Persistence and memory timescales in root-zone soil moisture dynamics

Khaled Ghannam,[1] Taro Nakai,[2,3] Athanasios Paschalis,[4] Christopher A. Oishi,[5] Ayumi Kotani,[6] Yasunori Igarashi,[2] Tomo'omi Kumagai,[2] Gabriel G. Katul[1]

[1]Nicholas School of the Environment and

Abstract. The memory timescale that characterizes root-zone soil moisture remains the dominant measure in seasonal forecasts of land-climate interactions. This memory is a quasi-deterministic timescale associated with the losses (e.g. evapotranspiration) from the soil column and is often interpreted as persistence in soil moisture states. Persistence, however, represents a distribution of time periods where soil moisture resides above or below some prescribed threshold, and is therefore inherently probabilistic. Using multi-

Earth Sciences, Duke University, Durham, North Carolina, USA.

[2]Hydrospheric Atmospheric Research Center, Nagoya University, Nagoya, Aichi, Japan.

[3]International Arctic Research Center, University of Alaska Fairbanks, Fairbanks, Alaska, USA.

[4]Faculty of Engineering and the Environment, University of Southampton, Southampton, UK.

[5]Coweeta Hydrologic Laboratory, USDA Forest Service, Otto, NC, USA.

[6]Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Japan.

ple soil moisture datasets collected at high resolution (sub-hourly) across different biomes and climates, this paper explores the differences, underlying dynamics, and relative importance of memory and persistence timescales in root-zone soil moisture. A first-order Markov process, commonly used to interpret soil moisture fluctuations derived from climate simulations, is also used as a reference model. Persistence durations of soil moisture below the plant water-stress level (chosen as the threshold), and the temporal spectrum of up- and down-crossings of this threshold, are compared to the memory timescale and spectrum of the full time series, respectively. The results indicate that despite the differences between meteorological drivers, the spectrum of threshold-crossings is similar across sites, and follows a unique relation with that of the full soil moisture series. The distribution of persistence times exhibits an approximate stretched exponential type and reflects a likelihood of exceeding the memory at all sites. However, the rainfall counterpart of these distributions shows that persistence of dry atmospheric periods is less likely at sites with long soil moisture memory. The cluster exponent, a measure of the density of threshold crossings in a time frame, reveals that the clustering tendency in rainfall events (on-off switches) does not translate directly to clustering in soil moisture. This is particularly the case in climates where rainfall and evapotranspiration are out of phase, resulting in less ordered (more independent) persistence in soil moisture than in rainfall.

1. Introduction

Water storage within the soil pores is governed by nonlinear interactions among multiple hydro-meteorological and biophysical processes (e.g. rainfall, evapotranspiration, surface runoff, and subsurface flow). These storage effects, particularly within the root-zone, tend to last for several weeks and are perceived as a principal 'modulator' of short-term atmospheric anomalies and 'driver' of longer term seasonal forecasts of over-land atmospheric states (e.g. summer rainfall), droughts, and floods. The timescales that characterize root-zone soil moisture variability associated with these nonlinear interactions are of significance to a variety of disciplines. This is apparent when noting the wide range of studies addressing the role of soil moisture in land-atmosphere feedbacks and rainfall [*Delworth and Manabe*, 1988; *Parlange et al.*, 1992; *Entekhabi et al.*, 1996; *Findell and Eltahir*, 1997; *Koster and Suarez*, 2001; *Wu et al.*, 2002; *Wu and Dickinson*, 2004; *Alfieri et al.*, 2008; *Juang et al.*, 2007], biogeochemical cycling and ecosystem resilience [*D'Odorico et al.*, 2003; *Porporato et al.*, 2004; *Guan et al.*, 2011; *Parolari et al.*, 2014; *Paschalis et al.*, 2015], overland and streamflow generation [*Thompson and Katul*, 2012; *Paschalis et al.*, 2014a], large-scale floods [*Milly et al.*, 2002], ponding and onset of water-born diseases [*Montosi et al.*, 2012], agriculture-food security [*Parent et al.*, 2006; *Lauzon et al.*, 2004], and soil microbial processes [*Daly et al.*, 2008; *Manzoni and Katul*, 2014].

One key characteristic of the soil water storage effect is the 'memory' timescale, which is a rough measure of the time needed by the soil column to 'forget' an imposed anomaly (such as a rainfall event or lack thereof). Commonly calculated from the corresponding time-lagged auto-correlation function, memory is typically on the order of a week to few months (depending on soil properties and meteorological/biophysical variables), and reflects the tendency of the temporal statistics of soil moisture to maintain a finite temporal correlation. In analogy with oceans as heat reservoirs in ocean-atmosphere coupling, soil moisture memory is invariably relied upon as a measure for seasonal projections of land-climate interactions. Examples of its use include studies on soil moisture feedback on convective rainfall [*Alfieri et al.*, 2008], summer heat waves [*Fischer et al.*, 2007; *Lorenz and Seneviratne*, 2010], and general impact on the climate system [*Seneviratne et al.*, 2006, 2010].

Fairly often, this 'memory' timescale is treated as a surrogate for 'persistence' in soil moisture states and the two terms are used interchangeably to emphasize that the effects of a short-term forcing, such as a storm event, may persist within the soil column long after the forcing ceases [e.g. *Seneviratne et al.*, 2006, 2010]. Nevertheless, persistence in non-equilibrium systems (e.g. soil moisture) represents a different timescale in its definition and underlying dynamics. In simple words, for a process that evolves in time according to some dynamics, persistence represents the probability that this process remains in a prescribed state (e.g. below/above some threshold or within a certain range) [*Majumdar*, 1999]. Driven by external forcing, non-equilibrium systems tend to exit and re-enter such states in the course of time, and hence persistence theory encompasses the probability distribution of the time periods spent below/above the prescribed threshold. The density of switching between states per unit time and the threshold-crossing statistics (in time and spectral domains) are indicative of clustering and intermittency in the process (see *Bershadskii et al.* [2004] and *Sreenivasan and Bershadskii* [2006] for applications in turbulence and convection). These concepts are widely used in non-equilibrium systems and stochastic models to characterize the time periods where a system dwells below/above some threshold [*Majumdar*, 1999]. Theoretical and experimental studies for many systems showed that this persistence probability decays as a power law at late times, $P_0(t) \sim t^{-\kappa}$ as $t \to \infty$ [*Majumdar*, 1999], where $P_0(t)$ is the probability that the system remains in the prescribed state up to time (t) and the exponent κ is usually nontrivial.

The concepts of persistence below or above some threshold, and the crossing properties of this threshold are not uncommon in hydrological time series analysis and modeling [*Bras and Rodríguez-Iturbe*, 1985]. For instance, durations where a river flow remains above some design threshold are equivalent to flooding periods. Similarly, the duration between two consecutive up-crossings (down-crossings) of this threshold represents the time between successive floods (droughts). The distribution of inter-arrival times between rainfall events (dry periods) used in rainfall and eco-hydrological studies [*Laio et al.*, 2001; *Molini et al.*, 2009; *Paschalis et al.*, 2013, 2014b] is equivalent to a persistence probability. In the context of root-zone soil moisture, applications of these concepts include discussions on analytical approaches to estimate mean first passage times and crossing dynamics of a prescribed threshold [e.g. *Rodríguez-Iturbe and Porporato*, 2005; *Borgogno et al.*, 2010; *Vico and Porporato.*, 2013]. These approaches typically assume a probability distribution for the occurrence (marked Poisson process) and depth (exponential) of rainfall at the *daily* timescale. Perhaps due to the dearth of high frequency soil moisture measurements (such as sub-daily), the latter approximations may mask the significance of higher frequency dynamics such as storm durations and their own persistence, which constitute the atmospheric forcing on root-zone soil moisture.

The two timescales (memory and persistence) encode different information about the dynamics of root-zone soil moisture, where memory is largely dictated by evapotranspiration and drainage losses and is essentially quasi-deterministic [*Delworth and Manabe*, 1988], and persistence is primarily forced by rainfall and is therefore inherently probabilistic. Distinguishing between these timescales can have implications on the land-atmosphere interaction schemes used in climate models, which rely on soil moisture memory for improving their predictive skill in seasonal forecasts [*Seneviratne et al.*, 2006]. However, since the wetness/dryness of the soil column largely controls the energy fluxes at the surface, persistence timescales (indicative of wet/dry states) may be more relevant than memory (correlation timescale) as a measure of the land-atmosphere coupling strength. This is especially the case when noting that persistence represents a distribution of timescales below/above a threshold value, whereas regional and general circulation models (RCM and GCM) use simplified approximations of the auto-correlation function to estimate memory [*Koster and Suarez*, 2001]. The reliability of such approximations is often affected by the non-stationarity of the soil moisture time series and hence the (lack of) stability of the corresponding auto-correlation function (sensitivity to the length of the time series, sampling frequency, periodicity such as seasonality and interannual variability).

As a starting point for characterizing a persistence timescale, this work examines the statistics and scaling laws of persistence of dry/wet states for several root-zone soil moisture time series sampled at high resolution (sub-hourly) that experience different vegetation cover and climatic forcing (mainly quantified by phase relations between evapotranspiration and rainfall). The soil moisture level below which plants become water-stressed is chosen as a threshold for dry/wet states. In particular, the probability distribution of these persistence scales, the frequency of threshold-crossing (clustering properties), their temporal correlation (spectrum), and how do these scales compare to the widely used soil moisture memory are addressed here. Whenever applicable, similar analysis is conducted on rainfall (persistence of dry events) as a means for explaining soil moisture persistence. The soil moisture, meteorological variables, and rainfall datasets were measured at different locations encompassing a variety of soil properties, vegetation and climatic regimes to allow for an assessment of the impact of these mechanisms on persistence and memory. While a distribution of persistence times of soil moisture at high frequencies is not yet theoretically accessible, the work here serves to initiate a discussion on the characteristics and relative importance of persistence and memory timescales. Although persistence is seemingly a more relevant measure of land-atmosphere coupling, the question of how to use such a distribution, or characteristics thereof, in lieu of memory (single timescale) in land-climate models remains open for further investigation. Connections between persistence and memory in soil moisture content may be provided through analogies to other systems such as those exhibiting self-organized criticality and intermittency corrections thereto, although a rigorous treatment of such connections is still lacking and outside of the scope here.

2. Theory

A brief presentation of the governing equations and theoretical background used in the analysis of the soil moisture time series is first provided. Further details can be found in the work of *Majumdar* [1999] and *Perlekar et al.* [2011] on persistence in nonequilibrium dynamics and statistical mechanics, of *Bershadskii et al.* [2004], *Sreenivasan and Bershadskii* [2006], *Cava and Katul* [2009] and *Chamecki* [2013] on applications in turbulence research, and *Laio et al.* [2001], *Rodríguez-Iturbe and Porporato* [2005] and *Molini et al.* [2009] for applications in hydrological contexts.

2.1. Soil Water Balance

For planar homogeneous conditions, the vertically integrated mass conservation equation for soil moisture across the active root-zone depth is given by

$$\eta Z_{\mathrm{r}} \frac{\mathrm{d}s(t)}{\mathrm{d}t} = \Phi[s(t), t] - \chi[s(t), t], \tag{1}$$

where t [T] is time, Z_{r} [L] is the root-zone depth, η [L^3 L^{-3}] is the soil porosity, $s(t)$ [L^3 L^{-3}] is the effective soil moisture ($0 \leq s(t) \leq 1$), $\Phi[s(t), t]$ [L T^{-1}] and $\chi[s(t), t]$ [L T^{-1}] are rates of infiltration from rainfall and soil moisture losses from the active root-zone depth, respectively. The term $\Phi[s(t), t]$ is the stochastic component in equation (1) and is represented by

$$\Phi[s(t), t] = P(t) - Q[s(t), t], \tag{2}$$

where the net rainfall (henceforth throughfall) $P(t) = R(t) - I(t)$ is the difference between the rainfall rate $R(t)$, and the fraction of $R(t)$ intercepted by the canopy cover, $I(t)$. The statistics (inter-arrival times and depth) of $P(t)$ and $R(t)$ are considered identical to each other, only censored and rescaled due to a loss fraction $I(t)$. The second term on the right hand side (r.h.s) of equation (2) ($Q[s(t),t]$) is the rate of surface runoff, which is significant when $P(t)$ exceeds the soil moisture saturation deficit and/or the soil saturated hydraulic conductivity. The dominant runoff mechanism at the sites considered here is saturation excess and the analyzed data (described later) show that measured $s(t)$ rarely reaches saturation at all sites. $Q[s(t), t]$ is hence neglected since the main interest here is in the fraction of $P(t)$ that reaches the root-zone. The loss function in equation (1) is given as

$$\chi[s(t),t] = ET[s(t),t] + D_{\mathrm{r}}[s(t)], \tag{3}$$

reflecting the sum of losses due to evapotranspiration (transpiration and soil evaporation) (ET) and subsurface drainage (D_{r}). The dependence of $\chi[s(t),t]$ on $s(t)$ is expressed as a piecewise function [see e.g. *Laio et al.*, 2001] controlled by characteristic soil moisture levels, namely the hygroscopic point s_{h}, the wilting point s_{w}, the plant water-stress level s^*, and the field capacity s_{fc}, with $s_{\mathrm{h}} < s_{\mathrm{w}} < s^* < s_{\mathrm{fc}}$. The characteristic value s_{h} (depends on soil type) represents the soil moisture level below which bare soil evaporation becomes negligible, whereas s_{w} (depends on soil and vegetation types) is the value below which plant stomata are completely closed and transpiration ceases. These soil moisture levels are small and no dynamics below them is further considered here. The value s^* depends on soil properties and vegetation type and represents the soil moisture level below which plants start reducing transpiration (control stomatal opening) to conserve water, i.e. become water-stressed. The values of s_{w} and s^* are commonly quantified by the plant-specific water potential (or equivalently by the site-specific soil matric potential) with typical values between -3 MPa to -0.03 MPa, respectively. The soil field capacity in the root-zone (s_{fc}) depends on soil and root-induced porosity and is the maximum water-holding capacity per unit volume of the soil. At hourly and daily timescales, the second term on the r.h.s of equation (3) is quasi-instantaneous and considerable only when soil moisture approaches its field capacity (s_{fc}). *Katul et al.* [2007] argued that this term (D_{r}) may still be important at longer timescales, and its nonlinear dependence on the variable s may contribute to low frequency variations in the soil moisture spectrum. An empirical representation of this term driven by gravitational drainage is [*Clapp and Hornberger*, 1978]

$$D_{\mathrm{r}}(s) = K_s s^c, \tag{4}$$

for $s > s^*$, where K_s is the soil saturated hydraulic conductivity at Z_{r} and c is an exponent that varies with pore-size distribution. Typical values of c range between ≈ 11 to ≈ 26 for (loamy) sands and clays, respectively [*Clapp and Hornberger*, 1978]. Below s_{fc}, ET is the dominant component in the loss function and is at its maximum value, the potential evapotranspiration (PET) for $s(t) > s^*$. Note that PET, while independent of soil moisture $s(t)$, is controlled by vegetation type and climatic factors (wind speed, radiation, air temperature, humidity, soil type). A common approximation of PET uses the Penman-Monteith equation [*Monteith*, 1965]. In the water-limited regime ($s < s^*$), in addition to the previous factors (vegetation and climatic), ET becomes a function of s, where in its simplest form, this dependence is assumed to be quasi-linear [*Laio et al.*, 2001; *Katul et al.*, 2007]

$$ET = PET \frac{s - s_{\mathrm{w}}}{s^* - s_{\mathrm{w}}}, \tag{5}$$

for $s_{\mathrm{w}} \leq s \leq s^*$, and s_{w} is the wilting point defined above. Another parametrization of the dependence of the loss function on soil moisture are sigmoidal functions (for instance hyperbolic tangents) [*Budyko*, 1961, 1974]. Such regime shifts in the dependence of the loss function on soil moisture are expressed by rewriting equation (1) as

$$\eta Z_{\mathrm{r}} \frac{\mathrm{d}s(t)}{\mathrm{d}t} = P(t) - PET \frac{s - s_{\mathrm{w}}}{s^* - s_{\mathrm{w}}}, \tag{6}$$

when $s_{\mathrm{w}} \leq s \leq s^*$, and as

$$\eta Z_{\mathrm{r}} \frac{\mathrm{d}s(t)}{\mathrm{d}t} = P(t) - PET - K_s \left(\frac{s - s^*}{s_{\mathrm{fc}} - s^*} \right)^c, \tag{7}$$

when $s^* < s \leq s_{\mathrm{fc}}$. While gravitational drainage is fast (sub-daily) when soil moisture exceeds s_{fc} (hence not included here), this term is retained in equation (7) to account for its possible role in slower soil moisture dynamics. In their interpretation of climate-model simulated soil moisture, *Delworth and Manabe* [1988] used a value of $s = 0.75 s_{\mathrm{fc}}$ for this shift from water-controlled dynamics to other environmental-controlled regime, which is equivalent to using $s^* = 0.75 s_{\mathrm{fc}}$ here. Equations (6) and (7) are stochastic ordinary differential equations with an intermittent and random component ($P(t)$) and a quasi-deterministic nonlinear loss term ($ET + D_{\mathrm{r}}$). On annual or longer timescales, these equations are often studied in a 'Budyko framework' that relates the actual ET to an aridity index (ratio of atmospheric evaporation demand to available water, PET/P) [e.g. *Li et al.*, 2013].

2.2. Memory and Spectra

The storage term ($\mathrm{d}s/\mathrm{d}t$) in equations (6) and (7) is known to introduce a statistical 'memory' into soil moisture, which in turn influences regional atmospheric processes on daily-to-seasonal timescales. This memory effect is manifested in the slow decay of the corresponding auto-correlation function, and is often determined from the integral timescale (τ) of the auto-correlation function [*Priestley*, 1981]

$$\rho_s(t, \alpha) = \frac{\overline{s'(t)s'(t+\alpha)}}{\sigma_s^2}, \tag{8}$$

where the subscript s denotes soil moisture as a state variable, primes indicate fluctuations around the mean value, α is the time lag, and σ_s^2 is the soil moisture variance. Direct approaches to estimate this function have been studied by *Koster and Suarez* [2001] and *Seneviratne et al.* [2006], mostly at monthly time lags and at global scales. Local to regional spatial scales and higher frequency (subdaily to interannual) dynamics are the focus of this work. Note that ρ_s is expressed in equation (8) as a function of t and time lag (α) to signify the non-stationarity of the soil moisture time series.

When assuming stationarity (i.e. $\rho_s(t, \alpha) = \rho_s(\alpha)$), the integral timescale (memory) of soil moisture can be defined in multiple ways. These definitions include the first time lag (α) at which $\rho_s(\alpha)$ crosses zero, the lag at which it drops to $1/e \approx 0.37$ (e-folding) of its initial value ($= 1$) at zero lag (assumes an exponential decay of ρ_s), or most commonly as the area under ρ_s [*Priestley*, 1981]

$$\tau = \int_0^{+\infty} \rho_s(\alpha)\mathrm{d}\alpha, \tag{9}$$

where ρ_s is assumed to decay to zero and remain negligible as $\alpha \to \infty$. The lack of stationarity in the soil moisture time series and the sensitivity of memory estimation to different treatments (such as removing periodicity) are discussed in detail in supplementary material S1. In essence, the analysis in S1 reveals that the auto-correlation function of soil moisture is affected by de-trending (removing monthly, seasonal, or annual means) the time series, and while this leads to shorter memory timescale estimates, it also yields losses in the variance of the process.

The normalized temporal spectrum of soil moisture $E_{\mathrm{ns}}(f)$, where f is frequency (in cycles per unit time), is the Fourier transform of $\rho_s(\alpha)$ (Wiener-Khinchin theorem)

$$E_{\mathrm{ns}}(f) = 2\int_{-\infty}^{+\infty} \rho_s(\alpha)e^{-i2\pi f\alpha}\mathrm{d}\alpha, \tag{10}$$

and thus

$$E_{\mathrm{ns}}(0) = 4\int_0^{+\infty} \rho_s(\alpha)\mathrm{d}\alpha = 4\tau, \tag{11}$$

since $\rho_s(\alpha)$ is a real and even function. Therefore, estimating soil moisture memory as $\tau = E_{\rm ns}(0)/4$ for a measured or modeled finite time series requires ad-hoc extrapolations of the spectral behavior of $s(t)$ as $f \to 0$. The above concepts were pioneered by *Delworth and Manabe* [1988] who hypothesized that this memory stems from evapotranspiration by studying equation (1) as a first-order Markov process, where $s(t)$ is governed by a white-noise spectrum of rainfall and a linear dependence of $\chi[s(t)]$ on s. This simplified model results in the well-known Lorentzian stationary soil moisture spectrum (red-noise), where $E_{\rm ns}(f) \sim ((2\pi f)^2 + \beta^2)^{-1}$, and $\beta = PET/(\eta Z_{\rm r})$ (see *Halley* [1996] for a review on $1/f$ noises in ecological contexts). While this f^{-2} scaling received some support from long-term measurements [*Vinnikov et al.*, 1996; *Wu et al.*, 2002] and climate model runs *Delworth and Manabe* [1988], recent theoretical efforts and models with varying complexity have shown that the temporal spectrum of soil moisture deviates from its Lorentzian form (decays faster than f^{-2}) at high frequencies, resembling black- instead of red-noise [e.g. *Katul et al.*, 2007; *Nakai et al.*, 2014]. These results were attributed to the fact that the rainfall spectrum exhibits a power-law decay ($f^{-0.5}$ to f^{-1}) at the storm scales [*Fraedrich and Larnder*, 1993; *Molini et al.*, 2009]. Deviations from Lorentzian spectra were also reported when including a nonlinear dependence of the drainage term on soil moisture and/or including net radiation variability at lower frequencies [*Nakai et al.*, 2014].

2.3. Persistence and Clustering

Formally, persistence in a stochastic field $\phi(x,t)$ fluctuating around its ensemble average (indicated by brackets) $\langle \phi(x,t) \rangle$ according to some prescribed dynamics and at a fixed point x is defined as the probability that the quantity $\mathrm{sgn}[\phi(x,t) - \langle \phi(x,t) \rangle]$ does not change up to time t [*Majumdar*, 1999; *Perlekar et al.*, 2011]. Henceforth, the field $\phi(x,t)$

represents the effective soil moisture $s(x, t)$ considered at a fixed location or averaged over nearby locations (see section 3), and the ensemble average can be replaced by any other relevant threshold, such as s_{w}, s^* or s_{fc} in this context of soil moisture dynamics. The alternations between long quiescent dry phases (such as $s < s^*$) and wet excursions in soil moisture are clearly forced by nonlinear interactions with $P(t)$ and $\chi[s(t), t]$. The two phenomenological components of these 'switches' are the amplitudes of excursions above or below the threshold and the local frequency of oscillations around it. The former is related to the strength of an imposed forcing (e.g. (non)occurrence of rainfall) and the latter is defined as the tendency of events to 'cluster' together. The separation of the amplitude variability from oscillatory behavior for the time-dependent variable $s(t)$ can be achieved using the telegraphic approximation ($TA[s(t)]$) [*Sreenivasan and Bershadskii*, 2006]

$$TA[s(t)] = \frac{1}{2}\left(\frac{s(t) - s^*}{|s(t) - s^*|} + 1\right), \tag{12}$$

where $TA(s)$ is binary (value of 0 or 1) depending on whether $s(t)$ at any time exceeds s^* (wet state $TA = 1$) or resides below it (dry state $TA = 0$). Figure 1 provides an example of the telegraph approximation of a soil moisture time series. Within this framework, the TA masks amplitude variations but retains the on-off and off-on switches in the time series. Time correlations between these switches, if any, and the distribution of inter-pulses between them define persistence. On the other hand, 'memory' in a hydrological context is the time needed for the system to dissipate/recover from wet/dry states.

Because such switches need not be entirely random in time, the connection between the spectral exponent of the full series $s(t)$ (controlled by both amplitude variation and clustering) and that of its TA (controlled by clustering only) reveals the magnitude of fractional variance explained by amplitude fluctuations and more importantly, the timescales at which they contribute to this variance. If both spectra exhibit power-law decays, with $E_{TA}(f) \sim f^{-m}$ and $E_{\mathrm{ns}}(f) \sim f^{-n}$, then an empirical relation between m and n, studied in turbulence research [*Bershadskii et al.*, 2004; *Sreenivasan and Bershadskii*, 2006; *Cava and Katul*, 2009], and proved analytically for a range of stochastic processes, is given by

$$m = \frac{n+1}{2}, \tag{13}$$

such that a Markov-Lorentzian spectrum of soil moisture ($n = 2$) will result in $m = 1.5$ for the TA, and therefore the on-off switches will have a larger spectral content. For $n > 1$ (which is the case for soil moisture), the TA spectrum exhibits slower decay (more randomness with $m < n$) than that of the full series of soil moisture. The usefulness of equation (13) lies in the fact that for a wide range of stochastic processes, analytical tractability of TA dynamics and the exponent m may be less challenging than that of the full dynamics.

Another dimension to persistence in non-equilibrium dynamical systems used here is clustering (density of crossings) and its scaling behavior. Let $\psi(T)$ be the running density of crossing the threshold s^* in a time interval T, i.e. $\psi(T) = N(T)/T$, where $N(T)$ is the number of crossings (upward or downward) of s^* in the interval T, and let its fluctuations be $\delta\psi(T) = \psi(T) - \langle\psi(T)\rangle$ (where the brackets indicate averaging over a long period), then the quantity $\langle\delta\psi(T)^2\rangle^{1/2}$ represents the local standard deviation of the series $\psi(T)$, and is assumed to decay as

$$\langle\delta\psi(T)^2\rangle^{1/2} \sim T^{-\omega}, \tag{14}$$

where ω is known as the cluster exponent and is a measure of the tendency of crossing events to cluster together. In the case of rainfall clustering, the rain- no rain (distribution of dry periods) is used instead of threshold-crossing. As a reference, we note that white noise presumably has no clustering properties with $\omega = 0.5$, while a smaller ω (< 0.5) indicates an increased clustering tendency with respect to white noise.

The concepts discussed above are first explored for the first-order Markov process that remains widely used as an idealized model for soil moisture dynamics in climate systems and was introduced by *Delworth and Manabe* [1988] when analyzing GCM outputs. This process is represented by

$$\frac{\mathrm{d}y}{\mathrm{d}t} + \lambda y = F(t), \tag{15}$$

where $y(t)$ is a stochastic process (analogous to dimensionless effective soil moisture $s(t)$), $F(t)$ (T^{-1}) is assumed to be a white noise process (analogous to $P(t)/(\eta Z_\mathrm{r})$), and $\lambda = (PET/(\eta Z_\mathrm{r}))$ (analogous to $1/\tau$ and independent of y) is a decay constant that represents a linear dependence of the loss function on y. In this framework, the timescale $1/\lambda$ is also the e-folding time lag (memory) of the exponentially decaying auto-correlation function of the process $y(t)$ in the absence of forcing. Equation (15) is analogous to equation (6) (water-limited regime) when assuming that rainfall exhibits a white-noise spectrum. The first-order Markov process is used here as a guiding model for the behavior of persistence and memory scales for the process $s(t)$, and hence two scenarios are considered. The first examines equation (15) under a white-noise rainfall forcing (similar to *Delworth and Manabe* [1988]), and the second uses a measured rainfall time series as the forcing on the r.h.s of equation (15). In supplementary material S1, the effect of using a constant or periodic PET (and hence τ) with each of the two forcing scenarios is also investigated.

This supplementary discussion (S1) concluded that both the 'redness' of the spectrum of $s(t)$ (white-noise rainfall) and the deviations from this 'redness' (measured rainfall) is not affected by this change of decay timescales. This same result was also pointed out by *Delworth and Manabe* [1988].

3. Data and Methods

The concepts of persistence, clustering and memory timescales presented above are explored for several datasets of high-frequency (half-hourly) root-zone soil moisture measurements collected at Mae Moh forest (Teak plantation in Thailand, Mar 2006 – Feb 2012) [*Yoshifuji et al.*, 2006, 2014], Duke forest (both a Loblolly pine plantation (PP) and a second-growth oak-hickory hardwood (HW) forest near Durham, NC, USA, Jan 2001 – Dec 2006) [*Katul et al.*, 2007; *Oishi et al.*, 2013], and the Seto forest (second-growth deciduous forest in Japan, Jan 2005 – Dec 2009) [*Matsumoto et al.*, 2008]. Additionally, eddy-covariance measurements of ET and other meteorological variables at 30-min timescales are available at all the sites. Table 1 summarizes the soil, canopy, and climate characteristics at each site. The long-term mean annual temperature and rainfall are 15.5°C and 1100 mm at Duke forest, 15.1°C and 1615 mm at Seto forest, and 25.8°C and 1284 mm at Mae Moh forest (2000-2004 only). Volumetric soil water content (m^3 m^{-3}) was measured at several depths covering the root-zone at each site, and at several spatially-extended locations (only at Duke forest) using time domain reflectometry (TDR) sensors (CS-615, Campbell Scientific, Logan, UT) at Duke PP and Mae Moh sites, vertically oriented frequency domain sensors (ThetaProbe ML2x, Delta-T Devices, Cambridge, UK) at Duke HW site, and TRIME-FM2/P2 (TDR with intelligent MicroElements, IMKO, Germany) at Seto forest. The measurement depths beneath the surface were 0.1, 0.2, 0.4, and 0.6 m at Mae Moh site (1 location), and 0.02, 0.05, 0.1, 0.2, and 0.5 m at Seto site (1 location). Vertically-arrayed rods of 30-cm length integrating soil moisture across the root-zone were deployed at Duke PP (24 locations) and Duke HW (6 locations). These measurements were averaged both vertically and spatially (when applicable) resulting in one multi-year 30-min soil moisture time series at each site.

The Mae Moe forest is situated in the subtropical region subject to a tropical monsoonal climate, while the Duke and Seto forests are in the mid-latitude zone characterized by a warm-temperate climate [*Nakai et al.*, 2014]. These datasets offer a unique opportunity to examine the individual impact of vegetation and soil type as well as rainfall regimes on persistence and memory timescales in soil moisture dynamics. The co-location of the pine and hardwood stands at Duke forest, which have comparable rooting zone depth restriction (formed by a hard clay pan due to prior agricultural practices at the site), and are subjected to the same climatic forcing and soil texture, allows an evaluation of how differences in vegetation cover may affect persistence and memory.

Figure 2 shows the time series of the effective soil moisture and rainfall measurements of the four datasets. Seasonality in rainfall is mostly evident at Mae Moh forest and less pronounced at the other sites, where it is distributed almost evenly around the year. The memory timescale τ for each soil moisture series computed from the empirical auto-correlation function is 47.5, 44.6, 38.8 and 24.4 days for Duke-HW, Duke-PP, Mae Moh, and Seto forests respectively.

4. Results and Discussion

To address the study objectives, the probability distributions of soil moisture and rainfall at each site are first described to further illustrate the effects of seasonality across the datasets. The analysis demonstrates that the soil moisture states primarily reside away from the mean value and exhibit bi-modality associated with seasonality. The plant water-stress level s^* (described above) at each site is chosen as the threshold when employing the telegraphic approximation needed for persistence and clustering analysis. The physical basis of soil moisture memory within the root-zone followed by a dynamical interpretation of s^* as the threshold for the computations of TA are presented. The spectral scaling and distribution of persistence times and their relation to soil moisture memory are then determined and discussed.

4.1. Soil Moisture and Rainfall Distributions

On annual scales with seasonal signatures, the soil moisture PDF is typically bi-modal and dependent on whether rainfall and temperature/radiation are in phase, i.e. whether the wet season coincides with the growing season [*Miller et al.*, 2007; *Viola et al.*, 2008; *Feng et al.*, 2012, 2014]. Here, a qualitative discussion on such distributions is presented to illustrate site differences in terms of seasonality and rainfall depth characteristics. Figure 3 shows the PDF of effective soil moisture and the probability of exceedance of rainfall depth (above 1 mm) at the four sites at half-hourly timescale. There is a seasonal signature characterized by bi-modality at all sites (especially at Mae Moh forest) except the Seto forest, with a tendency for prevalence of wet states at Duke forest due to the evenly distributed rainfall around the year. Note that this distribution of soil moisture is also controlled by the loss function ($\chi[s(t), t]$) through a 'regime shift' type of dependence, with a linear relation between ET and s in the water-limited case and a PET otherwise. This loss function, within any regime, is primarily responsible for the mode in the distribution at low soil moisture levels, while the wet season dominates the generation of the other mode. Examination of the soil moisture PDF in Figure 3 also reveals that the mean of the distribution falls between these modes of wet and dry states, which is indicative of the prevalence of transient dynamics, where the system resides away from the mean for most of the time. The rainfall distributions (right panel of Figure 3) are comparable for the four datasets with extreme events associated with the strongly seasonal Asian monsoons more likely at the Mae Moh forest site.

4.2. Physiological Water Stress and Dynamical Equilibria

The difficulty in studying the dynamics of equations (6) and (7) emanate from the regime shifts in the dependence of the loss function (ET and D_r) on the variable s, the explicit dependence of most variables on time (t), and the intermittent and random behavior of rainfall. Nonetheless, a discussion of such dynamics is included here to examine how dynamical equilibria and their transient times compare to the water-stress level s^* and the memory scale. For steady state conditions ($\mathrm{d}s/\mathrm{d}t = 0$), equation (6) reduces to

$$\frac{P(t)}{\eta Z_\mathrm{r}} - \frac{PET}{\eta Z_\mathrm{r}}\left(\frac{s_\mathrm{o} - s_\mathrm{w}}{s^* - s_\mathrm{w}}\right) = 0, \quad (16)$$

where s_o represents an equilibrium state of the system, and the quantity ηZ_r is retained for dimensional consistency. Equation (16), with an initial condition $s_\mathrm{w} \leq s_\mathrm{i} \leq s^*$, results in an equilibrium soil moisture level given by

$$s_\mathrm{o} = (s^* - s_\mathrm{w})\frac{P}{PET} + s_\mathrm{w}, \quad (17)$$

and a linear stability analysis around this fixed point reveals that it is always stable (slope$= -PET/[\eta Z_\mathrm{r}(s^* - s_\mathrm{w})]$ from equation (16)). In the absence of forcing, where no rainfall occurs after $t = 0$, the stable fixed point is $s_\mathrm{o} = s_\mathrm{w}$ and by integrating equation

(6), the system approaches the wilting point s_w exponentially in time as

$$s(t) = (s_\mathrm{i} - s_\mathrm{w}) \exp\left(\frac{-t}{\tau}\right) + s_\mathrm{w}, \tag{18}$$

where the memory timescale $\tau = \eta Z_\mathrm{r}(s^* - s_\mathrm{w})/PET$ has been used. The latter timescale is a result of the formulations in the previous discussion (equation (11) in subsection 2.2), where $\tau = E_\mathrm{ns}(0)/4$ and $E_\mathrm{ns}(f) \sim ((2\pi f)^2 + \beta^2)^{-1}$, with $\beta = PET/(\eta Z_\mathrm{r})$, resulting in $\tau = 1/\beta = \eta Z_\mathrm{r}/PET$ (see *Nakai et al.* [2014] for more details). The only difference here is the factor $(s^* - s_\mathrm{w})$ that emphasizes the dynamics within the water-limited regime. The time needed to reach some value s starting from s_i can be determined as

$$t_\mathrm{w} = \tau \ln\left(\frac{s_\mathrm{i} - s_\mathrm{w}}{s - s_\mathrm{w}}\right), \tag{19}$$

such that as $s \to s_\mathrm{w}$, $t_\mathrm{w} \to \infty$ (the system approaches s_w asymptotically), and therefore the memory timescale τ is only a fraction of t_w. In fact, noting that the quantity $(s_\mathrm{i} - s_\mathrm{w})/(s - s_\mathrm{w}) \geq 1$, and from equations (18) and (19), τ represents the time needed to reach the e-folding of the initial departure $(s_\mathrm{i} - s_\mathrm{w})$ from equilibrium. This is the essence of the Markovian process in the absence of forcing, where it can be shown that the e-folding time τ in equation (18) and that of the corresponding exponentially decaying auto-correlation function $\rho_s(\alpha) = \exp(-\alpha/\tau)$ are identical. It is emphasized that a crossing lifetime (inter-pulse) of a threshold, or approaching a fixed point such as s_w, is typically longer than the aforementioned memory. The ratio of rainfall to the loss function in this regime P/PET controls the position of the stable fixed point on the linear ET-s dependence line. This equilibrium approaches the water-stress level ($s_\mathrm{o} = s^*$) when $P/PET \approx 1$, but the intermittent nature of rainfall prohibits further analytical tractability. When $P/PET > 1$, the system exits the linear dependence regime and equation (7) describes the dynamics. A similar analysis for this equation with an initial condition $s^* \leq s_{\mathrm{i}} \leq s_{\mathrm{fc}}$ results in the equilibrium

$$s_{\mathrm{o}} = (s_{\mathrm{fc}} - s^*) \exp\left(\frac{1}{c} \ln\left(\frac{P - PET}{K_s}\right)\right) + s^*, \tag{20}$$

and hence the ratio $(P - PET)/K_s$ controls the stable fixed point. As this ratio approaches unity (i.e. difference between rainfall and PET is comparable to saturated hydraulic conductivity), the fixed point approaches field capacity, $s_{\mathrm{o}} = s_{\mathrm{fc}}$. Note that the fast dynamics above s_{fc} were ignored, i.e. if the ratio $(P - PET)/K_s$ exceeds one, the decay to $s(t) = s_{\mathrm{fc}}$ is instantaneous. When $P - PET$ is very small compared to K_s, s_{o} approaches the water-stress level s^*. These fixed points (s^* and s_{fc}) are again approached asymptotically.

The above discussion provides a dynamical perspective on the role of the characteristic values s_{w}, s^*, and s_{fc} in soil moisture dynamics and memory. Here, the threshold s^* is estimated from the four datasets using a hydrological and a dynamical context (Figure 4). Daily averages (48 measurement records sampled at 30-min intervals) of all the variables are used in Figure 4. A hyperbolic tangent function of the form $ET/PET = a \tanh(s)$ is also used as a model for normalized evapotranspiration ET/PET, where PET is calculated using the Penman-Monteith equation from the corresponding micro-meteorological measurements and ET is determined from the available eddy-covariance measurements. While the Seto forest data shows small variance and negligible dependence of ET on s, the other sites exhibit comparable water-stress threshold, with s^* being 0.62, 0.6, 0.54, and 0.3 for Mae Moh, Duke-PP, Duke-HW, and Seto forests respectively. The right panel of Figure 4 shows soil moisture dynamics in the form ds/dt = f(s) (resembling a vector field representation), where ds/dt = Δs/Δt is the discretized time rate of change in s (differences in daily averages). The relatively large positive values (ds/dt > 0) are associated with rainfall events and their negative counterpart (ds/dt < 0) are due to drainage losses. These events are 'quasi-instantaneous' on the daily timescales. The small negative and positive fluctuations are attributed to ET losses (when ds/dt < 0) and weak rainfall events or otherwise moisture redistribution from below the root-zone (when ds/dt > 0). The features in Figure 4 are common to all datasets, where there appears to be an approximate balance between the rainfall input and the loss function. Recall that ds/dt $= 0 =$ f(s) represents the dynamical equilibrium, and in cases where P balances PET, this equilibrium approaches s^*. The latter is evident in the right panels of Figure 4, where the stable fixed point is close to the water-stress level approximated in the left panel ($ET/PET = a \tanh(s)$). The function $f(s)$ in the vector field is fit to a cubic function, $O(s^3)$, to capture the likely non-linearity in the dynamics that accommodates rainfall and drainage, but we emphasize that higher-order functions in s result in essentially the same stable fixed point.

4.3. Persistence and Clustering

Figure 5 shows the spectrum of the simulated process $s(t)$ and its corresponding TA in the Markovian framework (equation (15)) using two types of forcing $F(t)$, a white noise process (left panel) and the measured rainfall time series (normalized by ηZ_{r}) at one of the sites (right panel), selected here as the Duke forest site (section 3) only for illustration. Using measurements at the other sites did not result in any significant changes in the outcome of the analysis. Since there is no s^* defined for this idealized model, the TA here is calculated using equation (12) around the mean of $s(t)$. The decay constant τ is estimated from the measurements as the average of $\eta Z_{\mathrm{r}}/PET$ during the growing season and over five years (length of Duke forest measurements)(see also supplementary material S1). As discussed earlier, the f^{-2} and $f^{-2.7}$ scaling of the normalized spectrum of $s(t)$ at frequencies higher than $1/\tau$ are clear when forced by a white noise and measured rainfall spectra respectively. The decay timescale τ (also referred to as separation timescale) corresponds to the frequency that defines the transition of the soil moisture spectrum from a white noise type at low frequencies to a red (or black) noise type at higher frequencies. The relation between the spectral exponents of $s(t)$ and its TA given in equation (13) holds reasonably for the two types of rainfall forcing, which suggests some robustness to the particulars of the forcing variable. Figure 5 also shows the PDF of I_s, the inter-pulses below the mean of $s(t)$, normalized by the memory timescale τ. This PDF shows that persistence timescales can exceed memory ($I_s/\tau > 1$). Note that these persistence times are largely controlled by $P(t)$ (which initiates an up-crossing), while positive excursions above the mean of $s(t)$ depend on the interplay of $P(t)$ and τ.

The binary time series (TA) around the threshold s^* for each dataset is shown in Figure 6. The highest density of crossings is evident at the Seto forest, indicating shorter persistence times above or below the physiological threshold. Note that for this site, the threshold was estimated from the vector field analysis in Figure 4 rather than the water-stress level. On the other hand, longer persistence times are evident at the other sites, with Mae Moh forest, and due to seasonality in rainfall, exhibiting prolonged wet/dry states. The TA at Duke forest has a more pronounced seasonal trend, where persistence times are shortest (higher frequency of crossings) during the growing season. The temporal correlation between these crossing events at each site is shown in the bottom panel of Figure 6, through the normalized spectrum of these TA series ($E_{TA}(f) \sim f^{-m}$)

along with that of the full series ($s(t)$) spectrum ($E_{\mathrm{ns}}(f) \sim f^{-n}$). The latter exhibits a power-law decay steeper than the Lorentzian f^{-2} scaling at high frequencies (daily and sub-daily) owing to the correlated structure of rainfall at these short timescales, whereas the TA spectra have larger variances at all sites. The comparison between the two spectra at each site reveals that amplitude fluctuations in soil moisture, which are absent from $E_{TA}(f)$, are responsible for the imposition of steeper deterministic decay in $E_{\mathrm{ns}}(f)$, particularly at high frequencies, hence resulting in larger memory in $s(t)$ relative to its TA counterpart. At longer timescales, $E_{TA}(f)$ captures the low frequency fractional variance in $E_{\mathrm{ns}}(f)$, where at scales comparable to or longer than the soil moisture memory (solid vertical line in the bottom panel of Figure 6), the bulk of the variance stems from the crossing dynamics (persistence scales). In other words, the memory timescale is dictated by deterministic processes (such as $\tau = \eta Z_{\mathrm{r}}/PET$), while persistence scales are dominated by long-term 'de-correlated' forcing such as rainfall. The relation between the spectral exponents m and n also holds reasonably for the datasets featured here, with a deviation of ± 0.1 at most. The latter result, while empirical, was shown to be true for velocity and temperature statistics in turbulent flows and at different Reynolds numbers [*Sreenivasan and Bershadskii*, 2006].

Figure 7 shows the distribution of persistence times for both soil moisture (I_s) and rainfall (I_P) at each site. Note that I_P is the inter-arrival times between rainfall events and both I_s and I_P are normalized by the corresponding soil moisture memory τ. These distributions are fit to a stretched exponential (a multiplicative PDF of power law and exponential decay) of the form [*Laherrere and Sornette*, 1998]

$$\mathrm{PDF}(x) \sim x^{b-1} \exp\left(-x^{b}\right), \tag{21}$$

where $b < 1$ and x can represent I_s/τ or I_P/τ. The borderline case $b = 1$ recovers the exponential distribution. The PDF's in Figure 7 show that there is a tendency of persistence times of soil moisture below the threshold s^* to exceed the memory timescale at all sites, albeit as extreme events emphasized by the tails of the distributions. The stretched exponential functions fit to the data reflect a power-law behavior at short persistence times and an exponential decay at long times. These exponentially decaying long dry periods prevail for around two to four times the memory scale, and are indicative of the fact that anomaly dissipation (quantified by τ) does not necessitate a switching (transition from dry to wet states or vice versa). Another important aspect of the distributions shown in Figure 7 (for soil moisture) is that they exhibit negligible sensitivity to the magnitude of τ (note that the Duke forest sites have much longer memory). On the contrary, and except for Mae Moh site, the inter-arrival times between rainfall events rarely exceed the corresponding soil moisture memory, i.e. dry atmospheric anomalies are unlikely to persist longer than the 'de-correlation' time in soil moisture statistics (τ). The latter may be regarded as a necessary but not sufficient condition for causality between soil moisture and convective rainfall, or otherwise that atmospheric states are 'feeding off' on this memory. This is especially the case at Duke forest sites, where the rainfall persistence timescale at which the power-law ceases to exist is around 0.1τ, and therefore longer dry atmospheric anomalies decay exponentially fast before reaching τ. The analogous regime shift (power-law to exponentials) for soil moisture appears to be indifferent to the variability in memory across all sites (around 0.3τ). Those events within the exponential part of the distributions, for both soil moisture and rainfall, are likely to be statistically independent (memory-less property of exponential distributions). Hence, τ, being towards the tail of this part, is likely to be an overestimate of the 'de-correlation' time in soil moisture.

The clustering properties of both soil moisture and rainfall at each site are shown in Figure 8. The quantity $\langle \delta\psi(T)^2 \rangle^{1/2}$ indicates similar decay for all the datasets with a higher tendency for clustering at Duke forest. At all sites, the cluster exponent w ranges between 0.36 to 0.42 for soil moisture and 0.24 to 0.34 for rainfall. *Molini et al.* [2009] found similar clustering properties for rainfall occurrences at different sites, while *Sreenivasan and Bershadskii* [2006] found remarkably close cluster exponents for velocity signals in turbulent flows as those of soil moisture here. The differences between the cluster exponents of rainfall and soil moisture at each site, with the former exhibiting higher tendency of clustering of rainfall occurrence, show that rainfall persistence (or lack thereof) does not translate directly to soil moisture. In other words, rainfall occurrence alone cannot explain soil moisture switching events between wet and dry states, which suggests the significance of rainfall depth (storm strength and duration) relative to the storage capacity of the active soil layer on these persistence times. This tendency for clustering ceases to exist at all sites beyond seasonal scales (around 100 days), where at longer time intervals the cluster exponent approaches unity as a limiting value for both soil moisture and rainfall. This unity limit is indicative of statistical independence of rainfall occurrences and soil moisture crossing events.

5. Future Directions

Much of the memory-persistence results reported here remain diagnostic, not prognostic. The lack of a theoretical or concrete measure of persistence in soil moisture currently limits its direct use in land-atmosphere models instead of memory, especially that persistence represents a distribution of times (rather than a single timescale) between threshold crossings and involves clustering of these crossing events. Nevertheless, connections between memory and persistence is an on-going research topic in complex system sciences, where there exists a relation between the distribution of these persistence times and the corresponding spectrum of the threshold-crossings (spectrum of telegraph approximation (TA)) [*Jensen*, 1998; *Bershadskii et al.*, 2004]. Such relations have been derived for a restricted class of systems. For example, when invoking certain analogies with systems exhibiting or approaching a state of self-organized criticality (SOC), connections between TA spectral exponents (linked to the full spectrum of soil moisture content as evidenced by the analysis here) and the inter-pulse PDF can be made. While the latter concept applies in the context of spatially-extended dissipative dynamical systems *Bak et al.* [2004], efforts to generalize its characteristics have been made by *Jensen* [1998] and *Majumdar* [1999]. Examples of such systems are the classical sand pile model, turbulence and convection, river flow, electric currents through resistors, and many others [*Bak et al.*, 2004]. These systems evolve toward a self-similar (fractal) critical state with no intrinsic time or length scale. Whether soil moisture as a stochastically-forced process exhibits self-organized criticality is not fully known, but the system can be regarded as dissipative in the absence of rainfall. Since this topic is certainly interesting for future investigation, only a preliminary assessment of connecting persistence and memory within this SOC framework to the soil moisture datasets used here is provided. Let α and β be the exponents of the power-law decay of the PDF of persistence times, and that of the spectrum of the TA respectively, then

$$\beta = 3 - \alpha, \tag{22}$$

is a well-known relation for SOC systems. Note that β here is about 1.57-1.67 (see TA spectra in Figure 6) and α ranged between 0.5-0.8 (see power-law fits in Figure 7). Intermittency corrections to equation (22) are also studied in the context of turbulence and convection, where

$$\beta = 3 - \alpha - \mu, \tag{23}$$

and μ represents such corrections. The analysis here and equation (23) show that the intermittency explanation μ is of order 0.8. In analogy with intermittency in the turbulence convection problem studied by *Bershadskii et al.* [2004], where they addressed hot/cold plumes (temperature fluctuations), which are here equivalent to wet/dry states (soil moisutre fluctuations), this exponent μ is calculated from the intermittency in soil moisture fluctuations as

$$\chi = \|\frac{\mathrm{d}s^2}{\mathrm{d}t}\|, \tag{24}$$

where s here is soil moisture fluctuations around the threshold s^*. The local average in a time window T is

$$\chi_T = \frac{1}{T}\int_t^{t+T} \chi(t)\mathrm{d}t, \tag{25}$$

such that for several time windows T (e.g. 0.1, 0.5, 1, 5, 10, 20, 50, 100, ... days), the scaling

$$\frac{\langle \chi_T^2 \rangle}{\langle \chi_T \rangle^2} = T^{-\mu}, \tag{26}$$

describes such intermittency effects and μ is the intermittency exponent. Figure 9 shows the intermittency calculations for the soil moisture time series at each site. The exponent μ is also shown to be of order 0.8 as predicted by equation (23). While *Bershadskii et al.*

[2004] had a factor of two difference ($\beta = 3 - \alpha - \mu/2$) in their paper, if this correction can be verified, then it is possible to link the spectrum of soil moisture to its TA counterpart, and use such SOC analogy to infer the PDF of persistence timescales. This framework offers an *ad hoc* result for connecting memory (from spectra) and persistence (through SOC + intermittency). However, whether such analogies can be applied in the context of soil moisture dynamics, i.e. whether soil moisture exhibits features of an SOC system is a topic for a future examination.

6. Conclusions

This work addressed the different underlying mechanisms and relative importance of the concepts of memory and persistence timescales in root-zone soil moisture dynamics. While memory is a well-studied and a widely used timescale for soil moisture content in land-climate modeling, persistence times below or above some threshold (such as s^* used here) remain under-exploited. These persistence scales are more indicative of the wet and dry states of soil moisture, and are perhaps the principal measure of land-atmosphere coupling strength. In a comparative context with soil moisture memory, the characteristics of the distribution of such persistence times were explored for several high frequency soil moisture datasets collected in different biomes and climates. The clustering properties of the soil moisture time series (density of threshold-crossing per unit time) were also analyzed. The sites spanned tropical monsoon to warm-temperate climates, where rainfall was seasonal in the former and distributed almost evenly around the year in the latter. The threshold s^* (plant water-stress level) was estimated for each dataset by relating the water losses (mostly measured ET) to soil moisture using a sigmoid-like function, and independently from a data-based one-dimensional phase space reconstruction to infer the stable fixed point in the dynamics. The estimated threshold s^* was acceptably close when comparing these two methods, indicating that for these datasets, and due to the balance between the input (rainfall) and output (loss function), the system approaches s^* as a stable fixed point.

Despite the differences in the rainfall forcing and vegetation cover among the studied sites, the temporal correlations of threshold crossings were similar and followed a unique relation with the corresponding correlations in the measured soil moisture series (that includes amplitude fluctuations from the threshold). This relation is common in many stochastic models and has been shown to hold true for turbulence statistics. The distribution of the persistence times exhibited a stretched exponential behavior and reflected a likelihood of exceeding the memory timescale at all sites. However, the rainfall counterpart of these distributions showed that at sites with longer soil moisture memory, dry atmospheric anomalies become less likely. The cluster exponent revealed that the clustering tendency in rainfall events (on-off switches) does not translate directly to clustering in soil moisture. This is particularly the case in climates where rainfall and evapotranspiration are out of phase, resulting in less ordered (more independent) persistence in soil moisture than in rainfall.

Acknowledgments. The authors thank Amilcare Porporato for the helpful comments and suggestions that motivated the discussion and supplementary material. Ghannam thanks Natsuko Yoshifuji for providing the soil moisture and meteorological data at Mae Moh forest. Katul and Ghannam acknowledge the National Science Foundation (NSF-CBET-103347, and NSF-EAR-1344703), and the U.S. Department of Energy (DOE) through the office of Biological and Environmental Research (BER) Terrestrial Ecosystem Science (TES) Program (DE-SC0006967 and DE-SC0011461)). Paschalis acknowledges the financial support of the Swiss National Sciences Foundation (grant No P2EZP2-152244) and the Stavros Niarchos Foundation, through the SNSF Early Postdoc Mobility Fellowship. Nakai, Igarashi, and Kumagai acknowledge the framework of the "Precise Impact Assessments on Climate Change" of the Program for Risk Information on Climate Change (SOUSEI Program) supported by the Ministry of Education, Culture, Sports, Science, and Technology-Japan (MEXT). Additional information about the datasets used here can be obtained from the corresponding author at khaled.ghannam@duke.edu.

References

Alfieri, L., P. Claps, P. D'Odorico, F. Laio, and T. M. Over (2008), An analysis of the soil moisture feedback on convective and stratiform rainfall, *J. Hydrometeorol.*, *9*(2), 280–291.

Bak, P., C. Tang, and K. Wiesenfeld (1988), Self-organized criticality, *Phy. Rev. A.*, *38*(1), 364–374.

Bershadskii, A., J. Niemela, A. Praskovsky, and K. Sreenivasan (2004), Clusterization and intermittency of temperature fluctuations in turbulent convection, *Phy. Rev. E.*, *69*(5), 056,314.

Borgogno, F., P. D'Odorico, F. Laio, and L. Ridolfi (2010), A stochastic model for vegetation water stress, *Ecohydrology*, *3*(2), 177–188.

Bras, R.L., and I. Rodríguez-Iturbe (1985), Random functions and hydrology, 559 pp, *Addison-Wesley, Reading, MASS, USA.*.

Budyko, M. (1974), Climate and life, 508 pp, *Academic, San Diego, Calif*, pp. 72–191.

Budyko, M. I. (1961), The heat balance of the earth's surface, *Soviet Geography*, *2*(4), 3–13.

Cava, D., and G. Katul (2009), The effects of thermal stratification on clustering properties of canopy turbulence, *Boundary-Layer Meteorol.*, *130*(3), 307–325.

Chamecki, M. (2013), Persistence of velocity fluctuations in non-Gaussian turbulence within and above plant canopies, *Phys. Fluids.*, *25*(11), 115,110.

Clapp, R. B., and G. M. Hornberger (1978), Empirical equations for some soil hydraulic properties, *Water Resour. Res.*, *14*(4), 601–604.

Daly, E., A. C. Oishi, A. Porporato, and G. G. Katul (2008), A stochastic model for daily subsurface CO_2 concentration and related soil respiration, *Adv. Water Resour.*, *31*(7), 987–994.

Delworth, T., and S. Manabe (1988), The influence of potential evaporation on the variabilities of simulated soil wetness and climate, *J. Clim.*, *1*, 523–547.

D'Odorico, P., F. Laio, A. Porporato, and I. Rodriguez-Iturbe (2003), Hydrologic controls on soil carbon and nitrogen cycles. II. a case study, *Adv. Water Resour.*, *26*(1), 59–70.

Entekhabi, D., I. Rodriguez-Iturbe, and F. Castelli (1996), Mutual interaction of soil moisture state and atmospheric processes, *J. Hydrol.*, *184*, 3–17.

Feng, X., G. Vico, and A. Porporato (2012), On the effects of seasonality on soil water balance and plant growth, *Water Resour. Res.*, *48*(5).

Feng, X., A. Porporato, and I. Rodriguez-Iturbe (2014), Stochastic soil water balance under seasonal climates, *P ROY SOC LOND A MAT*, *471*(2174), doi: 10.1098/rspa.2014.0623.

Findell, K. L., and E. A. Eltahir (1997), An analysis of the soil moisture-rainfall feedback based on direct observations from Illinois, *Water Resour. Res.*, *33*(4), 725–735.

Fischer, E. M., S. I. Seneviratne, D. Lüthi, and C. Schär (2007), Contribution of land-atmosphere coupling to recent European summer heat waves, *Geophys. Res. Lett.*, *34*(6).

Fraedrich, K., and C. Larnder (1993), Scaling regimes of composite rainfall time series, *Tellus A*, *45*(4), 289–298.

Guan, K., S. E. Thompson, C. J. Harman, N. B. Basu, P. S. C. Rao, M. Sivapalan, A. I. Packman, and P. K. Kalita (2011), Spatiotemporal scaling of hydrological and agrochemical export dynamics in a tile-drained Midwestern watershed, *Water Resour. Res.*, *47*(10), doi:10.1029/2010WR009997.

Halley, J. M. (1996), Ecology, evolution, and $1/f$ noise, *Trends Ecol. Evol.*, *11*(1), 33–37.

Juang, J. Y., G. G. Katul, A. Porporato, P. C. Stoy, M. S. Siqueira, M. Detto, H. S. KIM, and R. Oren (2007), Eco-hydrological controls on summertime convective rainfall triggers, *Glob. Change Biol.*, *13*(4), 887–896.

Jensen, H. J. (1998), *Self-organized criticality: emergent complex behavior in physical and biological systems*, vol. 10, 129 pp., Cambridge university press.

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*(5), doi:10.1029/2006WR005356.

Koster, R. D., and M. J. Suarez (2001), Soil moisture memory in climate models, *J. Hydrometeorol.*, *2*, 558–570.

Laherrere, J., and D. Sornette (1998), Stretched exponential distributions in nature and economy:fat tails with characteristic scales, *Eur. Phys. J. B. Condensed Matter and*

Complex Systems, *2*(4), 525–539.

Laio, F., A. Porporato, L. Ridolfi, and I. Rodriguez-Iturbe (2001), Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress: II. probabilistic soil moisture dynamics, *Adv. Water Resour.*, *24*(7), 707–723.

Lauzon, N., F. Anctil, and J. Petrinovic (2004), Characterization of soil moisture conditions at temporal scales from a few days to annual, *Hydrol. Process.*, *18*(17), 3235–3254.

Li, D., M. Pan, Z. Cong, L. Zhang, and E. Wood (2013), Vegetation control on water and energy balance within the Budyko framework, *Water Resour. Res.*, *49*(2), 969–976.

Lorenz, E. B., R.and Jaeger, and S. I. Seneviratne (2010), Persistence of heat waves and its link to soil moisture memory, *Geophys. Res. Lett.*, *37*(9).

Majumdar, S. N. (1999), Persistence in nonequilibrium systems, *arXiv preprint cond-mat/9907407*.

Manzoni, S., and G. Katul (2014), Invariant soil water potential at zero microbial respiration explained by hydrological discontinuity in dry soils, *Geophys. Res. Lett.*, *41*(20), 7151–7158.

Matsumoto, K., T. Ohta, T. Nakai, T. Kuwada, K. Daikoku, S. Iida, H. Yabuki, A. V. Kononov, M. K. van der Molen, Y. Kodama, T. C. Maximov, A. J. Dolman, and S. Hattori (2008), Energy consumption and evapotranspiration at several boreal and temperate forests in the Far East, *Agr. Forest Meteorol.*, *148*(12), 1978–1989.

Miller, G. R., D. D. Baldocchi, B. E. Law, and T. Meyers (2007), An analysis of soil moisture dynamics using multi-year data from a network of micrometeorological observation sites, *Adv. Water Resour.*, *30*(5), 1065–1081.

Milly, P. C. D., R. T. Wetherald, K. A. Dunne, and T. L. Delworth (2002), Increasing risk of great floods in a changing climate, *Nature*, *415*, 514–517.

Molini, A., G. Katul, and A. Porporato (2009), Revisiting rainfall clustering and intermittency across different climatic regimes, *Water Resour. Res.*, *45*(11).

Monteith, J. L. (1965), *Evaporation and the environment, Symp. Soc. Exp. Bio*, vol. 19, 205–234 pp., Cambridge university press.

Montosi, E., S. Manzoni, A. Porporato, and A. Montanari (2012), An ecohydrological model of Malaria outbreaks, *Hydrol. Earth Syst. Sc.*, *16*, 2759–2769.

Nakai, T., G. G. Katul, A. Kotani, Y. Igarashi, T. Ohta, M. Suzuki, and T. Kumagai (2014), Radiative and rainfall controls on root zone soil moisture spectra, *Geophys. Res. Lett.*, *41*(21), 7546–7554.

Oishi, A. C., S. Palmroth, J. R. Butnor, K. H. Johnsen, and R. Oren (2013), Spatial and temporal variability of soil CO_2 efflux in three proximate temperate forest ecosystems, *Agr. Forest Meteorol.*, *171*, 256–269.

Parent, A. C., A. Francois, and L.-E. Parent (2006), Characterization of temporal variability in near-surface soil moisture at scales from 1-h to 2 weeks, *J. Hydrol.*, *325*, 56–66.

Parlange, M. B., G. G. Katul, R. H. Cuenca, M. L. Kavvas, D. R. Nielsen, and M. Mata (1992), Physical basis for a time series model of soil water content, *Water Resour. Res.*, *28*, 2437–2446, doi:10.1029/92WR01241.

Parolari, A. J., G. G. Katul, and A. Porporato (2014), An ecohydrological perspective on drought-induced forest mortality, *J. Geophys. Res.: Biogeosc.*, *119*(5), 965–981.

Paschalis, A., S. Fatichi, G.G. Katul and V.Y. Ivanov (2015), Cross-scale impact of climate temporal variability on ecosystem water and carbon fluxes, *J. Geophys. Res.: Biogeosc.*, *120*(9), 1716–1740.

Paschalis, A., P. Molnar, S. Fatichi, and P. Burlando (2013), A stochastic model for high-resolution space-time rainfall simulation, *Water Resour. Res.*, *49*(12), 8400–8417.

Paschalis, A., S. Fatichi, P. Molnar, S. Rimkus, and P. Burlando (2014a), On the effects of small scale space–time variability of rainfall on basin flood response, *J. Hydrol.*, *514*, 313–327.

Paschalis, A., P. Molnar, S. Fatichi, and P. Burlando (2014b), On temporal stochastic modeling of rainfall, nesting models across scales, *Adv. Water Resour.*, *63*, 152–166.

Perlekar, P., S. S. Ray, D. Mitra, and R. Pandit (2011), Persistence problem in two-dimensional fluid turbulence, *Phys. Rev. Lett.*, *106*(5), 054,501.

Porporato, A., E. Daly, and I. Rodriguez-Iturbe (2004), Soil water balance and ecosystem response to climate change, *Am. Nat.*, *164*(5), 625–632.

Porporato, A., G. Vico, and P. A. Fay (2006), Superstatistics of hydro-climatic fluctuations and interannual ecosystem productivity, *Geophys. Res. Lett.*, *33*(15).

Priestley, M. B. (1981), *Spectral Analysis and Time Series*, 877 pp., Elsevier, New York City, New York.

Rodríguez-Iturbe, I., and A. Porporato (2005), *Ecohydrology of water-controlled ecosystems: soil moisture and plant dynamics*, 416 pp., Cambridge University Press.

Rodriguez-Iturbe, I., A. Porporato, L. Ridolfi, V. Isham, and D. R. Coxi (1999), Probabilistic modeling of water balance at a point: the role of climate, soil and vegetation, *P. Roy. Soc. Lond. A. MAT*, *455*(1990), 3789–3805.

Seneviratne, S. I., R. D. Koster, Z. Guo, P. A. Dirmeyer, E. Kowalczyk, D. Lawrence, P. Liu, D. Mocko, C.-H. Lu, K. W. Oleson, et al. (2006), Soil moisture memory in AGCM simulations: Analysis of global land-atmosphere coupling experiment (GLACE) data, *J. Hydrometeorol.*, *7*(5), 1090–1112.

Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling (2010), Investigating soil moisture-climate interactions in a changing climate: A review, *Earth-Sci. Rev.*, *99*(3), 125–161.

Sreenivasan, K., and A. Bershadskii (2006), Clustering properties in turbulent signals, *J. Stat. Phys.*, *125*(5-6), 1141–1153.

Thompson, S. E., and G. G. Katul (2012), Multiple mechanisms generate Lorentzian and $1/f^a$ power spectra in daily stream-flow time series, *Adv. Water Resour.*, *37*, 94–103.

Vico, G., A. Porporato, (2013), Probabilistic description of crop development and irrigation water requirements with stochastic rainfall, *Water Resour. Res.*, *49*(3), 1466–1482.

Vinnikov, K. Y., A. Robock, N. A. Speranskaya, and C. A. Schlosser (1996), Scales of temporal and spatial variability of midlatitude soil moisture, *J. Geophys. Res.*, *101*(D3), 7163–7174.

Viola, F., E. Daly, G. Vico, M. Cannarozzo, and A. Porporato (2008), Transient soil-moisture dynamics and climate change in mediterranean ecosystems, *Water Resour. Res.*, *44*(11).

Wu, W., and R. E. Dickinson (2004), timescales of layered soil moisture memory in the context of land-atmosphere interaction, *J. Climate.*, *17*(14), 2752–2764.

Wu, W., M. A. Geller, and R. E. Dickinson (2002), The response of soil moisture to long-term variability of rainfall, *J. Hydrometeorol.*, *3*, 604–613.

Yoshifuji, N., T. Kumagai, K. Tanaka, N. Tanaka, H. Komatsu, M. Suzuki, and C. Tantasirin (2006), Inter-annual variation in growing season length of a tropical seasonal forest in northern Thailand, *Forest Ecol. Manag.*, *229*(13), 333 – 339.

Yoshifuji, N., Y. Igarashi, N. Tanaka, K. Tanaka, T. Sato, C. Tantasirin, and M. Suzuki (2014), Inter-annual variation in the response of leaf-out onset to soil moisture increase in a teak plantation in northern Thailand, *Int. J. Biometeorol.*, doi:10.1007/s00484-013-0784-2.

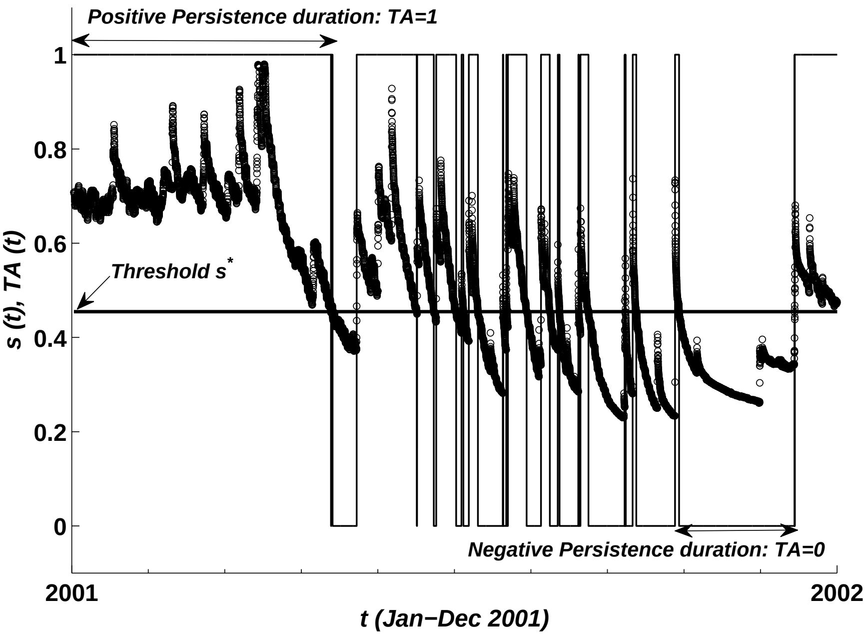

Figure 1. A one-year time series of measured effective soil moisture at the Duke-Hardwood site along with the corresponding telegraphic approximation (TA). The TA has a value of 1 when soil moisture is above the threshold s^* and a value of 0 when its below.

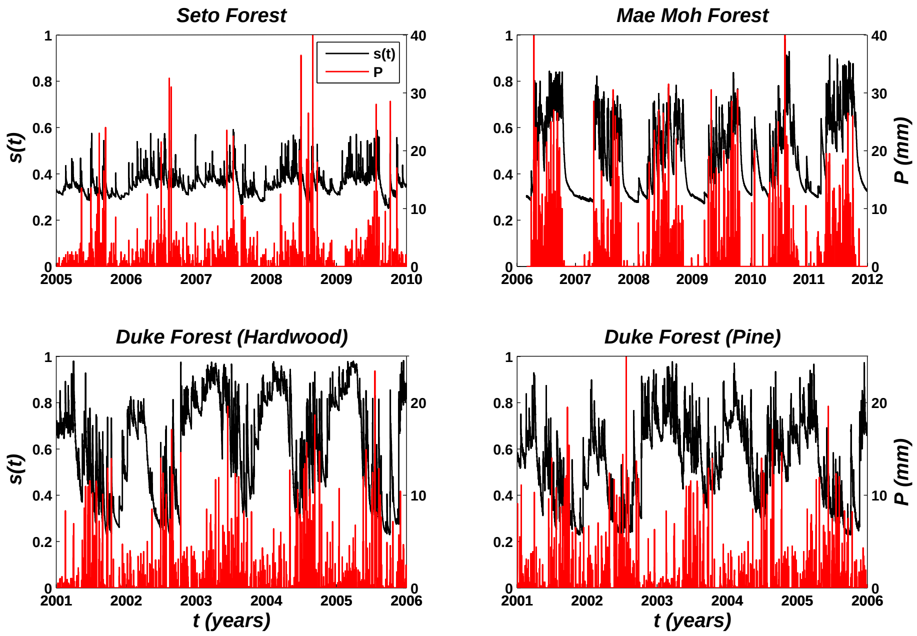

Figure 2. Time series of measured effective (dimensionless and depth-averaged) soil moisture within the root-zone and rainfall at each site sampled at 30-min intervals.

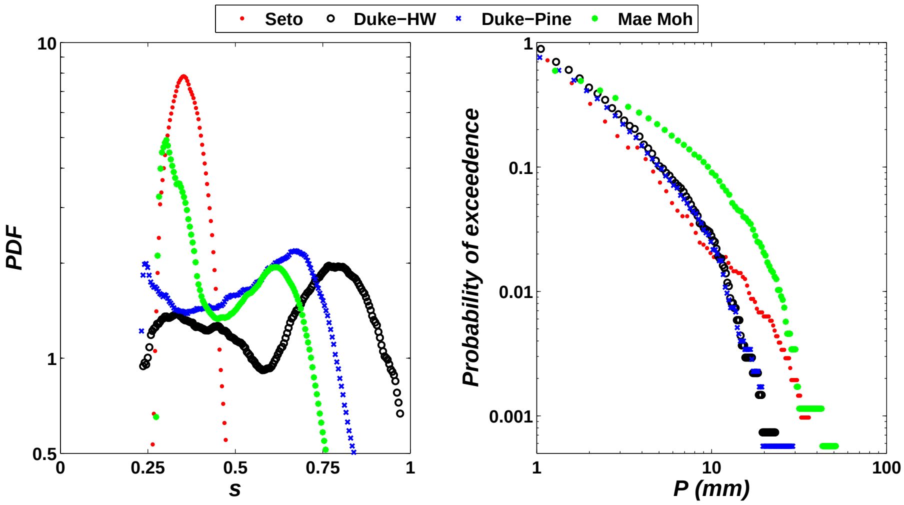

Figure 3. The probability density function (PDF) of soil moisture and probability of exceedance of rainfall (> 1mm) for the measurements in Figure 2. Note the bimodality at all sites except at Seto forest, while extreme rainfall events are more likely at Mae Moh and Seto forests.

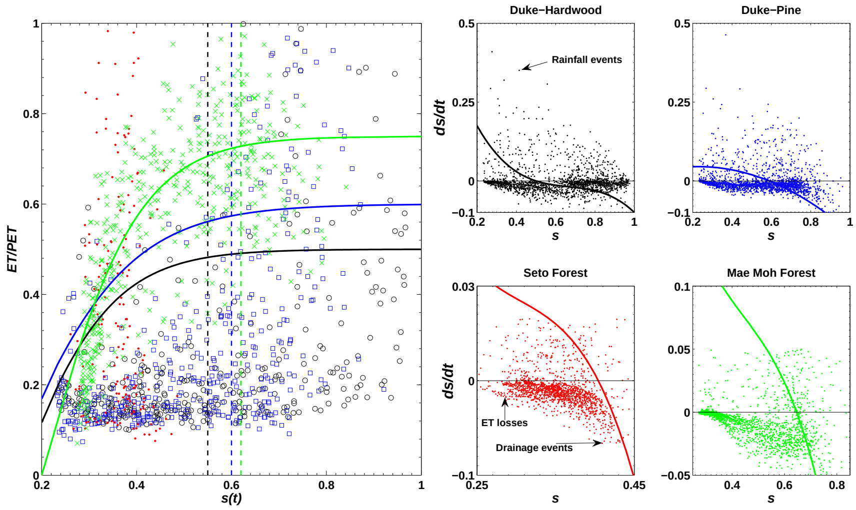

Figure 4. Left panel: The dependence of measured ET on measured effective soil moisture (symbols) along with a sigmoidal function ($a \tanh(s)$) fit (solid lines). PET is potential evapotranspiration calculated from the micro-meteorological measurements and aggregated to daily values. The dashed vertical lines correspond to the threshold s^* for each dataset. The colors in the left and right panels are equivalent, i.e. Duke-Hardwood (black), Duke-Pine (blue), Seto forest (red), and Mae Moh forest (green). Right panel: vector field representation of ds/dt as a function of s aggregated to the daily timescale. Large positive and negative values of ds/dt are associated with quasi-instantaneous rainfall and drainage events. The lines in the right panel are cubic fits to the function $f(s)$ in the equation $ds/dt = f(s)$. The intersection between these cubic functions and the $ds/dt = 0$ line represents a stable fixed point for each data set.

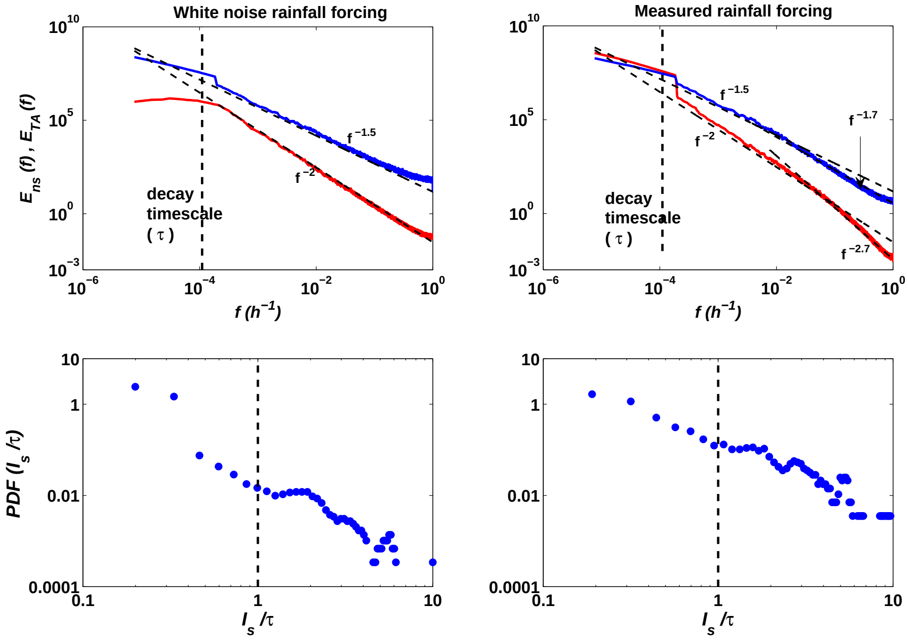

Figure 5. Persistence and crossing dynamics for a simulated first-order Markov process (equation (15)) with a white noise rainfall forcing (left column) and measured time series of rainfall at Duke forest-HW (right column). Upper panel: The normalized spectra of the stochastic process $s(t)$ (red color) and its TA (blue color) (shifted vertically for clarity). The vertical dotted line represents the decay frequency ($1/\tau$). Bottom panel: The PDF of the normalized persistence times (I_s/τ) for the two forcing cases.

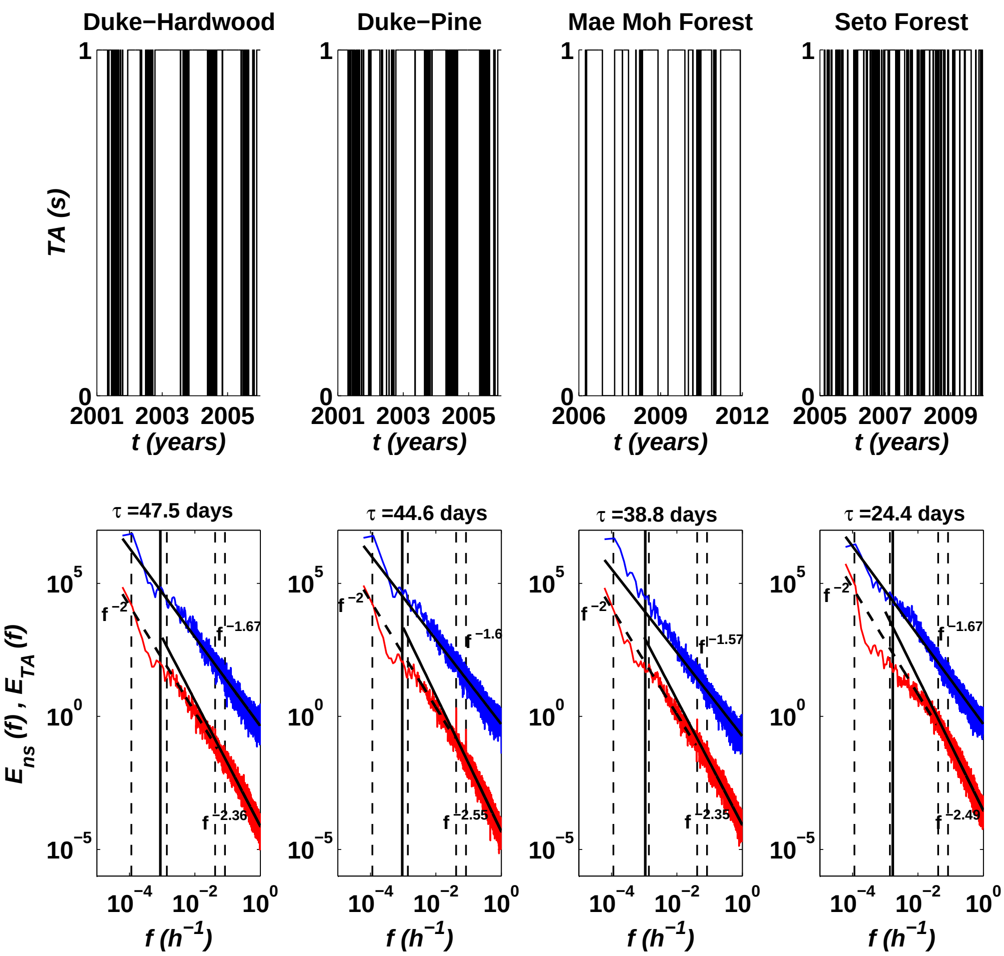

Figure 6. Top panel: Time series of the telegraph approximation of soil moisture $TA(s)$ around the threshold s^* for each site. TA is binary assuming values of 0 (below) or 1 (above) when comparing s with s^*. Bottom panel: The normalized power spectra of soil moisture $E_{\rm ns}(f)$ (red color) along with its TA spectrum, $E_{TA}(f)$ (blue color) for the sites in the upper panel. The TA spectrum was shifted on the y-axis to illustrate power-law exponents. The dashed vertical lines in each plot represent, from right to left, frequencies corresponding to diurnal (12 h), daily (24 h), monthly (720 h), and annual (8760 h) timescales, respectively. The solid vertical lines are the corresponding memory timescale τ for each site. The power-law fits are also shown.

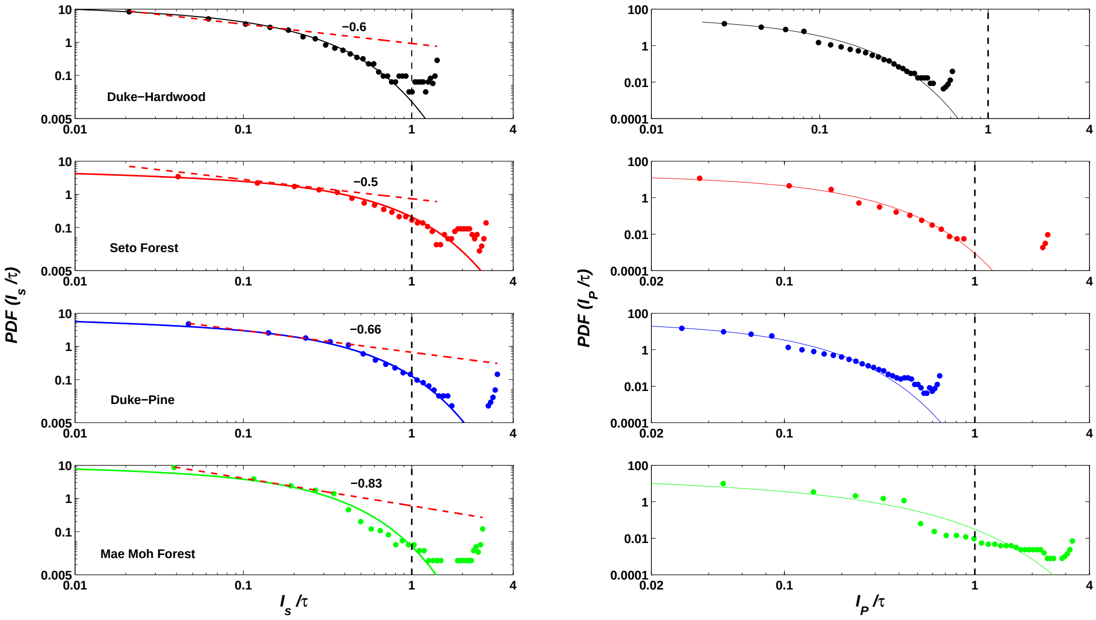

Figure 7. The probability density functions of persistence times of soil moisture below s^* (I_s) (left panel) and inter-arrival times between rainfall events (I_P) (right panel), both normalized by the corresponding soil moisture memory τ at each site. The solid lines are stretched exponential fits (see equation 21) to these distributions, with a value of b ranging from 0.8 to 0.9 for all sites. The dashed lines (red color) in the left panel are power-law fits (slope shown) to the first part of the PDF. The memory timescale τ for each soil moisture series is 47.5, 44.6, 38.8 and 24.4 days for Duke-HW, Duke-PP, Mae Moh, and Seto forests respectively.

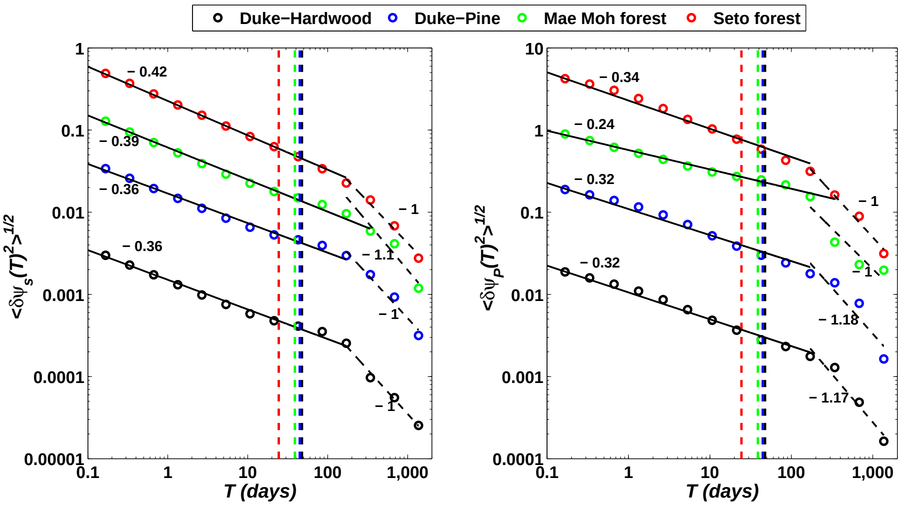

Figure 8. The relation $\langle \delta\psi(T)^2 \rangle^{1/2} \sim T^{-\omega}$ for soil moisture (left panel) and rainfall (right panel). The slopes (log-scale) of the power-law fits represent the cluster exponent w. The vertical dashed lines are the soil moisture memory scales for each site.

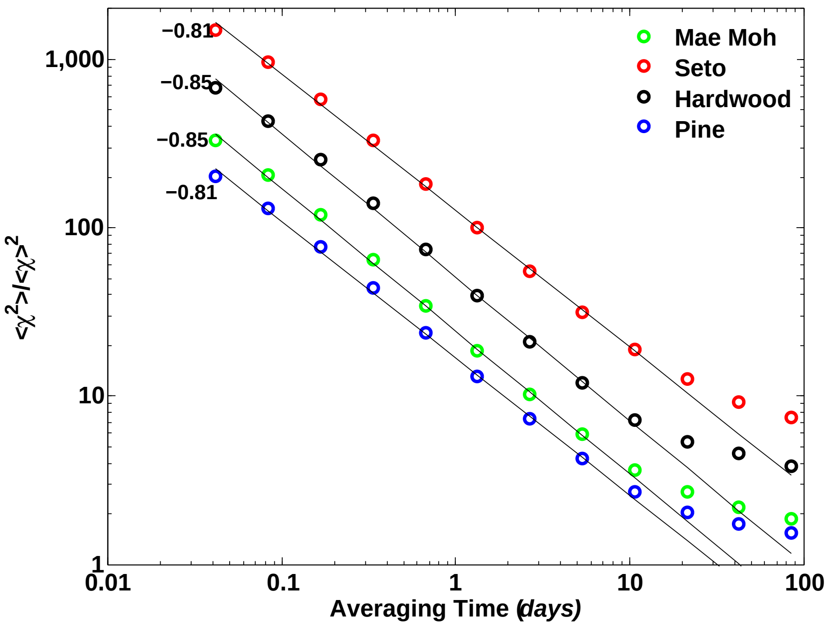

Figure 9. The intermittency exponent analysis for all sites. See text and equations (25) and (26) for explanation. The solid black lines represent power-law fits to the data. The scaling of the y-axis for Seto and Hardwood data was shifted vertically for clarity.

Table 1. Site description of the three forest sites.

Latitude band	Subtropical	Midlatitude		
Site	Mae Moh forest	Duke forest		Seto forest
Country	Thailand	USA		Japan
Climate	Tropical monsoon	Warm-temperate		Warm-temperate
Land use	Teak plantation	Hardwood stand	Loblolly pine plantation	Second-growth forest
Location	18°25′23″N	35°58′41″N	35°58′41″N	35°15′29″N
	99°43′05″E	79°08′ 39″W	79°05′ 39″W	137°04′54″ E
Forest type	Deciduous broadleaf	Mixed-species deciduous	Overstory: evergreen	Evergreen and
			Understory: mixed	deciduous mixed
Forest age[a] (year)	38	85-105	23	70–80
Dominant species	*Tectona grandis* Linn. f.	*Carya*	*Pinus taeda*	*Quercus serrata*
		Quercus	*Liquidambar styraciflua* L.	*Evodiopanax innovans*
		Other deciduous	Understory:	*Ilex pedunculosa*
			26 different species	*Symplocos prunifolia*
Stand density (trees ha^{-1})	343	930	3200	1900
Canopy height (m)	21.2[a]	35.0	20.0	9
Throughfall ratio	0.925	0.6	0.6	0.8
Root-zone depth R_L (mm)	400	300	300	650
Soil porosity η (–)	0.84[b]	0.55	0.55	0.62
Data period	Mar 2006 – Feb 2012	Jan 2001 – Dec 2006	Jan 2001 – Dec 2006	Jan 2005 – Dec 2009
References	*Yoshifuji et al.* [2006, 2014]	*Katul et al.* [2007]; *Oishi et al.* [2013]		*Matsumoto et al.* [2008]

[a]As of 2006.

[b]Determined from the maximum of the observed soil moisture data.