simple, physically based model of SSM, which only uses precipitation and downwelling surface shortwave radiation as inputs. Several models of the

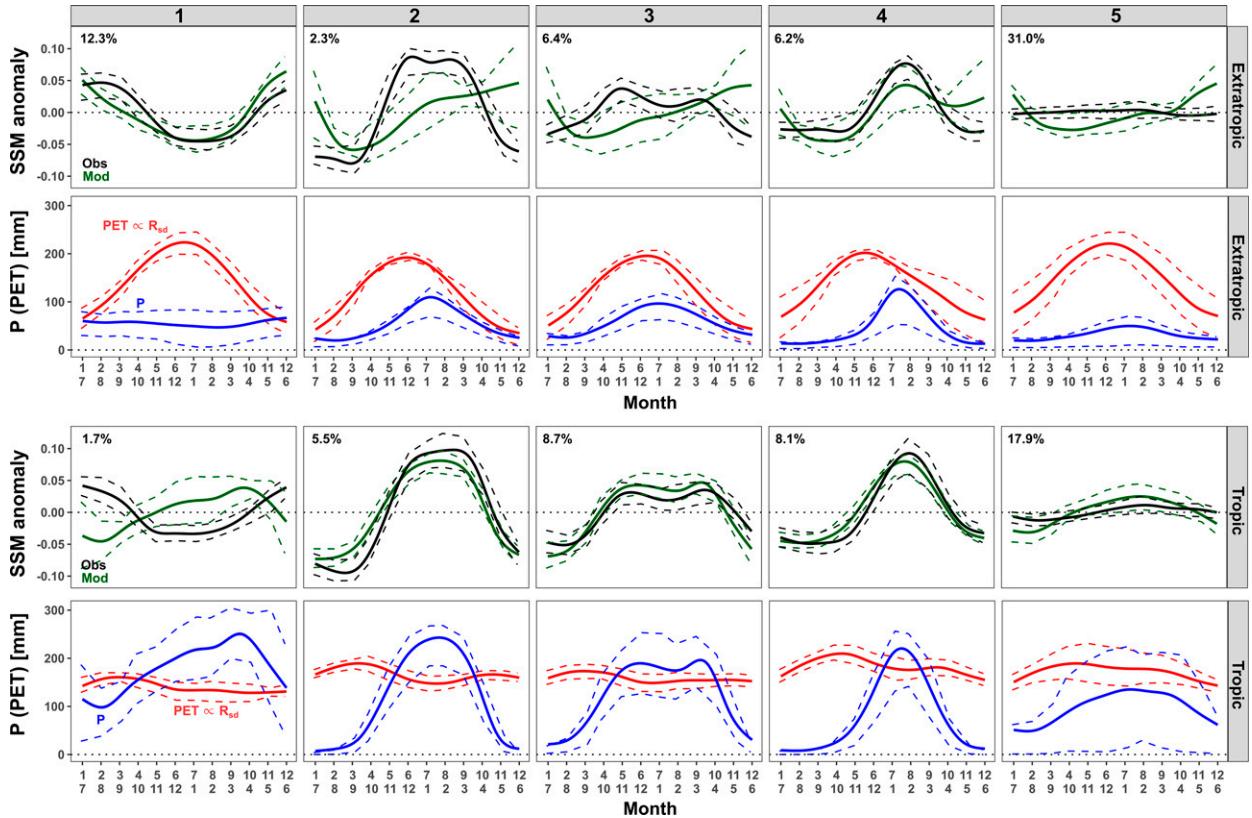


FIG. 5. Precipitation and PET [where $PET = 0.8R_{sd}(t)/(\rho\lambda)$] seasonal cycle and soil moisture seasonal cycle by region (rows) for each soil moisture seasonality regime (columns). Precipitation (blue lines) and PET (red lines) are shown for (second row) the extratropics and (fourth row) the tropics. Precipitation and downwelling surface shortwave radiation (R_{sd}) data are from TerraClimate gridded monthly normals (1981–2010) (Abatzoglou et al. 2018). Observed (SMAP) soil moisture anomalies (black lines) and modeled soil moisture anomalies (green lines) by regime (columns) are shown for (first row) the extratropics and (third row) the tropics. Modeled soil saturation from Eq. (2) was converted to soil moisture using gridded estimates of soil field capacity and wilting point obtained from the HiHydroSoil (v2.0) database (Simons et al. 2020). In all panels, solid lines represent the LOESS-smoothed monthly values for all pixels in a given regime and hemisphere and dashed lines represent the 25th- and 75th-percentile values. The upper row of month values on the horizontal axis applies to the Northern Hemisphere and the lower row applies to the Southern Hemisphere (data in the Southern Hemisphere were phase-shifted by six months; see section 2b). The percentages in the top-left corner of the panels represent the proportion of classified land area that falls into a given soil moisture regime and region.

seasonal dynamics of SSM have been proposed in previous studies (e.g., Milly 1994a; Laio et al. 2002; Feng et al. 2012, 2015). As the integral of the sum of surface water fluxes constrained by soil and vegetation, soil moisture exhibits substantial spatial variability; as a result, these models typically require multiple parameters. For example, the model results shown by Laio et al. (2002) in their Fig. 3 require the specification of seven parameters related to soil hydraulic properties, in addition to information on forcings (potential evapotranspiration and precipitation). While earlier studies have provided considerable insight, we will show that a substantially simpler model is sufficient to represent the main features of the SSM seasonal cycle regimes identified in previous sections.

The water balance of a vertically averaged, horizontally homogeneous control volume of soil extending from the land surface down to a depth Δz (m) (Fig. 4a) can be modeled as

$$\begin{aligned} \Delta z(\theta_{fc} - \theta_w) \frac{ds}{dt} &= P(t) - E(s, t) - Q(s, t) \\ &\approx P(t) - PET(t)s(t) - P(t)s(t), \end{aligned} \quad (1)$$

where t is time (s), P is precipitation at the land surface (m s^{-1}), E is evapotranspiration from the land surface (m s^{-1}), and Q is the sum of drainage (vertical transport to deeper soil layers) and runoff (horizontal transport) (m s^{-1}). We treat Q as equal to the product of precipitation and soil saturation, so that higher-intensity precipitation will increase Q , all else being equal, resulting in greater partitioning of precipitation to surface runoff and infiltration. We treat E as equal to the product of soil saturation and potential evapotranspiration (PET; m s^{-1}). PET is used as a forcing in this study rather than E because PET is not causally influenced by soil moisture, whereas E is itself strongly determined by soil moisture in water-limited environments. PET is, therefore, the more fundamental external forcing.

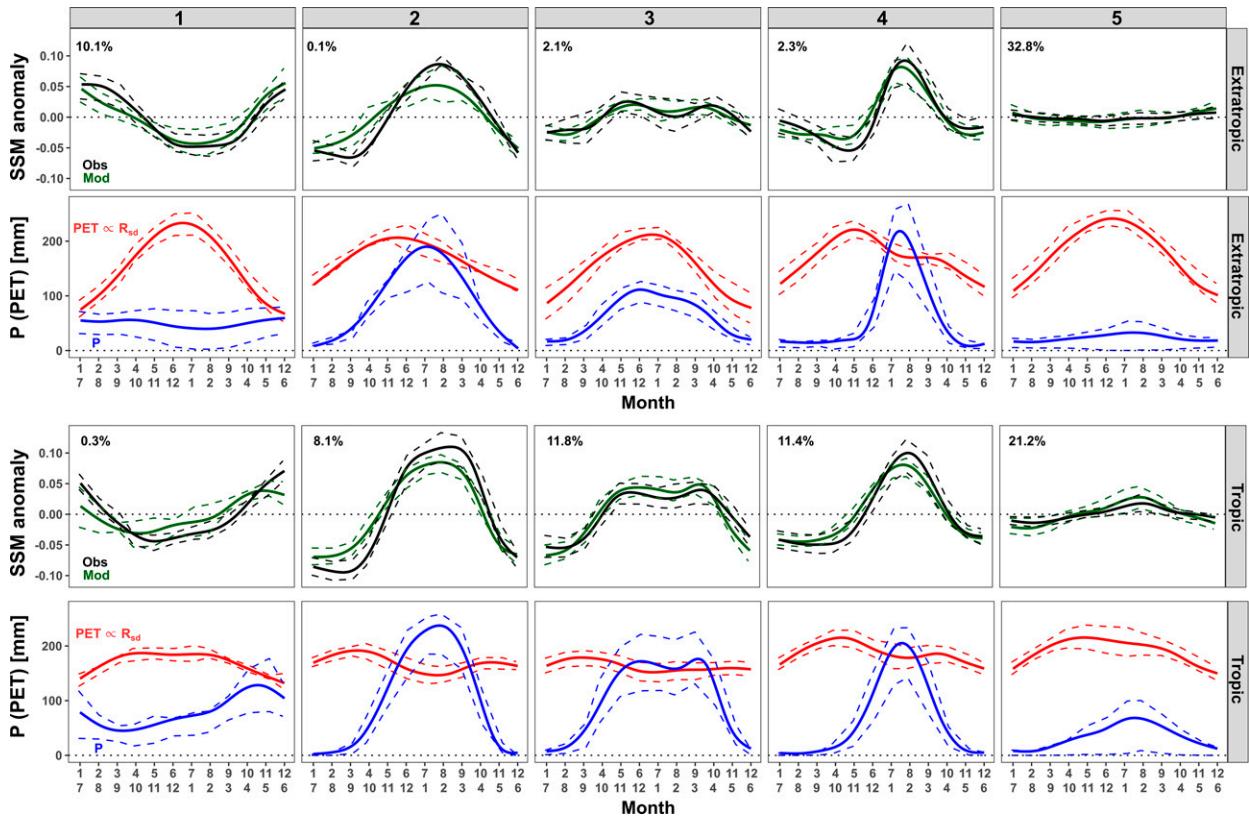


FIG. 6. As in Fig. 5, but restricting the analysis to regions with the highest-quality SMAP retrievals (see appendix B for further details).

Nonlinear variants of this simple model have been used in previous studies (e.g., Koster and Mahanama 2012). Our linear model does not necessarily sacrifice much model accuracy: for example, see Figs. 6d, 7d, and 8d of Koster and Mahanama (2012), which show that a linear variant of their nonlinear model performs reasonably well in reproducing results from a much more complex model. Our model prioritizes 1) maximum simplicity and 2) aspects relevant to the seasonal cycle (approximately monthly time scales).

Recent work (Koster and Mahanama 2012; Milly and Dunne 2016; Maes et al. 2019) has shown that PET can be estimated very well using the simple relation $PET \propto (R_n/\rho\lambda)$, where R_n is net surface radiation (W m^{-2}), ρ is the density of water (kg m^{-3}), and λ is the latent heat of vaporization (J kg^{-1}). This relation, which is substantially simpler than alternatives, has also been shown to be more accurate when validated against ground observations (Maes et al. 2019). Since the seasonal cycle of net surface radiation is largely determined by the seasonal cycle of downwelling surface shortwave radiation (R_{sd} ; W m^{-2}), we further simplify this relation to $PET \propto (R_{sd}/\rho\lambda)$. We use the same constant of proportionality (0.8) as that used in Milly and Dunne (2016). The model is not calibrated.

Taking the limit as $\Delta z \rightarrow 0$ (a relevant approximation for surface soil moisture), and rearranging gives

$$s(t) = \frac{P(t)}{PET(t) + P(t)} = \frac{\rho\lambda P(t)}{0.8R_{sd}(t) + \rho\lambda P(t)}. \quad (2)$$

To our knowledge, this model of surface soil saturation is new. Since Koster and Mahanama (2012) do not focus on the seasonal cycle, the most similar previous work is, arguably, Laio et al. (2002), which is itself similar to, for example, Milly (1994a,b) and Feng et al. (2015). More specifically, our model is nearly equivalent to a special case of Eqs. (14) and (15) of Laio et al. (2002): in their notation, the case in which the active soil layer depth $Z_r = 0$ (since the focus of our study is on surface soil moisture), the wilting point soil saturation $s_w = 0$, and the critical soil saturation $s^* = 1$. In this case, the loss function (the sum of E and Q , in our notation) is a linear function of soil saturation, and so the higher-order corrections discussed at length in Laio et al. (2002) are not required. While the assumption of a linear loss function is strong, it is better justified at larger spatial scales relevant to satellite soil moisture retrievals, rather than the point scales of interest to Laio et al. (2002). As we will show, it is also empirically successful, at least in reproducing the first-order behavior of the SSM seasonal cycle. The main difference between our model and that of Laio et al. (2002) is the choice of estimate for PET. They use a variant of the Thornthwaite equation (Thornthwaite 1948), which uses air temperature as an input, whereas our study uses a variant of the model proposed by

TABLE A1. Observed and random forest (full model) predicted regimes for all pixels. Values in the table indicate the number of pixels in each category.

Predicted regime	Observed regime				
	1	2	3	4	5
1	4844	12	54	15	800
2	5	2484	360	253	136
3	60	287	4516	126	1008
4	7	244	109	4664	803
5	669	90	716	535	17644

Milly and Dunne (2016) with R_{sd} used as the only input. While temperature-based estimates of PET have been widely used, they are empirically less accurate than radiation-based estimates (Maes et al. 2019), and also less physically justified. Our own random forest model analysis, presented in the previous section, demonstrates that R_{sd} explains substantially more spatial variability in the SSM seasonal cycle than air temperature, further justifying our approach.

Our model is a transformation of the well-known aridity index (often written as $AI = PET/P$), a common proxy of land surface “dryness.” The AI reproduces spatial and intra-annual temporal variability in surface soil moisture reasonably well (although not interannual and multidecadal variability; see, e.g., Greve et al. 2019; Berg and McColl 2021; Scheff et al. 2021). In contrast to the AI, our expression is derived from basic physical principles, and scales between 0 and 1. In addition, the Budyko curve relates E/P to AI by a single curve, at least for annual mean values (Budyko 1958, 1974). Our model predicts $E/P = (1 + AI^{-1})^{-1}$, which is identical to the functional form of the Budyko curve proposed in several earlier studies for the special case in which $n = 1$ (Choudhury 1999; Yang et al. 2008).

The simple model shows that the seasonal cycle of SSM is controlled, to first order, by the seasonal cycles of both precipitation and PET (which is reasonably approximated as proportional to R_{sd}). For the case in which precipitation dominates— $P(t) \gg PET(t) = [0.8R_{sd}(t)]/\rho\lambda$ —our simple model predicts $s(t) \approx 1$ (i.e., soils are saturated and soil moisture is at field capacity). For the case in which R_{sd} dominates— $P(t) \ll PET(t) = [0.8R_{sd}(t)]/\rho\lambda$ —our model predicts $s(t) \approx 0$ (i.e., soils are dry and soil moisture is at wilting point). Even for this highly simplified model, the predicted seasonal cycle of SSM is nontrivial. For example, if precipitation and PET both have seasonal cycles with large (constant) amplitudes, but their seasonal cycles are in phase

with one another, then this simple model would predict no seasonal cycle in SSM (Fig. 4b). If the seasonal cycles of precipitation and PET are not in phase with one another, then the seasonal cycle of SSM can be qualitatively different from that of either P or PET (Fig. 4c). While the model presented here is intentionally highly idealized, adding greater complexity would only make the seasonal cycle of SSM more complex.

Despite its simplicity, the conceptual model is broadly successful in reproducing the SSM seasonal cycle regimes (Fig. 5). The conceptual model’s performance is poorest for regime 1 in the tropics (comprising 1.7% of the classified area) and for regimes 2 and 3 in the extratropics (comprising 2.3% and 6.4% of the classified area, respectively). However, when considering only pixels with the highest-quality SSM retrievals (i.e., excluding regions with dense forests or with periodically frozen soil, as is common; see appendix B), the conceptual model performance improves considerably in these regions (Fig. 6). Using a substantially more complex model of PET within the conceptual model only decreases its accuracy (Fig. A1). Plausible alternatives of comparable complexity to our model turn out to be insufficient. For example, the alternative model $s(t) \propto P(t) - PET(t)$ [a common proxy of surface “dryness” used in, e.g., Slessarev et al. (2016)] fails to adequately reproduce regime 5 in the extratropics, which covers the largest area of any regime in the classification (Fig. A2). The alternative model $s(t) \propto P(t)$ similarly fails to identify regime 1 in the extratropics, which covers the second largest area in the extratropics of any regime in the classification (Fig. A3). In addition, both models lack a clear derivation based on the land surface water budget, neither model constrains $s(t)$ to lie between zero and one, and they both require a constant of proportionality parameter to convert to the right units.

We now use the conceptual model to interpret and explain the observed SSM seasonal cycles. Each of the observed SSM seasonal cycle regimes are present in both the Northern and Southern Hemispheres. For each regime, the hemisphere-averaged seasonal cycles of P and PET for each regime do not vary considerably between the Northern and Southern Hemispheres (not shown), indicating that the basic mechanisms linking P , PET, and SSM are similar in both hemispheres after phase-shifting the Southern Hemisphere by 6 months.

In the tropics, there is little seasonal variability in R_{sd} (and, therefore, PET), so the seasonal cycle of P dominates the seasonal cycle of SSM. Regimes 2, 3, and 4 all have a larger presence in the tropics than the extratropics, and all include a clear seasonal cycle in P , with P sometimes exceeding PET (Fig. 5). This results in a pronounced SSM seasonal cycle for these

TABLE A2. Random forest (full model) model performance statistics by regime. The statistics reported in the table are defined below.

	Regime 1	Regime 2	Regime 3	Regime 4	Regime 5
Sensitivity	0.85	0.77	0.75	0.80	0.90
Specificity	0.98	0.98	0.96	0.97	0.87
Positive prediction value (PPV)	0.87	0.80	0.78	0.83	0.87
Negative prediction value (NPV)	0.97	0.98	0.96	0.97	0.90

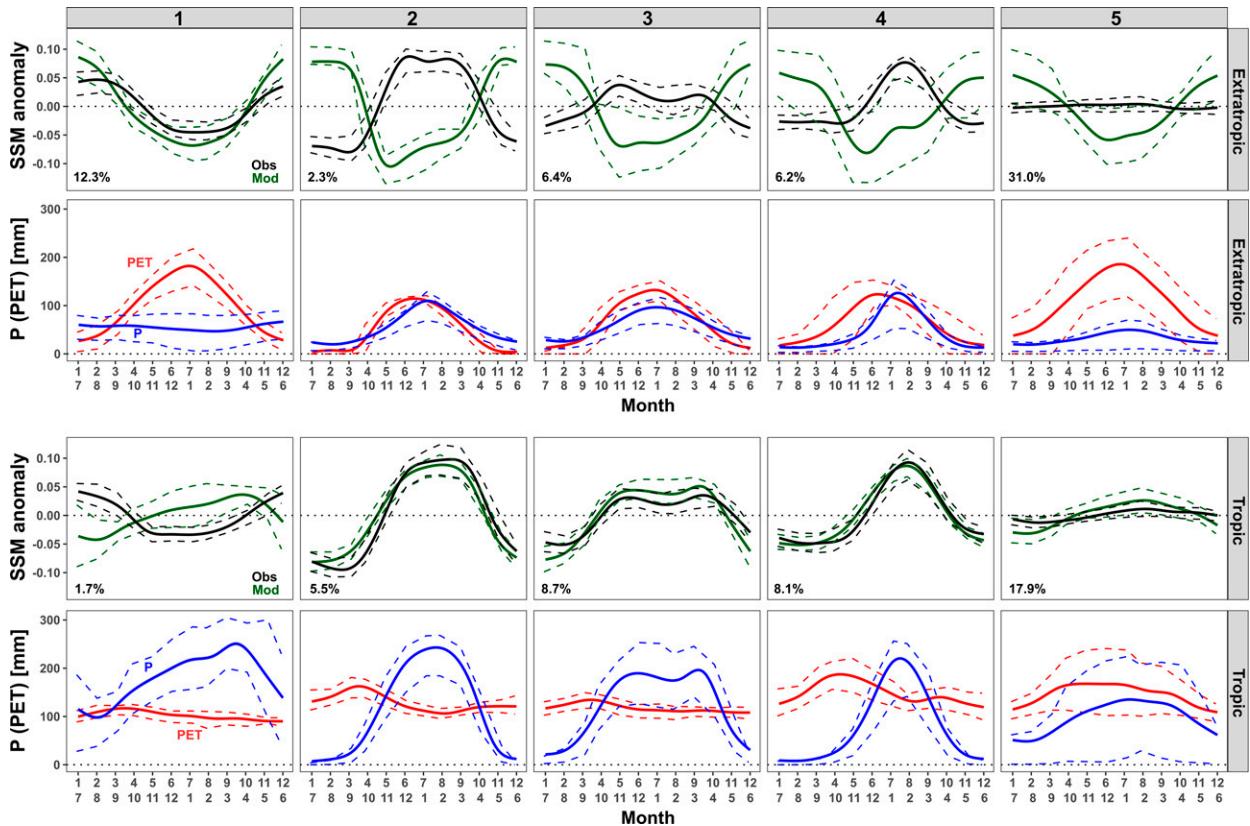


FIG. A1. As in Fig. 5, but using gridded TerraClimate PET instead of the simple formulation PET [where $PET = 0.8R_{sd}(t)/(\rho\lambda)$].

regimes in the tropics that follows the seasonal cycle of P . The most common regime in the tropics (regime 5) has a more muted seasonal cycle in P , and P does not exceed PET, on average, in this regime at any point in the year. This results in low seasonal variability in SSM in regime 5 in the tropics.

In contrast, outside the tropics, the strong seasonal cycle in R_{sd} (and, therefore, PET) dominates the seasonal cycle of SSM. Regimes 1 and 5 are both mainly observed outside the tropics, and both feature a relatively flat seasonal cycle in P and pronounced seasonal cycles in PET. However, regime 1 exhibits a strong seasonal cycle in SSM whereas regime 5 does not. Why? According to the conceptual model [Eq. (2)], variability in $s(t)$ will be small if either $PET(t) \gg P(t)$ [in which case, $s(t) \approx 0$] or $P(t) \gg PET(t)$ [in which case $s(t) \approx 1$]. PET is substantially larger than P throughout the year in regime 5, whereas PET and P have comparable magnitude in the winter in regime 1, explaining the difference between the two regimes.

4. Environmental significance

A surprising result of this study is that much of the temporal variability of surface soil saturation at seasonal time scales can be explained without reference to information on vegetation or land cover type (Figs. 5 and 6). As noted previously, even idealized models of the SSM seasonal cycle require multiple parameters that are dependent on land surface properties (e.g., Laio et al. 2002). While soil hydraulic properties are

unlikely to vary considerably on seasonal time scales, vegetation will in many regions, impacting evapotranspiration and thus soil moisture. We speculate that eco-evolutionary adaptation by plants to their environment (e.g., Eagleson 2002; Harrison et al. 2021) results in strong correlation between climate and vegetation, allowing climate variables to adequately explain variability that undeniably must include biological contributions, at least to first order.

As noted earlier, absolute values of soil moisture differ substantially between land surface models, even when driven by the same atmospheric forcings, but a linear rescaling is typically sufficient to dramatically improve the agreement between models (Koster et al. 2009). Koster et al. (2009) argued that this implies “the true information content ... of a model soil moisture product lies not in its *absolute* magnitudes but in its *time variations*” (p. 4333; italics from original). Our study provides a physical argument for why this is the case: given information on P and R_{sd} , even an extremely simple model is capable of simulating seasonal variations in SSM anomalies quite well. To the extent that models accurately simulate P and R_{sd} , they will likely accurately simulate seasonal time variations in SSM, regardless of differences in land surface model parameterizations.

5. Limitations

Beyond the limitations already discussed, we here note several other limitations of our analysis. Our results do not

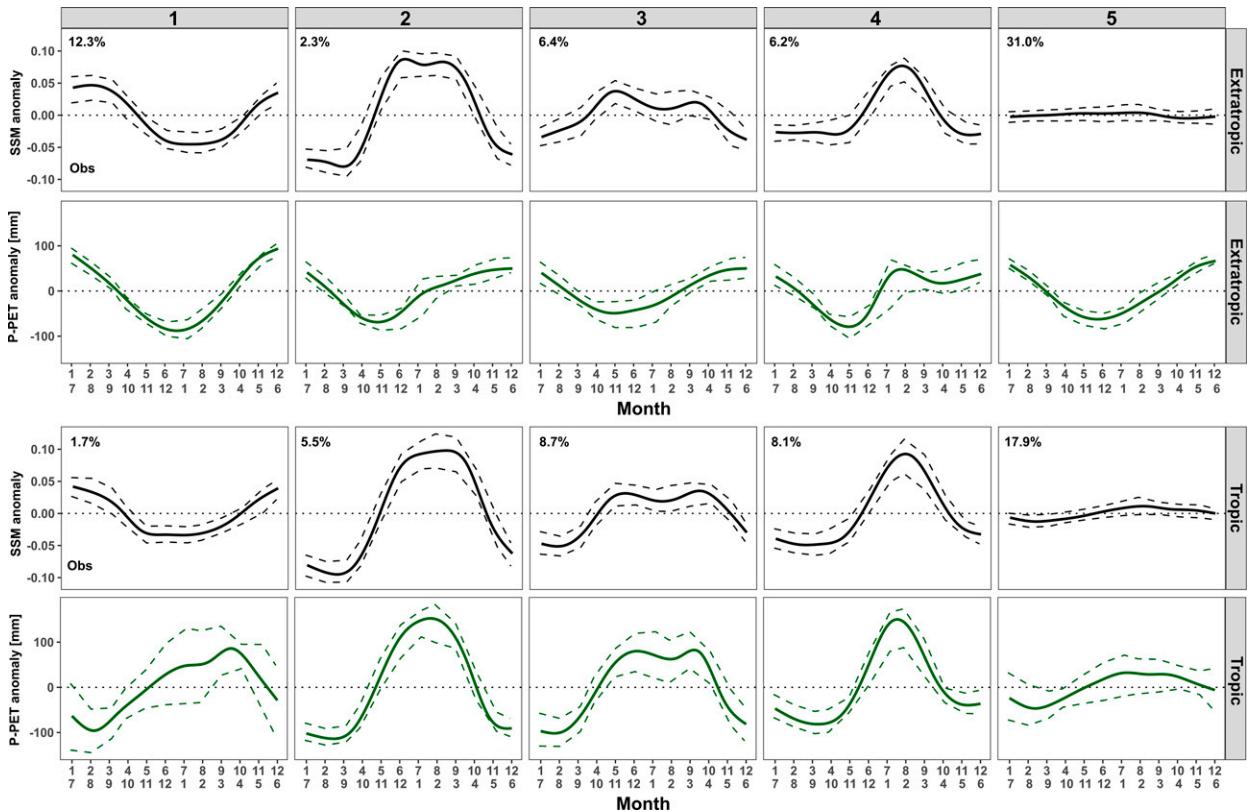


FIG. A2. Precipitation minus PET [where $PET = 0.8R_{sd}(t)/(\rho\lambda)$] anomalies seasonal cycle by region (second and fourth rows) and cluster (columns). Observed (SMAP) soil moisture anomalies by region (first and third rows) and cluster (columns). In all panels, solid lines represent the LOESS smoothed monthly values for all pixels in a given regime and hemisphere and dashed lines represent the 25th- and 75th-percentile values.

necessarily generalize to other temporal or spatial scales. Spatially, this study has focused on SSM, and its dynamics may differ substantially from those of deeper soil water storages. In our conceptual model, the approximation of $\Delta z \rightarrow 0$ (and the elimination of the time derivative in the land surface water balance) is reasonable for SSM but not for deeper soil layers. The presence of the time derivative introduces changes to the phase and amplitude of the seasonal cycle that become more important as Δz increases (e.g., Laio et al. 2002).

Temporally, this study has focused on the seasonal cycle of SSM, but the conceptual model will not necessarily generalize to other time scales. Satellite observations of soil moisture typically have a repeat time of ~ 3 days, meaning that variability at subweekly time scales is difficult to resolve globally. The satellite record is not currently long enough to examine temporal variability at longer time scales, including trends and interannual variability that have been of interest in studying surface temperatures (e.g., Stine and Huybers 2012). However, we can be confident that the conceptual model will *not* work at daily or subdaily time scales, in which case SSM exhibits sharp spikes in response to precipitation followed by exponential decay on time scales of days to weeks (McColl et al. 2017b, 2019a). The conceptual model includes no “memory” of past events, meaning its predicted SSM would spike rapidly at the onset of precipitation

before unrealistically returning to zero immediately once precipitation stops.

6. Conclusions

To our knowledge, this is the first study to characterize and explain the SSM seasonal cycle using global observations. An unsupervised clustering algorithm was used to summarize the observed seasonal cycles by partitioning them into one of five regimes (Fig. 1): one peaking in its hemisphere’s winter (regime 1), three peaking outside its hemisphere’s winter (regimes 2, 3, and 4), and one with no distinct seasonality (regime 5). Each regime appears on multiple continents, in both hemispheres, and across varied local climates. Regime 5 was most common, typically appearing in the most arid and least arid regions. Regime 1 appears mainly outside the tropics, while the reverse is true of regimes 2, 3, and 4. Furthermore, it is worth noting that the observed SSM seasonal cycle within a given regime agrees across both the tropics and extratropics (Fig. 5; see black lines for observed SSM cycles), highlighting that the cluster analysis (section 2b) on the SMAP observations performs well across all regions (i.e., the tropics and extratropics).

To understand the climatic causes of these spatial patterns, we trained a random forest model to reproduce the observed SSM regimes using climate data (Fig. 2). The model reproduced

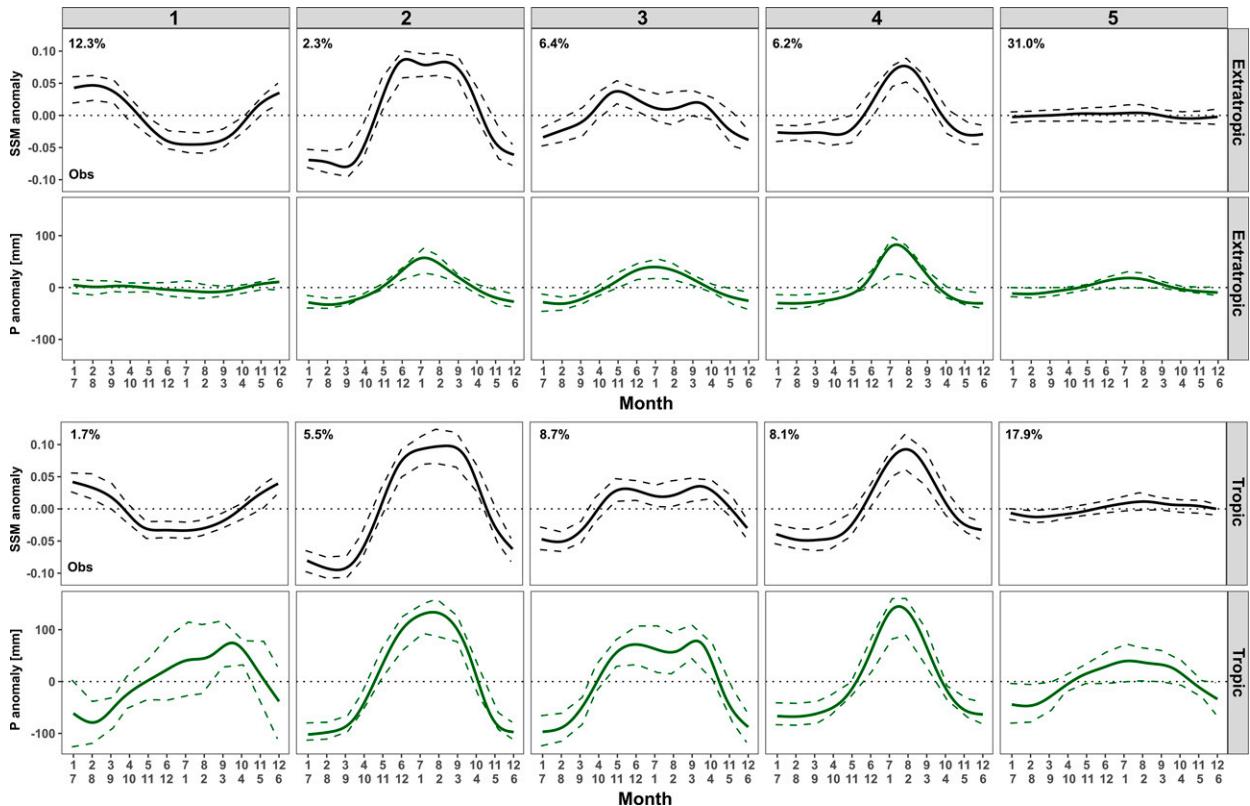


FIG. A3. Precipitation anomalies seasonal cycle by hemisphere (second and fourth rows) and cluster (columns). Observed (SMAP) soil moisture anomalies by hemisphere (second and fourth rows) and cluster (columns). In all panels, solid lines represent the LOESS smoothed monthly values for all pixels in a given regime and hemisphere and dashed lines represent the 25th- and 75th-percentile values.

the observed SSM regimes quite well with just P and R_{sd} as inputs (Fig. 3; Table 1). Information on air temperature was neither essential to, nor particularly effective in, achieving reasonable model accuracy.

To explain these results, we introduced an intentionally simple conceptual model of the SSM seasonal cycle [Eq. (2)], which only requires P and R_{sd} as inputs. The model is physically based, using pre-existing simplifications to the land surface water budget (Fig. 4), and includes zero calibration parameters. The conceptual model was broadly successful in reproducing the SSM seasonal cycle in each regime (Figs. 5 and 6), whereas alternative models of comparable complexity were not (Figs. A2 and A3). Regimes 2, 3, and 4 are mainly found in the tropics, where there is little seasonal variability in R_{sd} ; thus, seasonality in P determines seasonality in SSM in these regions. Regime 1 is mainly found outside the tropics, where seasonality in R_{sd} dominates seasonality in P , and thus determines the seasonal cycle of SSM in these regions. Regime 5 is prevalent both in the tropics and extratropics, but for different reasons. In the tropics, seasonality in both P and R_{sd} is muted for regions in regime 5, resulting in minimal SSM seasonality. In contrast, in the extratropics, seasonality in R_{sd} is large; however, $PET \propto R_{sd}$ is consistently much greater than P for extratropical regions in regime 5 throughout the year, resulting in minimal SSM seasonality.

We have shown that spatial variability in the seasonal cycle of SSM is explained well by a simple model [Eq. (2)], which requires only two inputs (P and R_{sd}) and no parameters requiring calibration. The model is surprisingly simple: for example, it does not require information on vegetation or land cover. Future work is planned to test the model at finer temporal and spatial scales. If impacts of clouds are neglected as a further approximation, it may be possible to further simplify the model by using top-of-atmosphere shortwave radiation as an input—which can be modeled accurately as a function of time and location—rather than R_{sd} . This would effectively reduce the model's required inputs to only precipitation. Since historical records of precipitation extend much further back in time than those of soil moisture, our model would effectively extend the soil moisture record back in time. This may prove useful in characterizing trends and interannual variability in the SSM seasonal cycle, two tasks that are currently not possible due to the short length of the satellite soil moisture record. In addition, to the extent that the model holds, predicted changes in global mean precipitation in a warming world based on fundamental physics (Held and Soden 2006) can be directly translated into predicted changes in global mean soil moisture; this would be useful since there are currently no simple theoretical constraints on changes in soil moisture in a warming world. More broadly, our work contributes to a growing recent literature on simple models of continental climate: in

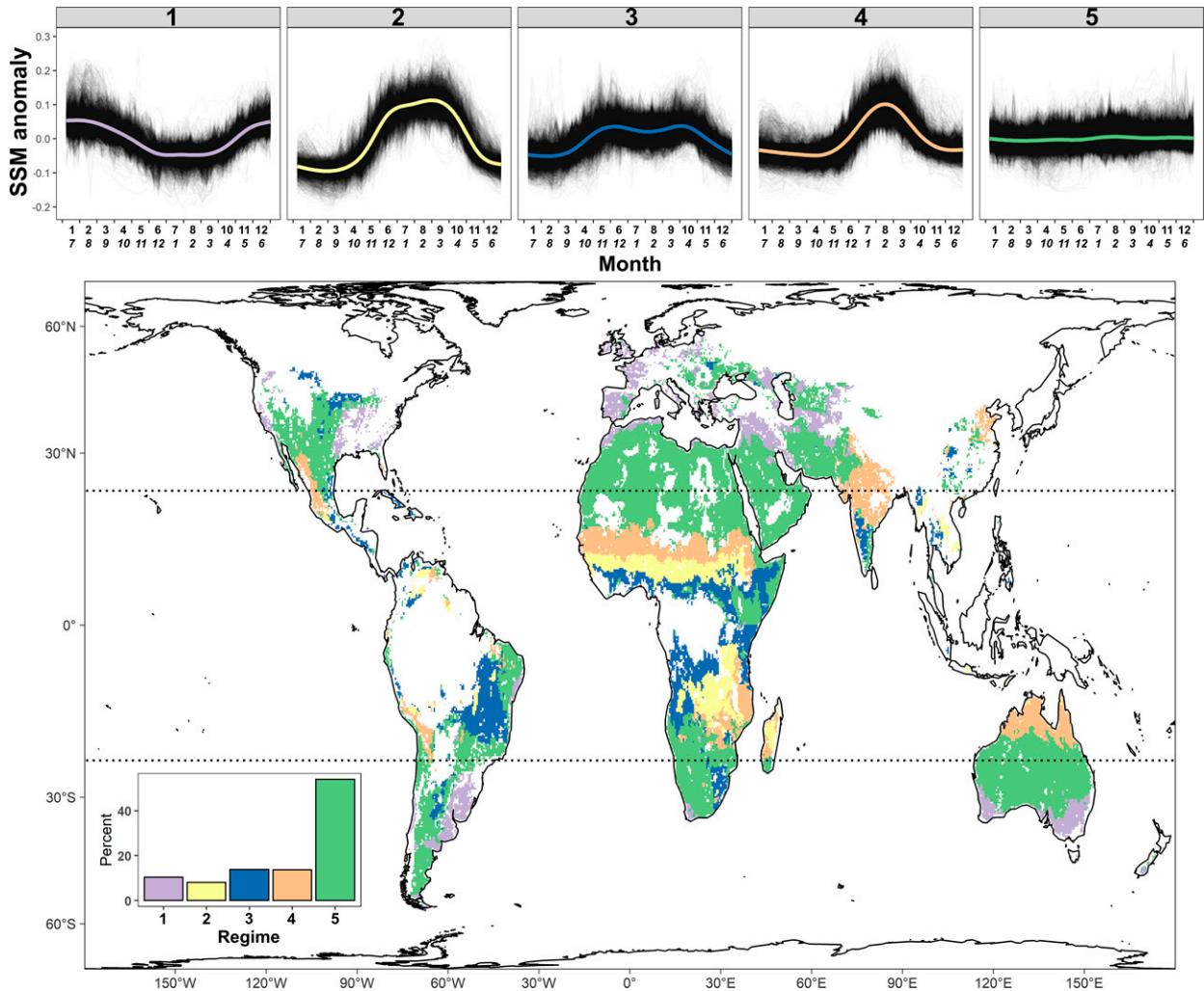


FIG. B1. As in Fig. 1, except that analysis was performed using only retrievals that had recommended quality (i.e., `retrieval_qual_flag` = 0 or 8) and complete weekly composite records (i.e., 52 weeks without any gap filling).

particular, models of evapotranspiration (McColl et al. 2019b; McColl and Rigden 2020; Chen et al. 2021), summertime surface temperature variability (Vargas Zeppetello et al. 2020a,b; Vargas Zeppetello and Battisti 2020), and relative humidity (Byrne and O’Gorman 2016, 2018), among others. Connecting these theories may provide a promising direction for future progress.

Acknowledgments. Feedback from three anonymous reviewers improved the manuscript. K.A.M. acknowledges funding from the Star-Friedman Challenge for Promising Scientific Research and the Dean’s Competitive Fund for Promising Scholarship at Harvard University. The authors declare no conflicts of interest.

Data availability statement. The data analyzed in this study are publically available. The SMAP soil moisture data are available from the NASA National Snow and Ice Data Center (NSIDC) at <https://nsidc.org/data/smap/smap-data.html>. The

TerraClimate precipitation, downwelling surface shortwave radiation, and air temperature data (Abatzoglou et al. 2018) are available on the THREDDS Data Server by the University Corporation for Atmospheric Research (UCAR) at http://thredds.northwestknowledge.net:8080/thredds/catalog/TERRACLIMATE_ALL/summaries/catalog.html. The gridded soil wilting point and field capacity data from the HiHydro-Soil (v2.0) database (Simons et al. 2020) are freely available after submitting a brief request form at <https://www.futurewater.eu/projects/hihydrosoil/> or can be directly accessed via Google Earth Engine at <https://code.earthengine.google.com/2b6f1aaf0b2acbf74a06144ed26ad606>.

APPENDIX A

Additional Figures and Tables Supporting Model Analyses

Table A1 shows the number of pixels in each category of the predicted and observed regimes. Table A2 provides

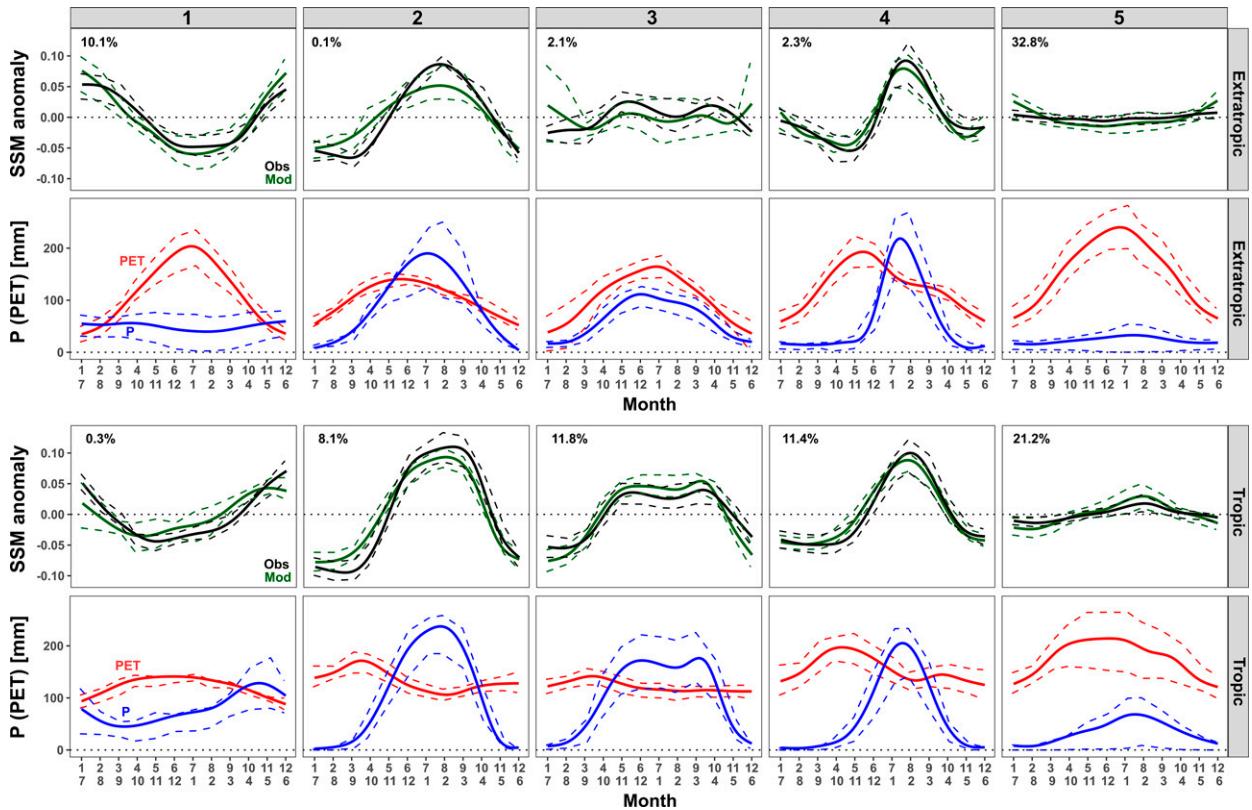


FIG. B2. As in Fig. 6, but using gridded TerraClimate PET instead of the simple formulation of $PET = 0.8R_{sd}(t)/(\rho\lambda)$.

performance statistics in each regime for the random forest model. Values are computed using the following equations:

$$\text{Sensitivity} = \frac{\text{Number of true positives}}{\text{Number of true positives} + \text{Number of false negatives}},$$

$$\text{Specificity} = \frac{\text{Number of true negatives}}{\text{Number of true negatives} + \text{Number of false positives}},$$

$$\text{PPV} = \frac{\text{Number of true positives}}{\text{Number of true positives} + \text{Number of false positives}},$$

$$\text{NPV} = \frac{\text{Number of true negatives}}{\text{Number of true negatives} + \text{Number of false negatives}}.$$

Figure A1 repeats the analysis presented in Fig. 5, except using gridded TerraClimate PET rather than the simple formulation used in Fig. 5. Figures A2 and A3 show the performance of two alternative models of $s(t)$: $P - PET$ (Fig. A2) and P (Fig. A3).

APPENDIX B

Figures Supporting Analysis with Highest-Quality Retrievals Only

Figures B1 and B2 correspond to the analysis performed using only pixels with the highest-quality retrievals (i.e., `retrieval_qual_flag = 0` or `8`) and with complete weekly composite records (i.e., 52 weeks without any gap filling). This

is approximately equivalent to excluding regions with dense forests or with periodically frozen soil.

REFERENCES

- Abatzoglou, J. T., S. Z. Dobrowski, S. A. Parks, and K. C. Hegewisch, 2018: TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci. Data*, **5**, 170191, <https://doi.org/10.1038/sdata.2017.191>.
- Akbar, R., D. Short Gianotti, K. A. McColl, E. Haghghi, G. D. Salvucci, and D. Entekhabi, 2018: Hydrological storage length scales represented by remote sensing estimates of soil moisture and precipitation. *Water Resour. Res.*, **54**, 1476–1492, <https://doi.org/10.1002/2017WR021508>.
- Berg, A., and K. A. McColl, 2021: No projected global drylands expansion under greenhouse warming. *Nat. Climate Change*, **11**, 331–337, <https://doi.org/10.1038/s41558-021-01007-8>.
- , and Coauthors, 2015: Interannual coupling between summertime surface temperature and precipitation over land: Processes and implications for climate change. *J. Climate*, **28**, 1308–1328, <https://doi.org/10.1175/JCLI-D-14-00324.1>.
- Bergen, K. J., P. A. Johnson, M. V. de Hoop, and G. C. Beroza, 2019: Machine learning for data-driven discovery in solid Earth geoscience. *Science*, **363**, eaau0323, <https://doi.org/10.1126/science.aau0323>.
- Berghuijs, W. R., S. Harrigan, P. Molnar, L. J. Slater, and J. W. Kirchner, 2019: The relative importance of different flood-generating mechanisms across Europe. *Water Resour. Res.*, **55**, 4582–4593, <https://doi.org/10.1029/2019WR024841>.

- Boehmke, B., and B. Greenwell, 2019: *Hands-On Machine Learning with R*. Chapman & Hall/CRC, 488 pp.
- Brunner, M. I., L. A. Melsen, A. J. Newman, A. W. Wood, and M. P. Clark, 2020: Future streamflow regime changes in the United States: Assessment using functional classification. *Hydrol. Earth Syst. Sci.*, **24**, 3951–3966, <https://doi.org/10.5194/hess-24-3951-2020>.
- Budyko, M. I., 1958: The heat balance of the Earth's surface. U.S. Department of Commerce, Weather Bureau, 259 pp.
- , 1974: *Climate and Life*. Academic Press, 508 pp.
- Byrne, M. P., and P. A. O’Gorman, 2016: Understanding decreases in land relative humidity with global warming: Conceptual model and GCM simulations. *J. Climate*, **29**, 9045–9061, <https://doi.org/10.1175/JCLI-D-16-0351.1>.
- , and —, 2018: Trends in continental temperature and humidity directly linked to ocean warming. *Proc. Natl. Acad. Sci. USA*, **115**, 4863–4868, <https://doi.org/10.1073/pnas.1722312115>.
- Chan, S. K., and Coauthors, 2016: Assessment of the SMAP Passive Soil Moisture Product. *IEEE Trans. Geosci. Remote Sens.*, **54**, 4994–5007, <https://doi.org/10.1109/TGRS.2016.2561938>.
- Chen, F., W. T. Crow, R. Bindlish, A. Colliander, M. S. Burgin, J. Asanuma, and K. Aida, 2018: Global-scale evaluation of SMAP, SMOS and ASCAT soil moisture products using triple collocation. *Remote Sens. Environ.*, **214**, 1–13, <https://doi.org/10.1016/j.rse.2018.05.008>.
- Chen, S., K. A. McColl, A. Berg, and Y. Huang, 2021: Surface flux equilibrium estimates of evapotranspiration at large spatial scales. *J. Hydrometeorol.*, **22**, 765–779, <https://doi.org/10.1175/JHM-D-20-0204.1>.
- Choudhury, B., 1999: Evaluation of an empirical equation for annual evaporation using field observations and results from a biophysical model. *J. Hydrol.*, **216**, 99–110, [https://doi.org/10.1016/S0022-1694\(98\)00293-5](https://doi.org/10.1016/S0022-1694(98)00293-5).
- Colliander, A., and Coauthors, 2017: Validation of SMAP surface soil moisture products with core validation sites. *Remote Sens. Environ.*, **191**, 215–231, <https://doi.org/10.1016/j.rse.2017.01.021>.
- , and Coauthors, 2020: SMAP detects soil moisture under temperate forest canopies. *Geophys. Res. Lett.*, **47**, e2020GL089697, <https://doi.org/10.1029/2020GL089697>.
- Connolly, C. T., M. O. Stahl, B. A. DeYoung, and B. C. Bostick, 2022: Surface flooding as a key driver of groundwater arsenic contamination in Southeast Asia. *Environ. Sci. Technol.*, **56**, 928–937, <https://doi.org/10.1021/acs.est.1c05955>.
- Daubenmire, R., 1968: Soil moisture in relation to vegetation distribution in the mountains of northern Idaho. *Ecology*, **49**, 431–438, <https://doi.org/10.2307/1934109>.
- Dwyer, J. G., M. Biasutti, and A. H. Sobel, 2012: Projected changes in the seasonal cycle of surface temperature. *J. Climate*, **25**, 6359–6374, <https://doi.org/10.1175/JCLI-D-11-00741.1>.
- Eagleson, P. S., 2002: *Ecohydrology: Darwinian Expression of Vegetation Form and Function*. Cambridge University Press, 484 pp.
- Entekhabi, D., and Coauthors, 2010: The Soil Moisture Active Passive (SMAP) mission. *Proc. IEEE*, **98**, 704–716, <https://doi.org/10.1109/JPROC.2010.2043918>.
- Feldman, A. F., D. J. Short Gianotti, A. G. Konings, K. A. McColl, R. Akbar, G. D. Salvucci, and D. Entekhabi, 2018: Moisture pulse-reserve in the soil–plant continuum observed across biomes. *Nat. Plants*, **4**, 1026–1033, <https://doi.org/10.1038/s41477-018-0304-9>.
- Feng, X., G. Vico, and A. Porporato, 2012: On the effects of seasonality on soil water balance and plant growth. *Water Resour. Res.*, **48**, W05543, <https://doi.org/10.1029/2011WR011263>.
- , A. Porporato, and I. Rodriguez-Iturbe, 2015: Stochastic soil water balance under seasonal climates. *Proc. Roy. Soc.*, **471A**, 20140623, <https://doi.org/10.1098/rspa.2014.0623>.
- Fernández-Long, M. E., M. Peretti, D. Carnelos, T. Della-Chiesa, and L. Spescha, 2021: Characterization of the seasonal variation of soil moisture in Argentina. *Int. J. Climatol.*, **41**, E537–E547, <https://doi.org/10.1002/joc.6705>.
- Ford, T. W., E. Harris, and S. M. Quiring, 2014: Estimating root zone soil moisture using near-surface observations from SMOS. *Hydrol. Earth Syst. Sci.*, **18**, 139–154, <https://doi.org/10.5194/hess-18-139-2014>.
- Greve, P., M. L. Roderick, A. M. Ukkola, and Y. Wada, 2019: The aridity index under global warming. *Environ. Res. Lett.*, **14**, 124006, <https://doi.org/10.1088/1748-9326/ab5046>.
- Harrison, S. P., and Coauthors, 2021: Eco-evolutionary optimality as a means to improve vegetation and land-surface models. *New Phytol.*, **231**, 2125–2141, <https://doi.org/10.1111/nph.17558>.
- Hartigan, J. A., and M. A. Wong, 1979: Algorithm AS 136: A K-means clustering algorithm. *Appl. Stat.*, **28**, 100–108, <https://doi.org/10.2307/2346830>.
- Held, I. M., and B. J. Soden, 2006: Robust responses of the hydrological cycle to global warming. *J. Climate*, **19**, 5686–5699, <https://doi.org/10.1175/JCLI3990.1>.
- Hijmans, R. J., 2020: Raster: Geographic data analysis and modeling. R package version 2.4-15, <http://CRAN.R-project.org/package=raster>.
- Hollinger, S. E., and S. A. Isard, 1994: A soil moisture climatology of Illinois. *J. Climate*, **7**, 822–833, [https://doi.org/10.1175/1520-0442\(1994\)007<0822:ASMC0I>2.0.CO;2](https://doi.org/10.1175/1520-0442(1994)007<0822:ASMC0I>2.0.CO;2).
- Homyak, P. M., and J. O. Sickman, 2014: Influence of soil moisture on the seasonality of nitric oxide emissions from chaparral soils, Sierra Nevada, California, USA. *J. Arid Environ.*, **103**, 46–52, <https://doi.org/10.1016/j.jaridenv.2013.12.008>.
- Illston, B. G., J. B. Basara, and K. C. Crawford, 2004: Seasonal to interannual variations of soil moisture measured in Oklahoma. *Int. J. Climatol.*, **24**, 1883–1896, <https://doi.org/10.1002/joc.1077>.
- James, G., D. Witten, T. Hastie, and R. Tibshirani, 2021: Tree-based methods. *An Introduction to Statistical Learning: With Applications in R*, Springer, 327–365.
- Katul, G. G., A. Porporato, E. Daly, A. C. Oishi, H. S. Kim, P. C. Stoy, J. Y. Juang, and M. B. Siqueira, 2007: On the spectrum of soil moisture from hourly to interannual scales. *Water Resour. Res.*, **43**, W05428, <https://doi.org/10.1029/2006WR005356>.
- Kirkyla, K. I., and S. Hameed, 1989: Harmonic analysis of the seasonal cycle in precipitation over the United States: A comparison between observations and a general circulation model. *J. Climate*, **2**, 1463–1475, [https://doi.org/10.1175/1520-0442\(1989\)002<1463:HAOTSC>2.0.CO;2](https://doi.org/10.1175/1520-0442(1989)002<1463:HAOTSC>2.0.CO;2).
- Koster, R. D., and M. J. Suarez, 2001: Soil moisture memory in climate models. *J. Hydrometeorol.*, **2**, 558–570, [https://doi.org/10.1175/1525-7541\(2001\)002<0558:SMMICM>2.0.CO;2](https://doi.org/10.1175/1525-7541(2001)002<0558:SMMICM>2.0.CO;2).
- , and S. P. Mahanama, 2012: Land surface controls on hydroclimatic means and variability. *J. Hydrometeorol.*, **13**, 1604–1620, <https://doi.org/10.1175/JHM-D-12-050.1>.
- , and Coauthors, 2006: GLACE: The Global Land–Atmosphere Coupling Experiment. Part I: Overview. *J. Hydrometeorol.*, **7**, 590–610, <https://doi.org/10.1175/JHM510.1>.
- , Z. Guo, R. Yang, P. A. Dirmeyer, K. Mitchell, and M. J. Puma, 2009: On the nature of soil moisture in land surface models. *J. Climate*, **22**, 4322–4335, <https://doi.org/10.1175/2009JCLI2832.1>.
- Laio, F., A. Porporato, L. Ridolfi, and I. Rodriguez-Iturbe, 2002: On the seasonal dynamics of mean soil moisture. *J. Geophys. Res.*, **107**, 4272, <https://doi.org/10.1029/2001JD001252>.
- Li, R. Y., J. L. Stroud, J. F. Ma, S. P. McGrath, and F. J. Zhao, 2009: Mitigation of arsenic accumulation in rice with water

- management and silicon fertilization. *Environ. Sci. Technol.*, **43**, 3778–3783, <https://doi.org/10.1021/es803643v>.
- Maes, W. H., P. Gentine, N. E. C. Verhoest, and D. G. Miralles, 2019: Potential evaporation at eddy-covariance sites across the globe. *Hydrol. Earth Syst. Sci.*, **23**, 925–948, <https://doi.org/10.5194/hess-23-925-2019>.
- Markham, C. G., 1970: Seasonality of precipitation in the United States. *Ann. Assoc. Amer. Geogr.*, **60**, 593–597, <https://doi.org/10.1111/j.1467-8306.1970.tb00743.x>.
- McColl, K. A., and A. J. Rigden, 2020: Emergent simplicity of continental evapotranspiration. *Geophys. Res. Lett.*, **47**, e2020GL087101, <https://doi.org/10.1029/2020GL087101>.
- , 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, <https://doi.org/10.1038/ngeo2868>.
- , 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, <https://doi.org/10.1002/2017GL072819>.
- , Q. He, H. Lu, and D. Entekhabi, 2019a: Short-term and long-term surface soil moisture memory time scales are spatially anticorrelated at global scales. *J. Hydrometeorol.*, **20**, 1165–1182, <https://doi.org/10.1175/JHM-D-18-0141.1>.
- , G. D. Salvucci, and P. Gentine, 2019b: Surface flux equilibrium theory explains an empirical estimate of water-limited daily evapotranspiration. *J. Adv. Model. Earth Syst.*, **11**, 2036–2049, <https://doi.org/10.1029/2019MS001685>.
- Milly, P. C. D., 1994a: Climate, soil water storage, and the average annual water balance. *Water Resour. Res.*, **30**, 2143–2156, <https://doi.org/10.1029/94WR00586>.
- , 1994b: Climate, interseasonal storage of soil water, and the annual water balance. *Adv. Water Resour.*, **17**, 19–24, [https://doi.org/10.1016/0309-1708\(94\)90020-5](https://doi.org/10.1016/0309-1708(94)90020-5).
- , and K. A. Dunne, 2016: Potential evapotranspiration and continental drying. *Nat. Climate Change*, **6**, 946–949, <https://doi.org/10.1038/nclimate3046>.
- Nandintsetseg, B., and M. Shinoda, 2011: Seasonal change of soil moisture in Mongolia: Its climatology and modelling. *Int. J. Climatol.*, **31**, 1143–1152, <https://doi.org/10.1002/joc.2134>.
- Orth, R., and S. I. Seneviratne, 2012: Analysis of soil moisture memory from observations in Europe. *J. Geophys. Res.*, **117**, D15115, <https://doi.org/10.1029/2011JD017366>.
- R Core Team, 2020: R: A language and environment for statistical computing, <https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing>.
- Reichle, R. H., R. D. Koster, J. Dong, and A. A. Berg, 2004: Global soil moisture from satellite observations, land surface models, and ground data: Implications for data assimilation. *J. Hydrometeorol.*, **5**, 430–442, [https://doi.org/10.1175/1525-7541\(2004\)005<0430:GSMFSO>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0430:GSMFSO>2.0.CO;2).
- Rickard, W. H., 1967: Seasonal soil moisture patterns in adjacent greasewood and sagebrush stands. *Ecology*, **48**, 1034–1038, <https://doi.org/10.2307/1934561>.
- Roberts, L. C., and Coauthors, 2010: Arsenic release from paddy soils during monsoon flooding. *Nat. Geosci.*, **3**, 53–59, <https://doi.org/10.1038/ngeo723>.
- , and Coauthors, 2011: Arsenic dynamics in porewater of an intermittently irrigated paddy field in Bangladesh. *Environ. Sci. Technol.*, **45**, 971–976, <https://doi.org/10.1021/es102882q>.
- Scheff, J., J. S. Mankin, S. Coats, and H. Liu, 2021: CO₂-plant effects do not account for the gap between dryness indices and projected dryness impacts in CMIP6 or CMIP5. *Environ. Res. Lett.*, **16**, 034018, <https://doi.org/10.1088/1748-9326/abd8fd>.
- Short Gianotti, D. J., G. D. Salvucci, R. Akbar, K. A. McColl, R. Cuenca, and D. Entekhabi, 2019: Landscape water storage and subsurface correlation from satellite surface soil moisture and precipitation observations. *Water Resour. Res.*, **55**, 9111–9132, <https://doi.org/10.1029/2019WR025332>.
- Simons, G. W. H., R. Koster, and P. Droogers, 2020: HiHydroSoil v2.0—High resolution soil maps of global hydraulic properties. FutureWater Report 213, 18 pp., <https://www.futurewater.nl/wp-content/uploads/2020/10/HiHydroSoil-v2.0-High-Resolution-Soil-Maps-of-Global-Hydraulic-Properties.pdf>.
- Slessarev, E. W., Y. Lin, N. L. Bingham, J. E. Johnson, Y. Dai, J. P. Schimel, and O. A. Chadwick, 2016: Water balance creates a threshold in soil pH at the global scale. *Nature*, **540**, 567–569, <https://doi.org/10.1038/nature20139>.
- Stine, A. R., and P. Huybers, 2012: Changes in the seasonal cycle of temperature and atmospheric circulation. *J. Climate*, **25**, 7362–7380, <https://doi.org/10.1175/JCLI-D-11-00470.1>.
- , —, and I. Y. Fung, 2009: Changes in the phase of the annual cycle of surface temperature. *Nature*, **457**, 435–440, <https://doi.org/10.1038/nature07675>.
- Strobl, C., A.-L. Boulesteix, A. Zeileis, and T. Hothorn, 2007: Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinf.*, **8**, 25, <https://doi.org/10.1186/1471-2105-8-25>.
- Thorntwaite, C. W., 1948: An approach toward a rational classification of climate. *Geogr. Rev.*, **38**, 55–94, <https://doi.org/10.2307/210739>.
- Tramblay, Y., G. Villarini, E. M. El Khalki, G. Gründemann, and D. Hughes, 2021: Evaluation of the drivers responsible for flooding in Africa. *Water Resour. Res.*, **57**, e2021WR029595, <https://doi.org/10.1029/2021WR029595>.
- Van den Berg, G. A., J. P. G. Loch, and H. J. Winkels, 1998: Effect of fluctuating hydrological conditions on the mobility of heavy metals in soils of a freshwater estuary in the Netherlands. *Water Air Soil Pollut.*, **102**, 377–388, <https://doi.org/10.1023/A:1004920700598>.
- Vargas Zeppetello, L. R., and D. S. Battisti, 2020: Projected increases in monthly midlatitude summertime temperature variance over land are driven by local thermodynamics. *Geophys. Res. Lett.*, **47**, e2020GL090197, <https://doi.org/10.1029/2020GL090197>.
- , —, and M. B. Baker, 2020a: A new look at the variance of summertime temperatures over land. *J. Climate*, **33**, 5465–5477, <https://doi.org/10.1175/JCLI-D-19-0887.1>.
- , É. Tétrault-Pinard, D. S. Battisti, and M. B. Baker, 2020b: Identifying the sources of continental summertime temperature variance using a diagnostic model of land–atmosphere interactions. *J. Climate*, **33**, 3547–3564, <https://doi.org/10.1175/JCLI-D-19-0276.1>.
- Wang, J. L., J. M. Chiou, and H. G. Müller, 2016: Functional data analysis. *Annu. Rev. Stat. Appl.*, **3**, 257–295, <https://doi.org/10.1146/annurev-statistics-041715-033624>.
- Wasko, C., R. Nathan, and M. C. Peel, 2020: Changes in antecedent soil moisture modulate flood seasonality in a changing climate. *Water Resour. Res.*, **56**, e2019WR026300, <https://doi.org/10.1029/2019WR026300>.
- Yang, H., D. Yang, Z. Lei, and F. Sun, 2008: New analytical derivation of the mean annual water-energy balance equation. *Water Resour. Res.*, **44**, W03410, <https://doi.org/10.1029/2007WR006135>.